Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Articulatory properties of period-doubled voice in Mandarin

Yaqian Huang

Department of Linguistics, University of California, San Diego, USA, 92093
yah101@ucsd.edu

Abstract

Period-doubled phonation is a type of creaky voice that contains two alternating periods. By presenting data from Mandarin Chinese read speech recordings, this study probes the articulatory properties of period-doubled phonation and its tonal distribution based on time-domain measures using electroglottography (EGG). Period doubling (PD) was found across all the tones (T3: 43% > T2 > T4 > T1: 11%), which was more prevalent than vocal fry, found mainly in T3 (48%) and T2 (43%), and only sporadically in T4 (7%) and T1 (2%). We calculated the two alternating glottal periods in PD, and they exhibited a ratio close to 3:2 or 2:1. The two pulses also alternated between higher and lower amplitudes with a mean ratio approximating 2 or 1.6. Women tended to produce more PD than men. Moreover, the contact quotient of PD, measured via EGG using the hybrid method, was around 0.5, similar to modal voice (0.54) and smaller than that of vocal fry (0.74), implying a more balanced opening and contact phase during phonation. Alternation of contact quotient and symmetry quotient was also seen in a few samples, suggesting that PD is likely articulated through two alternating pulses with distinct voice qualities and pitches.

Index Terms: period-doubled phonation, creaky voice, tone, pitch, fundamental frequency, electroglottography

1. Introduction

Period-doubled phonation refers to a subtype of creaky voice that contains two simultaneous periodicities, which contribute to an indeterminate pitch with a low and rough quality [1,2,3]. Also known as multiply-pulsed voice, it is often additionally characterized as having increased noise and glottal constriction [1], as well as by the presence of subharmonics (or interharmonics) in the spectrum [2,4,5,6]. Production studies that have documented period doubling (PD) have often found that the periods and amplitudes alternate between two states (high and low; long and short) in the speech signal [7,8].

However, apart from limited acoustic and perceptual descriptions, there lacks a systematic analysis of the physical properties of PD needed to quantify the regularly alternating pattern in this subtype of creaky voice in natural speech. The defining characteristics of PD have important implications for the interaction between pitch and voice quality, and between tone and phonation as linguistic functions. For example, [9] found that English listeners perceive lower pitch during utterances that at least partially contain creak with multiple pulsing. In Mandarin, though non-contrastive, laryngealization (including vocal fry and PD) has been found to carry phonological function by signifying low tones more than other tones [10], and boundary phenomenon such as being a marker of utterance finality [11]. Also found are correlations between creaky voice, vocal attractiveness [12], and sarcastic speech [13]. But whether PD and vocal fry differ in their phonological or social functions remains unclear. In this paper, we study the articulatory properties of PD, with reference to more established voice categories – vocal fry and modal voice – to better capture the alternating glottal cycles that occur during PD. The representative period-doubled voice samples used are utterances where two periods alternate in frequency and/or amplitude as they appear in the electroglottographic waveforms from a read speech corpus in Mandarin.

2. Methods

Here we describe the methodology and present instances of period-doubled phonation in Mandarin. The focus is on the identification of PD in the voice source, as measured by electroglottographic (EGG) signals, and on introducing measures for quantifying the EGG pulse shapes.

2.1. Materials

We recorded participants reading Mandarin sentences [14], which were designed for investigating tonal coarticulation in the language. The carrier phrase was *wǒ jiāo nǐ* STIMULUS *zěn me shuō* 'I teach you STIMULUS how to say.' The stimuli consisted of trisyllabic Mandarin compounds, where each of the four Mandarin tones was flanked by varying Mandarin Tones 1–4, for a full range of 4*4*4=64 combinations. Three sets of 64 sentences with two repetitions (384 sentences in total) were elicited per recording. The "neutral tone" was not investigated for the current purpose.

