Cepstral Peak Prominence Smoothed distribution as discriminator of vocal health in sustained vowel

Cepstral Peak Prominence Smoothed distribution as discriminator of vocal health in sustained vowel

A. Castellana, A. Carullo, S. Corbellini, A. Astolfi
Dipartimento di Elettronica e Telecomunicazioni e
Dipartimento di Energia
Politecnico di Torino, Torino, Italy
Email: antonella.castellana@polito.it

M. Spadola Bisetti, J. Colombini
Dipartimento di Scienze Chirurgiche
Università degli Studi di Torino, Torino, Italy
Email: massimo.spadolabisetti@unito.it

Abstract—**This paper focuses on Cepstral Peak Prominence Smoothed (CPPS) as a possible indicator of vocal health status, considering individual CPPS distribution and its descriptive statistics. 31 voluntary patients and 22 control subjects performed the same protocol, which includes the simultaneous acquisition of three repetitions of the sustained vowel /a/ with a microphone in air and a contact sensor, the perceptual assessment of voice and the videolaringoscopy examination. The best logistic regression models have been applied and preliminary results showed that the fifth percentile and the standard deviation of CPPS distributions are the best parameters that discriminate healthy and unhealthy voice for the microphone in air and the contact sensor, respectively. The Area Under Curve (AUC) revealed the diagnostic precision of the selected CPPS parameters: AUC of 0.96 and 0.83 have been found for the microphone in air and the contact sensor, showing strong to moderate discrimination power, respectively. The repeatability of the selected CPPS parameters has been also estimated. For each selected CPPS parameter, the Monte Carlo method has been implemented in order to evaluate the uncertainty of the threshold, which was identified by means of the Receiver Operating Curve analysis.**

Keywords — Dysphonia; Cepstral analysis; Sustained vowel; Repeatability; Monte Carlo method.

I. INTRODUCTION

Objective assessment of voice overcomes the subjectivity due to the interpretation of symptoms and medical standards. One of the objective tools commonly employed is the voice acoustic analysis, which is used to assess voice disorders thanks to its non-invasiveness, low cost and ease of application [1]. It provides a numerical output that is relatively easy to communicate to all stakeholders, such as voice clinicians, patients, third-party payers, and physicians [2] and allows tracking of vocal behavior, proving to be appealing for dysphonia prevention, diagnosis, and dysphonia treatment.

Many researches have studied acoustic analysis algorithms and methods to obtain an objective analysis of dysphonia and its severity (see Buder for an overview [3]). Time-based parameters, such as *jitter* and *shimmer*, have been the first investigated ones. They depend on accurately identifying cycle boundaries, i.e. where a cycle of vocal-fold vibration begins and ends, so they become unreliable with highly perturbed signals [4]-[5]. Furthermore, such traditional perturbation parameters are valid only for sustained vowels produced with steady pitch and loudness, since any purposeful changes will be read as increases in vocal perturbation [6]. To overcome the limitations of cycle boundary detection, current practice are considering spectral- and cepstral-based measures, which can be applied also to continuous speech that is able to represent everyday speaking patterns [7]. Among them, cepstral analysis has been considered as the most promises measure of dysphonia severity. According to the definition given by Hillenbrand and Houde [8], the cepstrum is a log power spectrum of a log power spectrum: the first power spectrum represents the frequency distribution of the signal energy, while the second spectrum shows how regular the harmonics peaks in the spectrum are. Two cepstral parameters have been defined, namely the Cepstral Peak Prominence (CPP) and its smoothed version (CPPS). CPP is a measure (in dB) of the amplitude of the cepstral peak, normalized for overall signal amplitude by means of linear regression line calculated relating quefrency to cepstral magnitude [9]. CPPS derived from two smoothing processes before calculating the cepstral peak prominence [8]. Maryn et al. [10] highlighted the relevance of CPPS: they performed a meta-analysis on correlation coefficients between acoustic measurements and perceptual evaluation of voice quality, stating that CPPS satisfied the meta-analytic criteria in sustained vowels as well as in continuous speech. Other studies have demonstrated the correlation of CPPS with perceptual ratings of overall grade of dysphonia and different types of voice quality [11]-[16]. Brinca et al. [17] assessed that CPPS measures were significantly different between dysphonic and control group in the vowel /a/, but in the existing literature there is a lack of investigations on diagnostic precision of CPPS. Such analysis has been done for multi-parametric indexes, e.g. the Acoustic Voice Quality Index (AVQI), which is a multivariate construct with CPPS and other four acoustic parameters [18]. All the above-mentioned works used cepstrum software packages to calculate CPPS from signals acquired with microphones in air. *Praat* [19], *SpeechTool* [20] and the Analysis of Dysphonia in Speech and Voice module [21] of *Multi-Speech* from KayPENTAX (Montvale, NJ) are the most popular packages. These programs only provide the mean of CPPS values and in some cases the standard deviation.

