THE INFLUENCE OF COMMON COLDS AND ITS INTERACTION WITH LEXICAL TONES ON VOICE QUALITY OF MANDARIN SPEAKERS

Yifan Mou[1], Lei Zhu[2]

[1, 2] Institute of Linguistics, Shanghai International Studies University, China
mouyifan@shisu.edu.cn, zhulei@shisu.edu.cn

ABSTRACT

This study examines the influence of common colds and its interaction with lexical tones on the voice quality of Mandarin speakers. We find that speakers with common colds have a significantly lower CPP and a higher Jitter compared with healthy speakers, implying a greater instability of vibrations and more noisy elements in voice production. Such changes are largely due to the inflammation of larynx and hypofunction of vocal folds caused by common colds. Besides, a significant interaction between health condition and tones is found for HNR and Jitter. Mandarin tones with curved pitch contours, including Tone2, Tone3 and Tone4, are more susceptible to colds, presenting a greater degradation of voice quality. This is mainly caused by the tendency for cold speakers to decrease their phonatory efforts and extra vocal efforts required by curved tones.

Keywords: voice quality, Mandarin tones, pathology

1. INTRODUCTION

Acoustic parameters obtained in non-invasive manners are objective indicators for speech-related diseases. The accuracy and robustness of acoustic parameters for disease detection have been confirmed in various studies, including asthma, Covid19, Parkinson's syndrome and Alzheimer's disease [1, 2]. Voice production requires the cooperation of vocal organs, such as pharynx, larynx and oral cavity [3]. However, speech-related diseases would disturb the function of vocal organs, causing significant acoustic differences between pathological speakers and normal ones [4]. These differences could be captured and quantified by various acoustic parameters [5].

Due to the connection with external environment, vocal organs are highly susceptible to viral infections [6]. Upper respiratory tract infection (URTI) is one of the most common diseases, causing three to five million serious infections each year [7]. Upper respiratory tract infection is an acute infection involving the sinuses, pharynx, larynx, trachea, and bronchi [8, 9]. The common cold is the most common symptom of upper respiratory tract infections. Adults are reported to catch two to five colds each year, while children might catch seven to ten colds annually [10, 11].

Early symptoms of common colds are headache, sneezing, chills and sore throat, and the later symptoms involve nasal congestion, cough and hoarseness [9, 12]. Even mild symptoms of colds could affect the function of vocal cords, causing temporary speech disorders including worse voice quality, vocal fatigue and reduced pitch range. Severe inflammation of the larynx caused by colds may lead to a long-term, or even persistent damage to vocal function [13, 14].

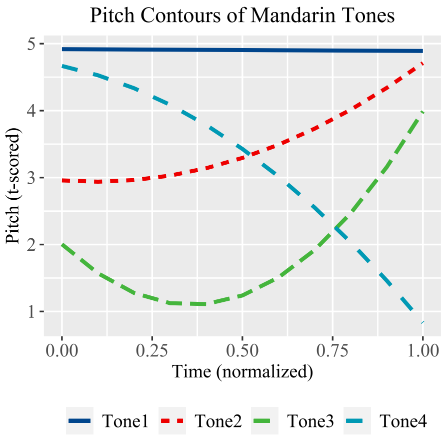
Previous phonetic studies on common colds focused on the variation of formants, fundamental frequency, spectral peaks and Mel-frequency cepstral coefficients between speakers with common colds and healthy ones. Related works have proven that frequencies of the first formant (F1) and the second formant (F2) of vowels get lower for the cold condition than for the healthy condition. The second Mel-frequency cepstral coefficient also gets significantly lower in the state of common cold [15]. The spectral peak near 1KHz in nasal consonants shows significant decrease for speakers with common colds as well. As a measurement of nasality, it suggests a higher nasal congestion for cold speakers [16]. Besides, the pitch of vowels is lowered during a cold, and an increase in noise is found for stops and liquids [6].

It has been noticed that speakers with common colds could have different voice quality compared with healthy people [17]. However, previous studies paid less attention to the effect of common colds on voice quality, which has been proven to be reliable indicator reflecting the state of the glottis and severity of voice disorder [18]. This study aims to fill this gap by exploring how common colds influence our voice quality, which helps uncover mechanisms of voice production for pathological conditions and improve the diagnosis of speech-related diseases. Based on previous studies, we hypothesized that there exists a significant main effect of health condition on voice quality. Common colds would lead to a worse performance of voice quality. According to the protocol of the European Laryngology Society [19] and American Speech-Language-Hearing Association [18], four parameters of voice quality, including harmonic-to-noise ratio (HNR), cepstral peak prominences (CPPs), Jitter and Shimmer, were included in the study.