2.2. Participants and procedure

Twenty native Mandarin speakers (10 females; mean age: 20.1; range: 18-22) participated. Fifteen were from northern provinces and five from southern provinces in China. No language or hearing disorders were reported. They were instructed to produce 384 sentences at a comfortable speed as if they were participating in a natural conversation. Speakers were recorded in a sound-attenuated booth at the UCSD Phonetics Lab. They wore a head-mounted microphone and a collar with two electrodes placed around the neck for recording the degree of contact between the vibrating vocal folds directly from the larynx. EGG was recorded using a Glottal Enterprises EG2-PCS. Both audio and EGG signals were pre-amplified through a Focusrite Scarlett preamplifier and digitized using the computer's sound card using Audacity at a 44,100 Hz sampling rate and 32-bit float rate.

2.3. Analysis

2.3.1. *Criteria for identifying period doubling*

We used PRAAT [15] to segment the phrases in the audio signal and used the EGG recordings to locate source pulses with period-doubled voice and vocal fry. The advantage of

 10.21437/SpeechProsody.2022-111

using the EGG signal instead of the audio to identify PD lies in the absence of formants, which avoids possible formant-induced interferences with the voicing signal. Based on [16], canonical period doubling is often characterized by sequences of two different cycles which differ in amplitude of the pulses, or length of the periods or frequency; in other words, by amplitude modulation and frequency modulation. For example, amplitude-modulated PD results in “high-low-high” alternating amplitudes, whereas frequency-modulated PD results in “long-short-long” alternating periods in the waveform. Figures 1 and 2 show sample EGG waveforms of typical-looking PD found in the corpus (from two women). Differences between sub-cycles of PD were on a continuum, which sometimes can be subtle. To be as inclusive as possible, we labeled any tokens that resemble Figures 1 or 2, or when it causes pitch halving in PRAAT’s f0 algorithm. The recordings were pre-processed using a built-in band-pass filter in PRAAT to remove the low-frequency EGG component below 40 Hz. Note that the instances of frequency-modulated PD tend to cooccur with amplitude modulation, similar to what [16] reported. Also, though period-doubled phonation in pathological speech and certain singing styles involves the vibrations of ventricular folds, [17] found that EGG exclusively encodes information about the two consecutive glottal cycles, so it is unlikely the PD observed here is due to ventricular fold vibrations.

2.3.2. *Frequency and amplitude ratio*

To quantify the differences in the two periods in period-doubled voice, we obtained the duration and amplitude of each alternating strong and weak adjacent cycles during samples of period doubling containing at least one pair of cycles.

In any pair of pulses, there are three periods. Following [17], who investigated period-doubled phonation in singing, here we define T_0 as the longest period, from strong-to-strong pulses, which is also the presumptive fundamental cycle. For the two shorter periods, the strong-weak cycle is defined as T_l and the other weak-strong cycle is defined as T_s. As (1) shows, the duration of the two shorter cycles adds up to the longer duration T_0, namely, $T_l + T_s = T_0$. The frequency-ratio R_T is defined as the ratio of the duration of the two alternating glottal cycles. Namely, the longer strong-to-weak cycle (T_l) divided by the shorter weak-to-strong one (T_s), or the higher frequency f_h over the lower frequency f_l; the formula of which is in (2). An illustration is shown in Figures 1 and 2. The frequency ratio R_T thus indicates the relative relationship between any of the two short periods, which together could contribute to the fundamental cycle and/or frequencies.

The amplitude ratio R_A, defined as the higher amplitude A_l divided by the lower one A_s, is shown in (3). Here, A_l and A_s are defined as the total displacement from the minimum to the maximum of a glottal cycle, to better capture the strength of each excitation. The top panels of Figures 1-2 illustrate the calculation of the respective amplitude values.