Recently, in-clinic short-term measurements have been replaced by in-field long-term monitorings, which allow for the characterization of the vocal behavior with distributional parameters [22]. Proper devices for such vocal monitoring have

Cepstral Peak Prominence Smoothed distribution as discriminator of vocal health in sustained vowel

voice disorders who are correctly identified as positive. *Specificity*, that is the true negative rate, is the percentage of people with healthy normal voice who are correctly classified as negative. The authors avoid the usual choice of selecting where the sensitivity and specificity curves cross, since they selected the cutoff giving priority to the sensitivity that corresponds to a greater true positive rate. All the analysis related to the logistic regression model has been performed using the statistical program RStudio (Version 0.99.489).

3) Intra-speaker variability

With the purpose of investigating the repeatability of the descriptive statistics for CPPS distribution that have been included in the empirical fitted models, CPPS distributions have been calculated in the three repetitions of the sustained vowel /a/ for each subject. Forty subjects repeated correctly the second task described in paragraph II.B, while wearing both the headworn microphone and the ECM.

4) Monte Carlo method

The Monte Carlo method has been implemented for the uncertainty estimation of the threshold values, which have been obtained for each empirical fitted model by means of the ROC analysis. The Maximum Likelihood Estimation has been implemented in MATLAB in order to determine the best fitting distribution for the distributions of CPPS parameters that were included in the models, both in healthy and unhealthy voices. In this analysis, the values of the CPPS parameters in the three repetitions of the vowel for each subject have been considered. Then, 1000 trials have been repeated by randomly sampling 50 values from each fitted distribution. For each Monte Carlo trial the best threshold value of the logistic model has been determined, setting the equality between the sensitivity and the specificity that were obtained from the ROC analysis.

III. Results

A. Microphone in air

The *p*-values of the Two-tailed Mann-Whitney U-test of the lists of descriptive statistics related to the two groups of subjects were lower than 0.05, which means null hypotheses rejected, except for skewness and kurtosis. These outcomes reveal that CPPS distributions are significantly different in location, with an average value of 15.4 dB and 18.4 dB for $CPPS_{mean}$ in patients and controls, respectively, and in variance, with an average value of 2.0 dB and 1.3 dB for $CPPS_{std}$ in patients and controls, respectively.

We assumed the presence/absence of dysphonia as dependent variable and the best logistic regression model between healthy and unhealthy voice includes $CPPS_{5prc}$ as independent variable. The best empirical fitted model is defined in terms of probability by the exponential expression:

$$P(Unhealthy) = \frac{e^{(28.1 - 1.87 \cdot CPPS_{5prc})}}{1 + e^{(28.1 - 1.87 \cdot CPPS_{5prc})}} \quad (1)$$