We also noticed that recording materials used in previous works on common colds were mainly sustained vowels with flat pitch contours. Mandarin has four lexical tones classified by their pitch heights and pitch contours. Mandarin tones include Tone 1 (high-flat), Tone 2 (mid-rising), Tone 3 (falling-rising) and Tone 4 (high-falling) [20], as shown in Figure 1.

In previous research on tonal languages, a significant interaction between tones and phonation was found. Different registers of tones show a clear phonation contrast with each other [21], which implies that variations in vocal effort exist for different tones. Considering the vocal fatigue and hypofunction of vocal folds for the cold condition, there is a tendency for speakers with common colds to adjust phonation strategy to reduce their phonatory efforts [16]. Thus, we proposed our second hypothesis that there exists an interaction between health condition and Mandarin tones. Some Mandarin tones might be more susceptible to common colds, presenting a greater degradation of voice quality.

Figure 1: Four types of Mandarin tones, plotted with normalized time and t-scored pitch height



2. METHOD

2.1 Data

Audio files in the study were obtained from the Common Cold Speech Dataset offered by Chinese Linguistic Data Consortium, containing a set of speech and corresponding transcriptions recorded by speakers with common colds and healthy speakers. Recordings were mono-channel speech saved in Wav format, with a sampling rate of 16KHz. Speech data were recorded with professional equipment in a quiet environment. Each phrase was repeated three times by participants to ensure a stable performance.

Recording materials included numbers, places and daily expressions. Almost all possible combinations of Mandarin syllables were covered in the design of recording materials, and there were sufficient recordings for all four Mandarin tones. All participants were native Mandarin speakers without any hearing loss or major voice damage. The dataset contained 92 participants with balanced gender and health condition distributions, including 45 common cold speakers and 47 healthy speakers. Half of participants were males, while another half were females. Each participant recorded approximately 600 words in average.

2.2 Data Processing

We used a pre-trained MFA [22] Mandarin acoustic model to perform forced alignment on the recorded speech and their corresponding transcriptions at the syllable level and the phoneme level. The Mandarin acoustic model has already been trained on various large-scale Mandarin datasets, with quite good performance and robust error rate. After obtaining the alignment, we then manually corrected the inaccurate boundaries of syllables and phonemes in the TextGrids. After all boundaries being marked and correctified, we used Praat's python interface *Parselmouth* [23] to extract parameters of voice quality from the voicing part (the rhyming part with lexical tones) of Mandarin syllables. Extracted data were subjected to outlier detection to ensure the results of linear mixed-effects models not being disturbed by outliers.

As we have mentioned in Section 1, four parameters of voice quality including harmonic-to-noise ratio (HNR), cepstral peak prominences (CPPs), Jitter and Shimmer were included in the study. HNR and CPPs mainly represent the proportion of noise in the sound, while Jitter and Shimmer mainly stand for the stability of voice production during the phonation stage. To improve the accuracy of parameter extraction in pathological condition, we adopted the cross-correlation method for the estimation of fundamental frequency, as suggested by the manual of Praat [24]. The default range for pitch detection was expanded to 50Hz (pitch floor) to 600Hz (pitch ceiling). Straight line was selected as the regression line to fit the overall cepstrum for calculating CPPs. Jitter (local) was measured by the average absolute difference between lengths of consecutive periods, divided by the average length. Shimmer (local) was measured by the average absolute difference between amplitudes of consecutive periods, divided by the average amplitude.

2.3 Statistical Analysis

In order to determine the effect of health condition and its interaction with lexical tones on voice quality of Mandarin speakers, we fitted four linear mixed-effects models for four parameters of voice quality respectively, using *lme4* package [25] in R [26]. Health condition (healthy vs. cold), tones (Tone1 to Tone4) and their interaction were included as fixed effects, while random effects were made of by-participant, by-gender, and by-item random intercepts. Health condition and tones were treatment-coded, with healthy condition and Tone1 as the reference level.