The time points for different periods at each rising zero-crossing, and the amplitude values at the extremes of the stronger and weaker pulses were extracted using a custom PRAAT script and the ratio of the periods/amplitudes was further calculated in R. After the calculation, we removed outliers if they had a z-score larger than 2.5 standard deviations away from the mean within each subject.

$$T_l + T_s = T_0 \quad (1)$$

$$R_T = \frac{T_l}{T_s} = \frac{f_h}{f_l} \quad (2)$$

$$R_A = \frac{A_l}{A_s} \quad (3)$$

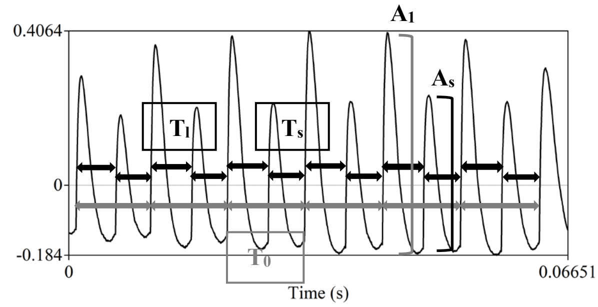

Figure 1: *EGG Waveform of amplitude-modulated period doubling. Short black lines represent the smaller periods T_l and T_s, and long grey lines represent the fundamental period T_0. The black and grey brackets show the overall displacement of the smaller pulse A_s and the larger pulse A_l respectively.*

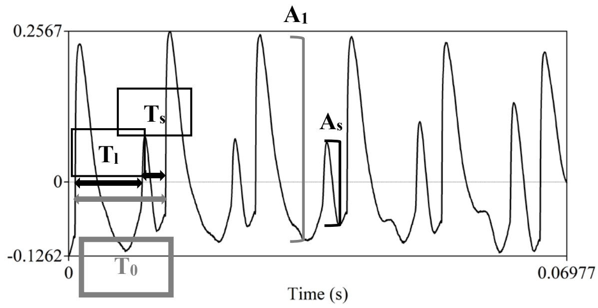

Figure 2: *EGG waveform of frequency and amplitude modulated period doubling. Note that there is a larger difference between the duration of the two shorter periods T_l and T_s (represented by the black lines) than the example in Figure 1. The black and grey brackets show the overall displacement of the smaller pulse A_s and the larger pulse A_l respectively.*

2.3.3. *Waveform analysis: contact and speed quotient*

Here we analyze pulse shapes in terms of their articulatory properties to determine the contact quotient (CQ) and speed quotient (SQ) measured from the EGG signal. Both were obtained from EGGWorks [18].

CQ is defined as the proportion of the cycle during which the vocal folds are in contact. We used a hybrid method which uses the derivative of the EGG to determine the onset of contact and an amplitude threshold of 25% for the onset of opening phase [19]. Higher values of CQ thus indicate a higher degree of constriction. Creaky voice is often characterized by vocal fold constriction, as in prototypical creak and vocal fry [1]. [1] also states that period-doubled pulses generally have a very long closed phase. To investigate the degree of constriction of period doubling, we compared the CQ of period doubling with modal voice and vocal fry found in the corpus. We extracted modal voice samples of ~50 ms from speech adjacent to period-doubled voice for comparison. We also identified vocal fry by its defining characteristics such as having low f0, glottal constriction, and high damping [1]. See Figure 3 for an example of vocal fry (from a woman) in the EGG waveform.

Speed Quotient (SQ) [20,21,22], defined as the ratio between the duration of the contacting and decontacting phases, is informative of the symmetry of the glottal pulses. In creaky voice, SQ is often smaller due to a rapid contacting gesture and a slow decontacting phase. Thus, we might expect vocal fry to have a more skewed glottal pulse shape comparing to modal voice. Period doubling, as a subtype of creaky voice, might be closer to the pattern for vocal fry.

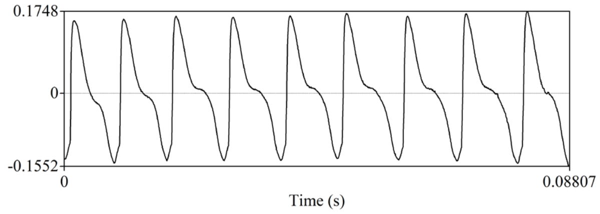

Figure 3: *EGG waveform of vocal fry (CQ = 0.77).*

3. Results and discussion

In this section, we summarize and discuss the findings of articulatory properties of period-doubled phonation. We discuss the frequency and amplitude ratios of the alternating adjacent cycles in period doubling, and the contact quotient and speed quotient of period doubling in comparison to vocal fry and modal voice.

