CEPSTRAL PEAK PROMINENCE IN NORMOPHONIC ADULTS: THE EFFECT OF GENDER AND SPEECH TASK SEGMENTAL COMPOSITION

Saoirse O'Regan & Irena Yanushevskaya

Trinity College Dublin
oregans1@tcd.ie, yanushei@tcd.ie

ABSTRACT

Cepstral Peak Prominence (CPP) is currently recommended for clinical use by speech and language therapists as an objective and robust measure of dysphonia. Its widespread use in clinic is stalled, however, by the relative conceptual complexity of CPP and by the lack of baseline normative data. This paper describes CPP values extracted from the audio recordings of normophonic adults 25-55 years old living in Ireland. Audio data was collected remotely, using mobile phones, from 42 participants (21 female). Speech tasks included sustained vowels as well as sentences with different segmental composition (vowels and approximants, nasals, voiced and voiceless obstruents). CPP values were extracted using Praat. We describe the effects of gender, speech task segmental composition and CPP extraction method (with and without voice activity detection) on the obtained CPP values. The results may serve as a normative baseline for clinical voice analysis and assessment in speech and language therapy practice.

Keywords: Cepstral Peak Prominence, normative data, gender, speech task, clinical voice analysis.

1. INTRODUCTION

Clinical voice analysis and evaluation is a multidimensional task and includes both subjective (auditory) and objective (instrumental) methods of assessment [1, 2]; it is expected that at least one objective measurement is included to support auditory assessment [3]. Cepstral Peak Prominence (CPP) is currently recommended by American Speech-Language-Hearing Association (ASHA) for use in clinic as the most robust and reliable measure of dysphonia [3]. Since CPP was first proposed as a measure of breathiness in the voice signal [4, 5], there has been growing interest in the clinical applications of CPP, e.g. [6, 7] and references therein. However, its use by clinicians is still not as widespread as the more traditional measures (jitter, shimmer and HNR). For example, a survey conducted by [8] found that while participant speech and language therapists (SLTs) working with voice consistently used acoustic analysis in voice assessment, none of them reported using CPP. The likely explanation is that CPP is conceptually more complex than commonly used acoustic measures: as stated in [9], 'CPP shares with other cepstral measures the lack of an intuitive interpretation relative to the underlying physiology of vocal fold vibration'. CPP values are influenced by recording conditions (e.g., using mobile phone for remote data collection), segmental composition of speech tasks [6, 10], the choice of software applications [11, 12], analysis settings such as window type and length [13] and the use of additional speech processing, e.g., voice activity detection (VAD) [7]. SPL and speaker f_0 are also contributing factors [13, 14]. Importantly, there is lack of comprehensive baseline normative CPP data by gender and age that would facilitate the interpretation of CPP in clinical voice assessment.

In this paper we describe CPP values extracted using Praat [15] from the audio recordings of various speech tasks (isolated vowels, connected speech) obtained from a sample of normophonic adults 25-55 years old living in Ireland. The data was collected and analysed in an effort to contribute to the growing body of information about normative CPP values which clinicians could use as a baseline when working with disordered voice clients in speech and language therapy clinics and other health care settings.

2. CEPSTRAL PEAK PROMINENCE AND ITS CLINICAL APPLICATIONS

Power cepstrum of a signal equals to the Fourier transform of the logarithm of its power spectrum [9]. Periodic signals will show energy at harmonically related frequencies in their spectrum, and their cepstrum will have a prominent peak at the time ('quefrency') corresponding to the fundamental period of the signal. Cepstral Peak Prominence (CPP) measures, in dB, the magnitude of the cepstral peak relative to a linear regression line fitted to the cepstrum to normalise the prominence of the cepstral peak to the overall amplitude of the signal [4, 5]. CPP is a measure of the periodicity of the spectrum of a speech signal rather than the periodicity of the signal per se [10, 13].

Over the past years, CPP has been extensively studied, e.g. a comprehensive review in [9] or [6] on clinical applications of CPP. It is well established that CPP is a more robust measure compared to traditional voice quality measures such as jitter, shimmer and HNR: its calculation does not require accurate pitch extraction algorithms and it can be reliably used in the analysis of both sustained vowels and connected speech even if recorded in a non-controlled environment [16, 17]. While CPP is a robust predictor of overall dysphonia severity, it may not perform well discriminating individual voice quality dimensions, such as roughness [10, 18].