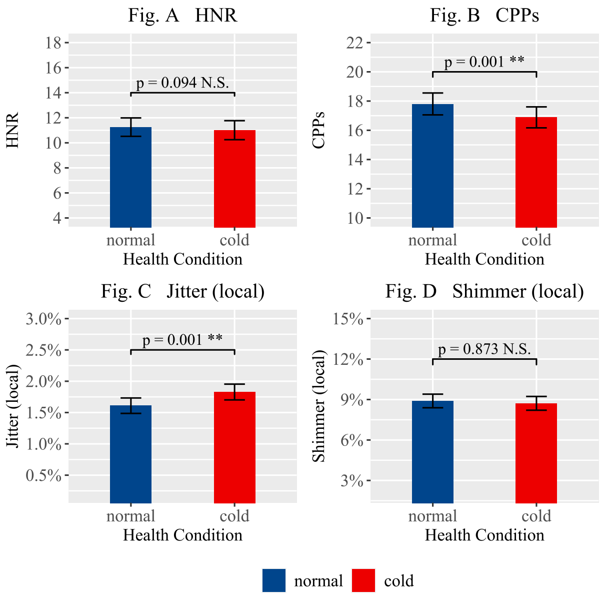
All models were fitted with the backward stepwise method by removing the random effect capturing the smallest variance until convergence [27]. Examination of main effects and interaction effects was based on the Type III analysis of variance, with *anova* function in *lmerTest* package [28]. Degree of freedom was calculated with the Satterthwaite's method. Multiple comparisons were calculated by *emmeans* package [29] with Bonferroni correction. The significance level was set at 0.05, and *p*-values were obtained with *lmerTest*.

3. RESULTS

3.1 Main effect of health condition on voice quality

Linear mixed effects analysis performed on the CPPs and Jitter shows significant differences for speakers under different health conditions, as shown in subplots B and C of Figure 2. Specifically, a significant main effect of health condition is found on CPPs [$F(1,90) = 11.441$, $p = 0.001^{**}$] and Jitter [$F(1,90) = 11.314$, $p = 0.001^{**}$], indicating that speakers with common colds have different CPPs and Jitter, compared with healthy speakers. Statistics reveals that health condition does not seem to cause significant differences in HNR [$F(1,90) = 2.856$, $p = 0.094$] and Shimmer [$F(1,90) = 0.0483$, $p = 0.873$], as shown in subplots A and D of Figure 2.

Figure 2: Error bar plots of HNR, CPPs, Jitter and Shimmer for different health conditions



Given the significant difference in CPPs and Jitter of speakers under different health conditions, a Bonferroni corrected post-hoc pairwise comparison was carried out. Results of post-hoc pairwise comparisons (see Table 1) show that speakers with common colds have a significantly lower CPP compared with healthy speakers [$\beta = -0.835$, $t = -3.383$, $p = 0.001^{**}$]. A lower CPP indicates an increase in noise during the process of voice production, probably caused by hoarseness and hypofunction of vocal folds. Besides, speakers with common colds have a significantly higher Jitter compared with healthy speakers [$\beta = 0.220$, $t = 3.364$, $p = 0.001^{**}$]. A higher Jitter indicates that more aperiodic vibrations are found in the phonation stage, which is largely due to asymmetrical changes in mass and tension of vocal folds under pathological condition. Coughs and inflammation of larynx, which cause discomfort and disturb vocal function, could lead to an increase in Jitter as well.

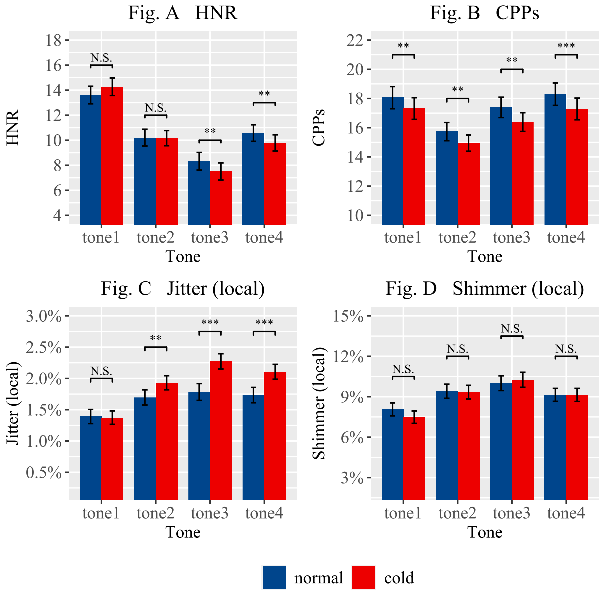
Table 1: Results of post-hoc pairwise comparisons for main effect of health condition. Estimates with significant p-values are put in bold.