3.1.1. Instances of period doubling in the corpus

There were numerous instances of period doubling found in the corpus. A t-test comparing the number of tokens identified in women vs. men's voice shows that women had produced more period-doubled utterances than men with a mean of 534 versus 157 tokens (p<.005), consistent with [2,23]. Women also had longer voice samples of period doubling than men for an average of 3.8 cycles per utterance (p<.001), and a maximum of 22 cycles. Within an utterance of a certain duration, there are fewer cycles if the frequency is lower; thus, this finding is expected, given that male voices have lower fundamental frequencies or longer periods.

3.1.2. Distribution of period doubling and vocal fry across Mandarin tones

We labeled the occurrences of period doubling and vocal fry in the trisyllabic stimuli. There were 1140 period-doubled tokens found across all the tones: T3: 43% > T2 > T4 > T1: 11%, which was more prevalent than vocal fry. Only 235 vocal fry tokens were found within the phrase-medial stimuli: T3 (48%) and T2 (43%), and only sporadically in T4 (7%) and T1 (2%). Figure 4 shows the tonal distributions. In T3, period doubling tended to occur throughout the contour whereas in other tones, it tended to occur at beginnings or ends.

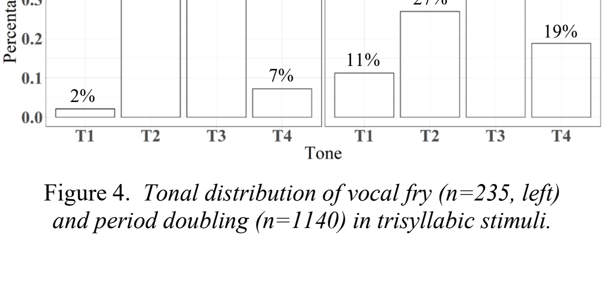

Figure 4. *Tonal distribution of vocal fry (n=235, left) and period doubling (n=1140) in trisyllabic stimuli.*

3.1.3. Frequency ratio

To test the relationship and the relative time courses of the two alternating cycles, we focused next on the ratio RT between the higher and lower glottal frequencies, given by T_l/T_s. (Note that we chose to start the analysis with the stronger and longer cycle T_l is an arbitrary choice.) Data from the 20 speakers showed the mean RT was around 1.26-2.04, such that the weaker pulses occur later in the course of articulation, after the midpoint of T_0. To preclude incorrectly identified period doubling from normal jitter found in modal voice, we calculated the pulse-to-pulse variation in duration based on the adjacent samples of modal voice and removed the period-doubled cycles whose values fall within the jitter range for modal voice (0.997~1.02), which were 0.05% of the data.

We then looked further into the frequency ratio in the two different types of period doubling, amplitude modulation and frequency modulation. In period-doubled voices like Figure 1, the average frequency ratio RT was 1.44 (~3/2), meaning that the sub-pulse fell closer to the midpoint of the fundamental cycle; in samples like Figure 2, the average RT was 1.9 (~2/1), closer to 2/3 of the fundamental cycle. The two alternating glottal cycles thus exhibit a ratio approximating 3:2 or 2:1, depending on the types of period doubling.

We also asked how the fundamental and the two glottal frequencies compared to f0 in modal voice. The results revealed the same pattern across speakers: in period doubling, both the f0 and the first higher frequency f_l (=1/T_l) was lower than the f0 during modal voice, which was in turn lower than the second higher frequency f_h (=1/T_s): f0 (period-doubled) < f_l < f0 (modal) < f_h. The f0 in period doubling is almost half of the f0 in modal voice, and the f0 in modal voice is near the mean of the two higher glottal frequencies.