Various factors may influence CPP values. Speaker gender and speech task effects have been reported in [19], e.g., significantly higher CPP values were reported in sustained vowels produced by males than by females, but the trend was opposite in sentences. Typically, CPP values in sustained vowels are higher than in connected speech, e.g. summary of studies in [6]. [20] reported higher CPP values for open vowels /ɑ/ and /æ/ than for close vowels /i/ and /u/. The impact of segmental composition of the speech task on CPP was examined by [6]. Their findings suggest that CPP values in CAPE-V sentences varied depending on the amount of voiced segments: the all-voiced sentence ‘We were away a year ago’ yielded the highest CPP and the sentence with voiceless aspirated plosives ‘Peter would keep at the peak’ the lowest. The presence of nasals has an effect of lowering CPP [10, 13, 21].

CPP values extracted from the same audio file may differ depending on the software used [12, 22], e.g., Praat [15] and Analysis of Dysphonia in Speech and Voice (ADSV, PENTAX Medical), which have different approaches to voice activity detection [6]. The use of voice activity detection presents a bit of a conundrum: If VAD is not used, the presence of voiceless segments in the speech samples selected for the analysis may artificially lower CPP. If VAD is used and the analysed voice is dysphonic with a lot of aperiodicity, the resulting CPP based on relatively small sections of data will be artificially inflated and not representative of the speaker’s voice. The solution proposed in [6] is to use all voiced sentences or to remove voiceless segments prior to the analysis.

Many clinical studies looked at correlations of CPP and auditory-perceptual evaluation of disordered speech and normal controls and at the correlations of CPP with the auditory-perceptual ratings using CAPE-V [23] or GRBAS [24], e.g. reviews in [6, 9]. CPP was found to reliably separate normal and disordered voices and to highly correlate with auditory-perceptual ratings. Several studies aimed to establish CPP cut-off values for both sustained vowels and connected speech, but the findings are varied. Cut-off CPP values in [6] are 11.46 dB (ADSV) and 14.45 dB (Praat) for the sustained /a/ vowels and 6.11 dB (ADSV) and 9.33 dB (Praat) for the Rainbow Passage [25]; they detected the presence of voice disorders with over 90% accuracy.

As mentioned earlier, there are limitations to the use of CPP in clinical settings due to the lack of comprehensive normative data that would assist clinical voice evaluation. One can infer these data from CPP values obtained from healthy controls in the clinical studies; however, there are differences in the reported findings. Studies conducted specifically to collect CPP normative data are still scarce, and the obtained results also vary, e.g., [19, 26].

In this study CPP values were extracted using Praat from the audio recordings of isolated vowels and connected speech obtained from a sample of vocally healthy English speaking adults 25-55 years old living in Ireland. The ultimate goal was to provide normative CPP values which SLTs could use as a baseline when working with disordered voice clients in speech and language therapy clinics. Of particular interest are the effects of speaker gender, speech task and CPP extraction method on CPP values.

3. MATERIAL AND METHOD

Participants The study was approved by Research Ethics Committee; informed consent was obtained from all participants prior to data collection. Forty-two vocally healthy participants (21 female) 25-55 years old living in Ireland, all English speaking, were recruited. Participants had no self-reported history of voice disorders or smoking. Participants’ age by gender distribution is shown in Fig. 1.

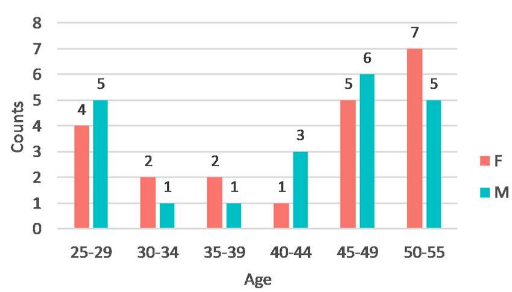

Figure 1: Recorded speakers by age and gender.

Speech task and recording procedure The speech tasks used here were those commonly used in clinical voice assessments and in studies on CPP [6, 12, 22]. The tasks included vowels [a] and [i] sustained for five seconds, CAPE-V [23] sentences with different segmental composition: ‘We were away a year ago’ (mainly vowels and vowel-like articulations), 'A man may mean many[1]' (vowels and nasals), 'How hard did he hit him?' (word-initial [h]),'Peter will keep at the peak' (voiceless aspirated stops), and the first two sentences from 'The Rainbow passage' [25]. In the discussion that follows we use shortened names for these tasks.