where *P(Unhealthy*) is the probability of having unhealthy voice, which ranges from zero to one. The negative coefficient of $CPPS_{5prc}$ shows that the probability to have unhealthy voice decreases as the $CPPS_{5prc}$ increases. The empirical model has a Mc Fadden's R^2 equal to 0.63 and an AUC of 0.96, thus highlighting that there is a clear separation between patients and controls. Fig. 2 shows the fitted values obtained for each subject and most of patients are in the upper part of the graph, where the probability of having unhealthy voice is near to one, while most of controls have lower scores, near to zero. We also calculated the best classification threshold of *P(Unhealthy*) = 0.48, that corresponds to 15.1 dB in terms of $CPPS_{5prc}$, with a sensitivity equal to 0.94 and a specificity of 0.86.

Fig. 3 shows the average values and the relative experimental standard deviations of $CPPS_{5prc}$ in the three repetitions of the vowel /a/ acquired with the headworn microphone for each subject. The average of the standard deviations of the $CPPS_{5prc}$ is equal to 1.0 dB for the patient group and 0.4 dB for the control group.

The best-fitted distributions of the parameter $CPPS_{5prc}$ for unhealthy and healthy voices acquired with the microphone in air are bimodal and normal, respectively. Their probability density functions have been used for the implementation of the Monte Carlo method. Fig. 4 shows the distribution of threshold-values, which has been obtained from 1000 trials. It

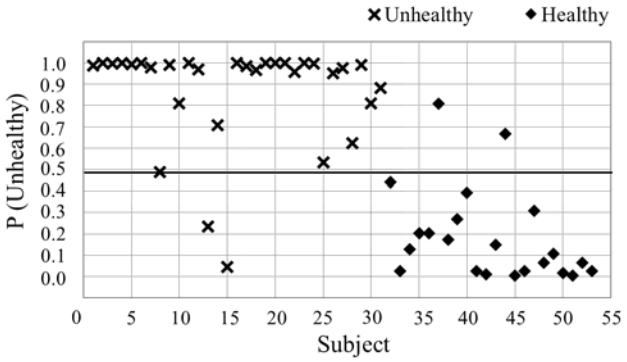

Fig. 2. Fitted values of the best logistic regression model, in terms of probability of having unhealthy voice, for vocalizations acquired with the headworn microphone Mipro MU-55HN. Cross points indicate the patient group; diamond points represent the control group. The bold line indicates the threshold value (0.48), which best separates patients and control subjects.

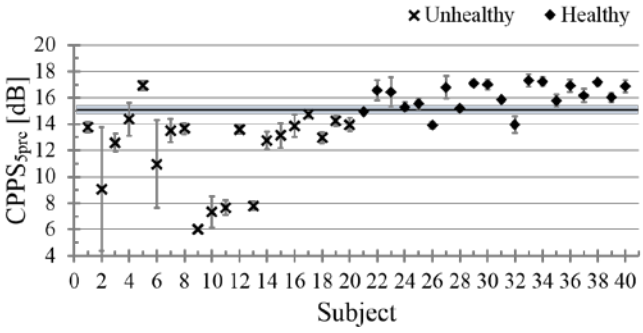

Fig. 3. Averaged values of $CPPS_{5prc}$ in the three repetitions of the vowel for each subject, acquired with the headworn microphone Mipro MU-55HN. Cross points indicate the patient group; diamond points represent the control group. Bars indicate the experimental standard deviation for each subject. The bold line indicates the threshold value (15.1 dB) and the gray area corresponds to its confidence interval obtained with a coverage factor $k = 2$.