3.1.4. Amplitude ratio

We also calculated shimmer from the normal pulse-to-pulse variation in amplitude in modal voice, and excluded the period-doubled cycles whose amplitude ratios were within the modal shimmer range (0.98~1.06), which were 5% of the data. The resulting average of the amplitude ratio is 2.03 (±0.71).

The amplitude ratio RA also varies according to the type of period doubling, depending on whether it is amplitude modulated or frequency modulated: results indicate that the difference in amplitude was larger in amplitude modulated tokens. In addition, the signals from women were lower in amplitude, which is expected because EGG signals tend to be stronger in men due to their narrower angles formed by the sides of the thyroid cartilage. Table 1 summarizes the descriptive statistics of frequency and amplitude ratios. In general, RA exhibits more variability by having a higher SD, especially in amplitude-modulated tokens and in women.

Table 1: *Means and standard deviations of frequency and amplitude ratio. "Am" stands for amplitude-modulated and "Fm" stands for frequency-modulated period-doubled tokens.*

	Overall mean	**Am**	**Fm**	**Women**	**Men**
RT	1.52	1.44	1.90	1.55	1.48
	(0.23)	(0.13)	(0.34)	(0.24)	(0.15)
RA	2.03	2.07	1.63	1.75	1.71
	(0.71)	(0.78)	(0.35)	(0.81)	(0.55)

3.1.5. *Contact quotient (CQ)*

We used linear mixed-effects model comparisons for CQ with and without the category of voice (period doubling vs. fry vs. modal), with a random intercept and slope by subject, which showed significant differences (p<.001). As expected, vocal fry had a higher CQ than modal voice (β=0.08, p<.05). Figure 5 shows the distinct distributions of CQ by phonation category. The distribution for period doubling overlapped largely with modal voice, and it had a slightly lower CQ than modal voice (β=-0.03, p<.01). The mean value of CQ for modal was 0.54 (SD: 0.07), for vocal fry was 0.74 (SD: 0.12), and for period doubling was 0.54 (SD: 0.11). The bimodal distribution of modal voice is by-speaker variation.

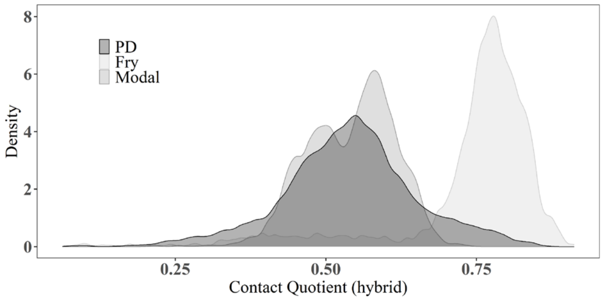

Figure 5: *Density distributions of contact quotient in modal voice (light grey), vocal fry (lighter grey), and period doubling (dark grey).*

Note that in amplitude-modulated tokens, the mean CQ of period doubling was 0.52, showing more balanced contact and open phases; in frequency-modulated tokens, the mean CQ of period doubling was 0.6, showing that the duration of vocal fold contact is slightly longer. This implies that in period doubling, the open and contact phases are more temporally balanced, and that the voice quality is not particularly constricted. This differs from what would typically be expected from other types of creaky voice such as vocal fry, which are characteristically constricted. Still, these findings are consistent with the high open quotient found in subharmonic vibration [24]. In addition, an alternation of CQ values was seen in period doubling. Figure 6 plots a sustained sample of such an alternation, suggesting that the degree of constriction changes from pulse to pulse in period doubling.