Participants recorded speech tasks in a quiet environment of their convenience using the default audio recording apps on their mobile phones (iPhone Voice Memos and Android Voice Recorder). They then sent their recordings to the researchers by email. To protect the identity of the participants, the audio files were anonymised. Analysis of data collected remotely, using mobile phones, provides valuable information to facilitate voice assessment via telepractice, whether during the recent pandemic or to accommodate people living in remote areas.

CPP data extraction The obtained M4A files were converted to WAV using Audacity® [27] at the default sampling frequency 48 kHz. The files were segmented in Praat so that each speech task was saved as its own WAV file. An informal auditory analysis was conducted by the authors to ensure that the audio files were of adequate quality and did not contain ambient noise or evidence of atypical phonation. There were 294 sound files in total (42 participants x 7 speech tasks [2 vowels, 4 sentences, 1 passage]).

CPP values were extracted automatically using the Praat plug-in described in [7]. The plug-in allows to extract smoothed CPP using the default Praat settings following [4] which have also been used in [6, 11, 12, 14] and other papers. The plug-in includes the method for extracting CPP with voice activity detection (VAD) performed as the first step. We used both VAD and 'traditional' noVAD extraction methods.

4. RESULTS AND DISCUSSION

The distribution of CPP values by speech task, gender and extraction method (VAD, noVAD) is shown in Fig. 2; to facilitate by-gender and by-method comparison the same data are plotted twice. Table 1 gives median, mean and 95% CI values (due to space constraints, for noVAD method only). As can be inferred from Fig. 2, CPP values get lower for both male and female speakers with the increase of the proportion of voiceless segments, particularly voiceless plosives, in the speech task. The sentence where 39% of segments are voiceless stops ('Peter will keep at the peak') showed the lowest CPP. Generally, it appears that male speakers yield lower mean CPP in most tasks (except sustained vowels) compared to female speakers in both extraction methods.

[1] Not a CAPE-V sentence.

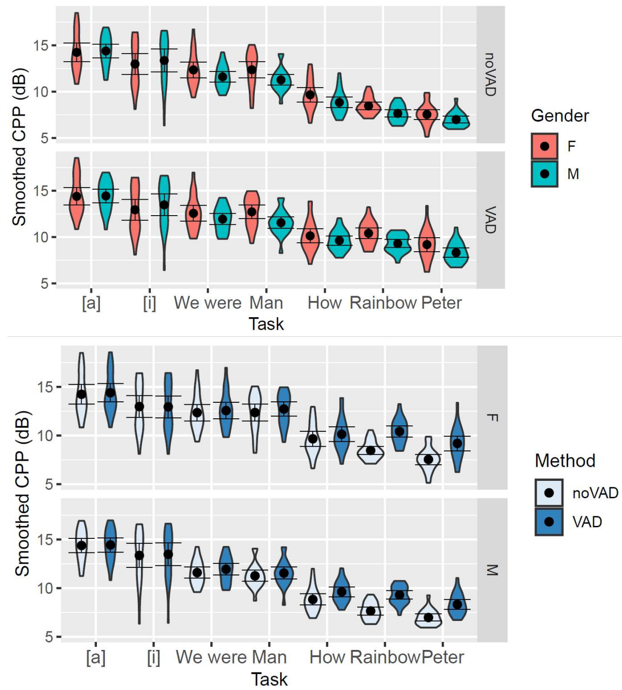

Figure 2: The distribution of CPP values by speech task, gender and extraction method. Black dots and whiskers show mean and 95% CI.

Gender	Task	CPP Median	CPP Mean	95% CI LL–UL
Female (n = 21)	[a]	13.86	14.23	13.24–15.23
	[i]	12.98	12.98	11.85–14.10
	We were	12.09	12.35	11.51–13.19
	Man	13.07	12.36	11.48–13.24
	How	9.18	9.66	8.90–10.43
	Rainbow	8.29	8.47	8.04–8.89
	Peter	7.53	7.54	7.02–8.07
Male (n = 21)	[a]	14.42	14.39	13.64–15.13
	[i]	14.47	13.36	12.13–14.59
	We were	11.46	11.60	11.03–12.17
	Man	11.15	11.27	10.70–11.84
	How	8.66	8.84	8.27–9.42
	Rainbow	7.68	7.66	7.25–8.06
	Peter	6.88	6.99	6.63–7.35

Table 1: CPP median, mean and 95% CI values. Due to space constraints, only noVAD method results are shown.