parameter	*Contrast*	*β*	*S.E.*	*t.ratio*	*p*
HNR	cold - healthy	-0.764	(0.452)	-1.690	0.094
CPPs	cold - healthy	**-0.835**	**(0.247)**	**-3.383**	**0.001**
Jitter	cold - healthy	**0.220**	**(0.066)**	**3.364**	**0.001**
Shimmer	cold - healthy	-0.075	(0.340)	-0.220	0.873

3.2 Interaction between health condition and tones

Significant interaction effects between health condition and lexical tones are found for HNR [$F = 190.85$, $p < 0.001^{**}$] and Jitter [$F = 328.12$, $p < 0.001^{**}$], as shown in subplots A and C of Figure 3. No significant interaction effects are found in CPPs [$F = 0.694$, $p = 0.556$] and Shimmer [$F = 2.398$, $p = 0.066$], as shown in subplot B and D. (No significant interaction effect was found in CPPs, despite the presence of significant simple main effects of health condition in all four lexical tones.) Significant interaction effects in HNR and Jitter indicate that there are different simple main effects of health condition for voice quality in four Mandarin tones.

Figure 3: Error bar plots of HNR, CPPs, Jitter and Shimmer of four tones by different health conditions



Post-hoc pairwise comparisons (Table 2) were carried out to further examine the interaction between health condition and tones. Table 2 shows that difference of HNR between cold speakers and healthy speakers is significant for Tone 3 [$\beta = -1.402$, $t = -3.075$, $p = 0.003^{**}$] and Tone 4 [$\beta = -1.20$, $t = -2.651$, $p = 0.009^{**}$], but not for Tone1 [$\beta = 0.08$, $t = 0.177$, $p = 0.86$] and Tone2 [$\beta = -0.533$, $t = -1.164$, $p = 0.247$]. Results also confirm that the difference of Jitter between speakers with common colds and healthy speakers is significant for Tone2 [$\beta = 0.196$, $t = 2.881$, $p = 0.005^{**}$], Tone3 [$\beta = 0.418$, $t = 6.165$, $p < 0.001^{***}$] and Tone4 [$\beta = 0.348$, $t = 5.297$, $p < 0.001^{***}$], but no significant difference is shown for Tone1 [$\beta = -0.082$, $t = -1.253$, $p = 0.213$]. Besides, pairwise comparisons of Jitter indicate a larger effect size of health condition in Tone3 [$\beta = 0.418$] and Tone4 [$\beta = 0.348$] than in Tone2 [$\beta = 0.196$].

Table 2: Results of post-hoc pairwise comparisons for the interaction effect. Differences between cold and healthy condition in each tone are listed. Estimates with significant p-values are put in bold.

Parm.	*Lexical Tones (cold - healthy)*				*Statistics*	
	Tone1	*Tone2*	*Tone3*	*Tone4*	*F*	*Sig.*
HNR	0.080	-0.533	**-1.402**	**-1.200**	**190.95**	**< 0.001**
CPPs	**-0.811**	**-0.784**	**-0.841**	**-0.904**	0.694	0.556
Jitter	-0.082	**0.196**	**0.418**	**0.348**	**328.12**	**< 0.001**
Shimmer	-0.535	-0.121	0.331	0.027	2.398	0.066

4. DISCUSSION

Results of linear mixed-effects models suggest that there is a significant main effect of health condition on CPPs and Jitter. Specifically, the CPPs of cold speakers is significantly lower, compared with healthy speakers. Cold speakers also have higher Jitter values, compared with healthy ones. As a measurement of overall noise level in the speech, a lower CPP implies that more aperiodic vibrations and noisy elements are found in the voice production of cold speakers. As for Jitter, it mainly presents the stability of vibration of the vocal folds. A higher Jitter suggests that the vocal folds show more tremors and a greater instability during the phonation stage.