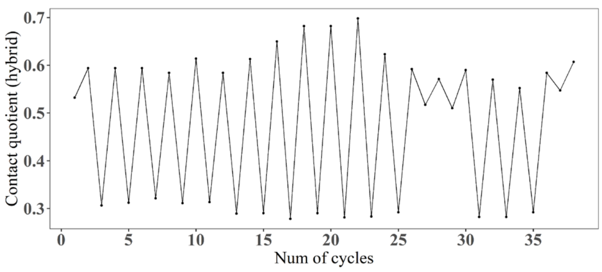

Figure 6: *The contact quotient of each cycle in a sustained sample of period doubling.*

3.1.6. *Speed quotient (SQ)*

SQ indicates the symmetry of pulse shapes, such that smaller values mean that the contacting phase is shorter than the decontacting phase, as occurs in vocal fry with a more rapid contacting phase. Using the same linear mixed-effects model structure as in CQ, vocal fry had a lower SQ than modal voice (β=-0.18, p<.001) whereas period doubling did not differ from modal. Figure 7 shows the distributions of SQ values in modal voice, vocal fry, and period doubling. Amplitude-modulated tokens had a mean SQ of 0.3, similar to the SQ of modal voice (0.32), and the frequency-modulated tokens had a mean SQ of 0.14, similar to the SQ of vocal fry (0.07).

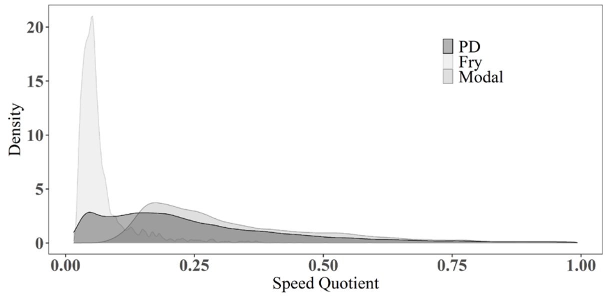

Figure 7: *Density distributions of speed quotient in modal voice (light grey), vocal fry (lighter grey), and period doubling (dark grey).*

In addition, alternation of SQ between different modes was seen in period doubling. Figure 8 plots a sustained example that shows alternation of SQ values, meaning that the relative duration of the contact and decontacting phases changes from pulse to pulse.

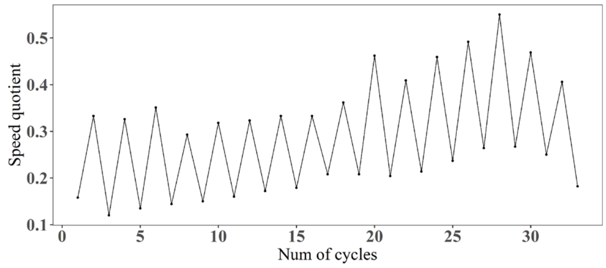

Figure 8: *The speed quotient of each cycle in a sustained sample of period doubling.*

In sum, we see that period-doubled phonation is found in all Mandarin tones, more abundant than vocal fry occurrences, and although it happens more frequently in women's voice, the frequency or amplitude ratios do not differ among different genders. More importantly, producing speech with period doubling involves phonation of two distinct glottal pulses which not only regularly alternate between their periods and amplitudes: strong vs. weak and/or short vs. long, but between more vs. less constriction and rapid vs. slow contacting. Importantly, these articulatory properties of period doubling essentially speak to the co-existence of two distinct glottal pulses with different voice qualities and pitches, which makes period doubling a distinct voice category from both vocal fry and modal voice. The concurrent voice qualities likely contribute to the indeterminate pitch – as well as quality – that one perceives during period doubling. In future work, we will probe the role of period doubling in pitch and tone perception, and the use of period doubling in various linguistic contexts.

4. Acknowledgements

The author would like to thank Marc Garellek, Will Styler, and Jody Kreiman for their feedback, Hira Rizvi and Shogo Nishimura for assisting the segmentation of EGG files, and three reviewers for suggestions. This material is based upon work supported by the NSF under Grant No. BCS-2141433.

5. References

[1] P. A. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," *Proc. 18th ICPhS*, pp. 0821-1, 2015.

[2] K. M. Yu, "Laryngealization and features for Chinese tonal recognition," in *Eleventh Annual Conference of the International Speech Communication Association*, pp. 1529–1532, 2010.