As it was of interest to establish to what extent speaker gender, the type and segmental composition of speech task and CPP extraction method have an effect on CPP values, we conducted linear mixed-effect model analyses. Our initial model included CPP values as the dependent variable, gender, speech task and extraction method as well as their interactions as the main predictors (fixed effects); random effects included by-subject random intercepts and slopes to account for speaker variability in speech tasks:

CPP~Gender*Task*Method+(1+Task|Participant). Analyses were conducted in R environment [28] using *lme4* [29] and *lmerTest* packages [30] for model fitting (using maximum likelihood method) and step-down model simplification by eliminating non-significant effects and interactions. The final model was CPP~Gender+Method+Task+Method:Task+(1+Task |Participant). *sjPlot* package [31] was used for model visualization. Post-hoc pairwise comparisons were conducted using the *emmeans* package [32] with Tukey correction for multiple comparisons. The results of the mixed model analyses are shown in Table 3; Fig. 4 shows model predicted CPP values.

Predictors	*Est.*	*CI*	*t*	*p*
(Intercept)$_{[F,\ noVAD,\ a]}$	14.81	14.17 – 15.44	**45.67**	**<0.001**
Gender$_{[M]}$	-0.98	-1.49 – -0.48	**-3.80**	**<0.001**
Method$_{[VAD]}$	0.10	-0.09 – 0.30	1.04	0.300
Task$_{[How]}$	-5.06	-5.66 – -4.46	**-16.54**	**<0.001**
Task$_{[i]}$	-1.14	-1.68 – -0.60	**-4.14**	**<0.001**
Task$_{[Man]}$	-2.49	-3.08 – -1.91	**-8.41**	**<0.001**
Task$_{[Peter]}$	-7.04	-7.61 – -6.48	**-24.35**	**<0.001**
Task$_{[Rainbow]}$	-6.25	-6.75 – -5.75	**-24.62**	**<0.001**
Task$_{[We_were]}$	-2.34	-2.89 – -1.78	**-8.20**	**<0.001**
Method$_{[VAD]}$× Task$_{[How]}$	0.52	0.25 – 0.80	**3.69**	**<0.001**
Method$_{[VAD]}$× Task$_{[i]}$	-0.06	-0.34 – 0.22	-0.42	0.669
Method$_{[VAD]}$× Task$_{[Man]}$	0.22	-0.06 – 0.50	1.53	0.128
Method$_{[VAD]}$× Task$_{[Peter]}$	1.39	1.11 – 1.67	**9.78**	**<0.001**
Method$_{[VAD]}$× Task$_{[Rainbow]}$	1.69	1.41 – 1.97	**11.91**	**<0.001**
Method$_{[VAD]}$× Task$_{[We_were]}$	0.17	-0.11 – 0.45	1.21	0.225
ICC		0.92		
N $_{Participant}$		42		
Observations		588		
Marginal R^2/Conditional R^2		0.651/0.973		

Table 2: Results of the mixed effect model analyses of the CPP data: coefficients and 95% CI (fixed effects only).

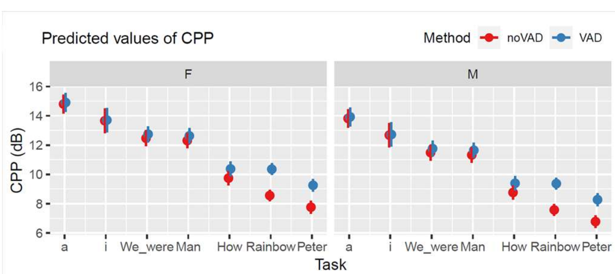

Figure 4: Model predicted CPP values in different speech tasks by gender and extraction method.