Degradation of voice quality is largely due to the inflammation of larynx caused by common colds, which changes the mass, length and tension of the vocal folds. Such changes could lead to a more deviant vibration. It becomes more difficult for speakers with common colds to maintain a consistent pitch or change their pitch with ease. Besides, asymmetrical changes in the shape and temporary hypofunction of vocal folds make it more difficult to form a complete closure during the contact phase. More airflows could be spilt out of the vocal folds, which accounts for an increase of noise and a greater instability of phonation in cold speakers. Other symptoms of common colds, such as coughing and hoarseness, irritate the vocal folds as well and bring more noisy portions to the sound.

Results of linear mixed-effects models suggest that a significant interaction effect between health condition and lexical tones was found for HNR and Jitter. To be specific, significant difference of HNR between cold speakers and healthy speakers was found for Tone 3 and Tone 4, but not for Tone1 and Tone2. Significant difference of Jitter was found for Tone2, Tone3 and Tone4, but not for Tone1. Significant interaction between health condition and Mandarin tones implies that lexical tones with curved pitch contours, including Tone2, Tone3 and Tone4, are more susceptible to common colds, presenting a more apparent degradation of voice quality.

Modification of fundamental frequency is achieved through changes in mass, length and tension of the vocal folds [30]. Therefore, Mandarin tones with curved pitch contours, including Tone 2 (mid-rising), Tone 3 (falling-rising) and Tone 4 (high-falling), might require extra vocal efforts and muscle controls during the process of voice production. Suggested by previous studies, there is a tendency for cold speakers to adjust their phonation strategy with a decrease of phonatory effort, considering the vocal fatigue caused by common colds [15]. Thus, curved tones which require more phonatory efforts, including Tone2, Tone3 and Tone4, show an apparently poorer voice quality compared to flat Tone1.

Besides, we notice that pairwise comparisons for Jitter show a larger effect size of health condition in Tone3 and Tone4 than in Tone2, despite they are all curved tones. It is largely due to the difference in the magnitude and rate for the change of pitch. As shown in Figure1, Tone3 and Tone4 have greater pitch ranges and steeper slopes of pitch contour, compared with Tone2. More phonatory efforts might be required for the production of Tone 3 and Tone4, to assure a timely and fast-enough change of the height and slope of the pitch. Besides, the production of Tone3 and Tone4 passes the lower part of possible pitch range of speakers, where chaotic oscillations are usually observed during lax phonation [31]. Thus, a greater degradation of voice quality could be found in Tone3 and Tone4, than in Tone2.

The study fills the gap of previous phonetic studies on common colds by exploring the influence of common colds on voice quality and its interaction with lexical tones in Mandarin speakers. The study offers a new interface between speech pathology and tone types, which could be further expanded in relevant studies on other tonal languages. Studies of the voice quality of colds speakers could help uncover mechanisms of voice production of speakers with speech disorders, and improve the performance of speech-related disease detection.

Additional investigation in future studies may be necessary to explore the physiological and articulatory factors underlying the dissimilarities of changing patterns observed in four parameters of voice quality.

5. REFERENCES

[1] V. M. Dixit and Y. Sharma, "Voice Parameter Analysis for the disease detection," *IOSR Journal of Electronics and communication Engineering,* vol. 9, no. 3, pp. 48-55, 2014.

[2] P. Mouawad, T. Dubnov, and S. Dubnov, "Robust detection of COVID-19 in cough sounds," *SN Computer Science,* vol. 2, no. 1, pp. 1-13, 2021.

[3] A. Vahedian-Azimi, A. Keramatfar, M. Asiaee, S. S. Atashi, and M. Nourbakhsh, "Do you have COVID-19? An artificial intelligence-based screening tool for COVID-19 using acoustic parameters," *The Journal of the Acoustical Society of America,* vol. 150, no. 3, pp. 1945-1953, 2021.

[4] J. Kutor, S. Balapangu, J. K. Adofo, A. A. Dellor, C. Nyakpo, and G. A. Brown, "Speech signal analysis as an alternative to spirometry in asthma diagnosis: investigating the linear and polynomial correlation coefficient," *International Journal of Speech Technology,* vol. 22, no. 3, pp. 611-620, 2019.

[5] A. Rybakovas, V. Beiša, K. Strupas, J. Kaukėnas, and G. Tamulevičius, "Inverse Filtering of Speech Signal for Detection of Vocal Fold Paralysis After Thyroidectomy," *Informatica,* vol. 29, no. 1, pp. 91-105, 2018.

[6] J. Wagner, T. Fraga-Silva, Y. Josse, D. Schiller, A. Seiderer, and E. André, "Infected Phonemes: How a Cold Impairs Speech on a Phonetic Level," in *Interspeech*, 2017.