[3] D. Schreibweiss-Merin, and L. M. Terrio, "Acoustic analysis of diplophonia: a case study," *Perceptual and motor skills*, 63(2), pp. 755–765, 1986.

[4] K. Omori, H. Kojima, R. Kakani, D. H. Slavit, and S. M. Blaugrund, "Acoustic characteristics of rough voice: subharmonics," *Journal of Voice*, 11(1), pp. 40–47, 1997.

[5] E. Kramer, R. Linder, and R. Schönweiler, "A study of subharmonics in connected speech material," *Journal of Voice*, 27(1), pp. 29 – 38, 2013.

[6] C. T. Herbst, "Performance Evaluation of Subharmonic-to-Harmonic Ratio (SHR) Computation," *Journal of Voice*, 2020.

[7] I. R. Titze, "Fluctuations and perturbations in vocal output," *Principles of voice production*, pp. 209–306, 1994.

[8] B. R. Gerratt, and J. Kreiman, "Toward a taxonomy of nonmodal phonation," *Journal of Phonetics*, 29(4), pp. 365–381, 2001.

[9] L. Davidson. "Contributions of modal and creaky voice to the perception of habitual pitch," *Language*, 96(1): e22-37, 2020.

[10] A. Belotel-Grenié, and M. Grenié, "Types de phonation et tons en chinois standard" ("Phonation types and tones in standard Chinese"), *Cah. Ling. Asie Orient*, 26, pp. 249–279, 1997.

[11] A. Belotel-Grenié, and M. Grenié, "The creaky voice phonation and the organisation of Chinese discourse," in *Proceedings of TAL-2004*, Beijing, China, pp. 5–8, 2004.

[12] A. Xu, and A. Lee, "Perception of vocal attractiveness by Mandarin native listeners", in *Proceedings of the International Conference on Speech Prosody*, Poznan, Poland, pp. 344–348, 2018.

[13] S., Li, W. Gu, L. Liu, and P. Tang, "The role of voice quality in Mandarin sarcastic speech: An acoustic and electroglottographic study," *Journal of Speech, Language, and Hearing Research*, 63(8), pp. 2578–2588, 2020.

[14] Y. Huang, and Y. Chai, "Voice quality of coarticulated Mandarin tones," *The Journal of the Acoustical Society of America*, 146(4), pp. 3010-3010, 2019.

[15] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.0.31, retrieved 21 August 2017 from http://www.praat.org/, 2017.

[16] M. Mazo, D. Erickson, and T. Harvey, "Emotion and expression: Temporal data on voice quality in Russian lament," in *The 8th Vocal Fold Physiology Conf., Kurume, Japan*, pp. 173–187, 1995.

[17] L. Bailly, N. Henrich, and X. Pelorson, "Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling," *The Journal of the Acoustical Society of America*, 127(5), pp. 3212 – 3222, 2010.

[18] H. Tehrani, "EGGWorks: A program for automated analysis of EGG signals," Program, 2009.

[19] D. M. Howard, "Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers," *Journal of Voice*, 9(2), pp. 163 – 172, 1995.

[20] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *Journal of Speech and Hearing Research*, vol. 38, pp. 1212–1223, 1995.

[21] C. Dromey, E. T. Stathopoulos, and C. M. Sapienza, "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," *Journal of Voice*, vol. 6, 44–54, 1992.

[22] J. Kuang, and P. A. Keating, "Vocal fold vibratory patterns in tense versus lax phonation contrasts", *The Journal of the Acoustical Society of America* 136(5), pp. 2784 – 2797, 2014.

[23] P. Hedelin, and D. Huber, "Pitch period determination of aperiodic speech signals," in *International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, NM, pp. 361–364, 1990.

[24] J. G. Švec, H. K. Schutte, and D. G. Miller, "A subharmonic vibratory pattern in normal vocal folds," *Journal of Speech, Language, and Hearing Research*, 39, no. 1, pp. 135–143, 1996.