The effects of gender, speech task and extraction method (marginal R^2) explain 65% of the data variance; combined fixed and random effects account for about 97% of the variance (conditional R^2). While the general trend is overall similar for male and female speakers, the results suggest significant effect of gender ($t = -3.80$, $p < 0.001$), with CPP values for male speakers being about 1 dB lower than those of female speakers. Unsurprisingly, our results suggest that CPP values are affected by Task. Similar to [20], CPP values of the [i] vowel are significantly lower than those of the [a] vowel for both male and female speakers. CPP values of ‘We were’ and ‘Man’ are significantly lower still, and the values get progressively (and significantly) lower as the number of obstruents and voiceless segments in the utterances increases, with the ‘Peter’ sentence showing the lowest CPP. However, we found significant Method:Task interaction (Fig. 4.). Extraction method has no significant effect on CPP in sustained vowels, but for other speech tasks CPP$_{[VAD]}$ values were significantly higher than CPP$_{[noVAD]}$, the effect gets larger as the number of the voiceless segments in the speech tasks increases.

5. CONCLUSIONS

The paper reported the analysis of audio data (sustained vowels and connected speech) collected with mobile phones from normophonic English speaking adults aged 25-55 living in Ireland. The ultimate goal was to contribute to the information on normative CPP values to facilitate the use and interpretation of CPP by SLTs in clinical voice assessment. The CPP normative values obtained in our study are similar to the ones reported in [6, 33] and the findings generally support earlier studies. Our data suggest significantly higher CPP for female speakers. Provided that the analysis settings are as described in [7], the choice of speech task emerged as another important factor for CPP analysis. CPP extraction method does not seem to matter for the analysis of sustained vowels; in all other speech tasks CPP$_{[VAD]}$ is higher than CPP$_{[noVAD]}$, similar to [6]. This is an important consideration in a busy clinic where pre-processing of the data (i.e. exclusion of unvoiced segments) is not desirable nor feasible unless fully automated.

The findings need to be confirmed on a larger sample and with parallel recording done in optimal conditions to compare mobile phone data with. A potential follow-up study would involve clinicians working with disordered voice clients who would submit mobile phone samples over the course of therapy to be evaluated using CPP. Developing clear guidelines for clinicians, incorporating normophonic data by age, gender, speech task, recording conditions and extraction method in intuitive and easy-to-use tools for clinicians will facilitate a wider use of CPP by speech and language therapists in clinical voice assessment.

6. ACKNOWLEDGEMENTS

The second author acknowledges the support of the Faculty of Arts, Humanities and Social Sciences, Trinity College Dublin (Research Fellows Award).

7. REFERENCES

[1] 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. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)," (in eng), *European Archives of Oto-Rhino-Laryngology,* vol. 258, no. 2, pp. 77-82, 2001.

[2] P. H. Dejonckere, "Assessment of voice and respiratory function," in *Surgery of Larynx and Trachea*, M. Remacle and H. E. Eckel Eds. Berlin Heidelberg: Springer-Verlag, 2010, pp. 11-26.

[3] 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.

[4] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech and Hearing Research,* vol. 37, no. August, pp. 769-778, 1994.

[5] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech," *Journal of Speech and Hearing Research,* vol. 39, pp. 311-321, 1996.

[6] O. Murton, R. Hillman, and D. Mehta, "Cepstral Peak Prominence values for clinical voice evaluation," *Am. J. Speech Lang. Pathol.,* vol. 29, no. 3, pp. 1596-1607, 2020.

[7] E. S. Heller Murray, A. Chao, and L. Colletti, "A practical guide to calculating Cepstral Peak Prominence in Praat," (in eng), *J. Voice,* 2022.

[8] S. McAlister and I. Yanushevskaya, "Voice assessment practices of speech and language therapists in Ireland," *Clinical Linguistics & Phonetics,* pp. 1-25, 2019.

[9] R. Fraile and J. I. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis," *Biomedical Signal Processing and Control,* vol. 14, pp. 42-54, 2014.

[10] C. A. Ferrer Riesgo, M. S. de Bodt, Y. Maryn, P. Van de Heyning, and M. E. Hernández-Díaz Huici, "Properties of the cepstral peak prominence and its usefulness in vocal quality measurements," in *MAVEBA*, Florence, Italy, 2007, pp. 93-96.

[11] Y. Maryn and D. Weenink, "Objective dysphonia measures in the program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index," *Journal of Voice,* vol. 29, no. 1, pp. 35-43, 2015.