[7] N. Cummins, A. Baird, and B. W. Schuller, "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning," *Methods,* vol. 151, pp. 41-54, 2018.

[8] R. Eccles, "Understanding the symptoms of the common cold and influenza," *The Lancet infectious diseases,* vol. 5, no. 11, pp. 718-725, 2005.

[9] M. J. Bové, S. Kansal, and C. A. Rosen, "Influenza and the vocal performer: Update on prevention and treatment," *Journal of Voice,* vol. 22, no. 3, pp. 326-332, 2008.

[10] A. S. Monto, "Epidemiology of viral respiratory infections," *The American journal of medicine,* vol. 112, no. 6, pp. 4-12, 2002.

[11] D. A. J. Tyrrell, S. Cohen, and J. E. Schilarb, "Signs and symptoms in common colds," *Epidemiology & Infection,* vol. 111, no. 1, pp. 143-156, 1993.

[12] G. G. Jackson, H. F. Dowling, I. G. Spiesman, and A. V. Boand, "Transmission of the common cold to volunteers under controlled conditions: I. The common cold as a clinical entity," *AMA archives of internal medicine,* vol. 101, no. 2, pp. 267-278, 1958.

[13] R. T. Sataloff, "Common infections and inflammations and other conditions," *Professional Voice: The Science and the Art of Clinical Care,* p. 429, 1997.

[14] P.-T. Lin, J. C. Stern, and W. J. Gould, "Risk factors and management of vocal cord hemorrhages: an experience with 44 cases," *Journal of Voice,* vol. 5, no. 1, pp. 74-77, 1991.

[15] R. G. Tull and J. C. Rutledge, "Analysis of “cold-affected”speech for inclusion in speaker recognition systems," *The Journal of the Acoustical Society of America,* vol. 99, no. 4, pp. 2549-2574, 1996.

[16] R. G. Tull, *Acoustic analysis of cold-speech: Implications for speaker recognition technology and the common cold*. Northwestern University, 1999.

[17] A. K. Suresh, M. SrinivasaRaghavanK., and P. K. Ghosh, "Phoneme State Posteriorgram Features for Speech Based Automatic Classification of Speakers in Cold and Healthy Condition," in *Interspeech*, 2017.

[18] R. R. Patel *et al.*, "Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function," *American journal of speech-language pathology,* vol. 27, no. 3, pp. 887-905, 2018.

[19] P. H. Dejonckere *et al.*, "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," *European Archives of Oto-rhino-laryngology,* vol. 258, no. 2, pp. 77-82, 2001.

[20] Y. R. Chao, (Mandarin Primer). Cambridge, MA and London, England: Harvard University Press, 1948.

[21] J. Kuang, "Production and perception of the phonation contrast in Yi," University of California, Los Angeles, 2011.

[22] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Interspeech*, 2017, vol. 2017, pp. 498-502.

[23] Y. Jadoul, B. Thompson, and B. De Boer, "Introducing parselmouth: A python interface to praat," *Journal of Phonetics,* vol. 71, pp. 1-15, 2018.

[24] P. Boersma, "Praat: doing phonetics by computer," http://www. praat. org/, 2006.

[25] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823,* 2014.

[26] R. C. Team, "R: A language and environment for statistical computing," 2013.

[27] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of memory and language,* vol. 68, no. 3, pp. 255-278, 2013.

[28] A. Kuznetsova, P. B. Brockhoff, and R. H. Christensen, "lmerTest package: tests in linear mixed effects models," *Journal of statistical software,* vol. 82, pp. 1-26, 2017.

[29] R. Lenth, H. Singmann, J. Love, P. Buerkner, and M. Herve, "Emmeans: Estimated marginal means, aka least-squares means," *R package version,* vol. 1, no. 1, p. 3, 2018.

[30] M. Asiaee, A. Vahedian-Azimi, S. S. Atashi, A. Keramatfar, and M. Nourbakhsh, "Voice quality evaluation in patients with COVID-19: An acoustic analysis," *Journal of Voice,* 2020.

[31] D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, "Interpretation of biomechanical simulations of healthy and chaotic vocal fold oscillations with empirical eigenfunctions," *The Journal of the Acoustical Society of America,* vol. 95, no. 6, pp. 3595-3604, 1994.