[12] C. R. Watts, S. N. Awan, and Y. Maryn, "A comparison of cepstral peak prominence measures from two acoustic analysis programs," (in eng), *J Voice,* vol. 31, no. 3, pp. 387.e1-387.e10, 2017.

[13] C. A. Ferrer Riesgo and E. Nöth, "What makes the Cepstral Peak Prominence different to other acoustic correlates of vocal quality?," (in eng), *J Voice,* vol. 34, no. 5, pp. 806.e1-806.e6, 2020.

[14] M. Brockmann-Bauser, J. H. Van Stan, M. Carvalho Sampaio, J. E. Bohlender, R. E. Hillman, and D. D. Mehta, "Effects of vocal intensity and fundamental frequency on Cepstral Peak Prominence in patients with voice disorders and vocally healthy controls," *Journal of Voice,* vol. 35, no. 3, pp. 411-417, 2021.

[15] *Praat: doing phonetics by computer [computer program]*. (2023). Accessed: 10/02/2023. [Online]. Available: http://www.praat.org/

[16] Y. Maryn, N. Roy, M. De Bodt, P. Van Cauwenberge, and P. Corthals, "Acoustic measurement of overall voice quality: a meta analysis," *Journal of the Acoustical Society of America,* vol. 126, no. 5, pp. 2619-2634, 2009.

[17] B. Barsties v. Latoszek, Y. Maryn, E. Gerrits, and M. De Bodt, "A Meta-analysis: Acoustic measurement of roughness and breathiness," *Journal of Speech, Language & Hearing Research,* Article vol. 61, no. 2, pp. 298-323, 2018.

[18] S. Y. Lowell, R. H. Colton, R. T. Kelley, and S. A. Mizia, "Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech," *Journal of Voice,* vol. 27, no. 4, pp. 393-400, 2013.

[19] C. Batthyany *et al.*, "A case of specificity: How does the Acoustic Voice Quality Index perform in normophonic subjects?," *Applied Sciences,* vol. 9, no. 12, p. 2527, 2019.

[20] S. N. Awan, A. Giovinco, and J. Owens, "Effects of vocal intensity and vowel type on cepstral analysis of voice," (in eng), *J Voice,* vol. 26, no. 5, pp. 670.e15-20, 2012.

[21] C. Madill, D. D. Nguyen, K. Yick-Ning Cham, D. Novakovic, and P. McCabe, "The impact of nasalance on Cepstral Peak Prominence and Harmonics-to-Noise Ratio," *The Laryngoscope,* vol. 129, no. 8, pp. E299-E304, 2019.

[22] C. Sauder, M. Bretl, and T. Eadie, "Predicting voice disorder status from smoothed measures of Cepstral Peak Prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV)," *Journal of Voice,* vol. 31, no. 5, pp. 557-566, 2017.

[23] G. B. Kempster, B. R. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, "Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol," (in eng), *Am J Speech Lang Pathol,* vol. 18, no. 2, pp. 124-32, 2009.

[24] M. Hirano, *Clinical Examination of Voice*. New York: Springer Verlag, 1981.

[25] G. Fairbanks, *Voice and articulation drillbook*. New York: Harper & Row, 1960.

[26] K. V. Phadke, A.-M. Laukkanen, I. Ilomäki, E. Kankare, A. Geneid, and J. G. Švec, "Cepstral and perceptual investigations in female teachers with functionally healthy voice," *Journal of Voice,* vol. 34, no. 3, pp. 485.e33-485.e43, 2020.

[27] *Audacity® software: free audio editor and recorder [computer application]*. (1999-2021). [Online]. Available: https://audacityteam.org/

[28] *R: A language and environment for statistical computing*. (2021). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.r-project.org/

[29] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[30] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software,* vol. 82, no. 13, pp. 1-26, 2017.

[31] *sjPlot: Data Visualization for Statistics in Social Science*. (2021). Accessed: 27/04/2023. [Online]. Available: https://CRAN.R-project.org/package=sjPlot

[32] *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.5*. (2023). . Accessed: 27/04/2023. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[33] J. Delgado-Hernández, N. León-Gómez, and A. Jiménez-Álvarez, "Diagnostic accuracy of the Smoothed Cepstral Peak Prominence (CPPS) in the detection of dysphonia in the Spanish language," *Loquens,* vol. 6, no. 1, p. e058, 2019.