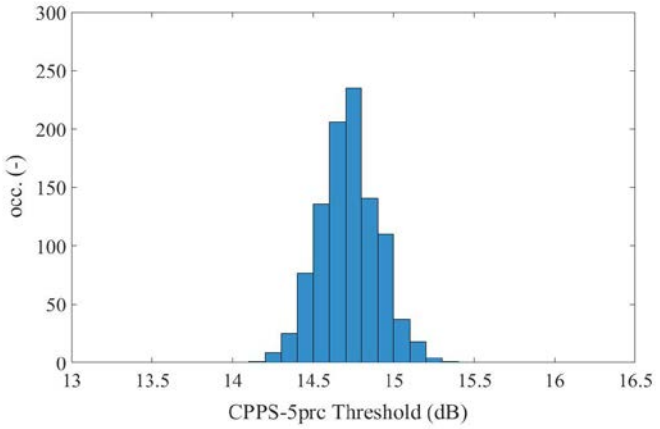

Fig. 4. Distribution of threshold-values between healthy and unhealthy voices for $CPPS_{5prc}$. It has been obtained from 1000 trials of the Monte Carlo method.

has a standard deviation of 0.18 dB that represents the standard uncertainty estimation of the $CPPS_{5prc}$ threshold value between healthy and unhealthy voices. The gray area around the $CPPS_{5prc}$ threshold in the Fig. 3 represents the confidence interval obtained with a coverage factor $k = 2$.

B. Contact microphone

The Two-tailed Mann-Whitney U-test stated that the lists of descriptive statistics for CPPS distributions related to the groups of patients and control subjects, who was recorded with ECM, resulted to be significantly different in $CPPS_{mean}$, $CPPS_{std}$, $CPPS_{range}$ and $CPPS_{5prctile}$ (p-values < 0.05). CPPS distributions were different in location, e.g. the average $CPPS_{mean}$ was equal to 18.2 dB for patients and 19.6 dB for controls, and in variance, e.g. the average $CPPS_{std}$ was equal to 1.8 dB and 1.0 dB for patients and controls, respectively.

The best empirical fitted logistic model for voice samples acquired with ECM includes $CPPS_{std}$ as independent variable and it is expressed as:

$$P(Unhealthy) = \frac{e^{(-5.31+4.60 \cdot CPPS_{std})}}{1+e^{(-5.31+4.60 \cdot CPPS_{std})}} \quad (2)$$

where *P(Unhealthy)* is the probability of having unhealthy voice, which ranges from zero to one. The positive coefficient of $CPPS_{std}$ shows that the probability to have unhealthy voice increases as $CPPS_{std}$ increases. The empirical model has a moderate discrimination power, with a Mc Fadden's R^2 equal to 0.31 and an AUC of 0.83. Therefore, Fig. 5 shows that the fitted values of the two groups are not clearly separated. The best classification threshold is *P(Unhealthy)* = 0.43, that corresponds to 1.1 dB in terms of $CPPS_{std}$, with a sensitivity of 0.79 and a specificity of 0.59.

Fig. 6 shows the average values and the relative experimental standard deviations of $CPPS_{std}$ in the three repetitions of the vowel /a/ acquired with the ECM for each subject. The average of the standard deviations of the $CPPS_{std}$ is equal to 0.3 dB for the patient group and 0.2 dB for the control group.

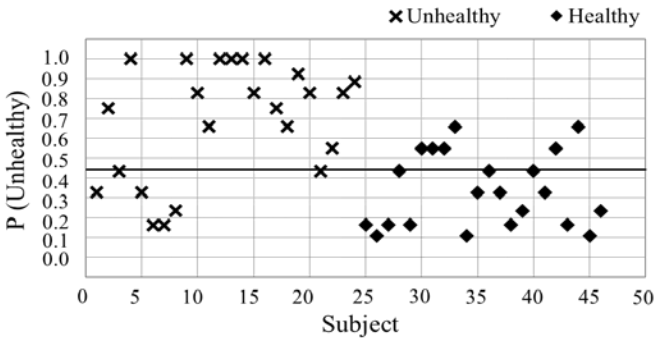

Fig. 5. Fitted values of the best logistic regression model, in terms of probability of having unhealthy voice, for samples acquired with the contact microphone ECM AE38. Cross points indicate the patient group; diamond points represent the control group. The bold line indicates the selected threshold value, that is 0.43, which best separates patients and control subjects.

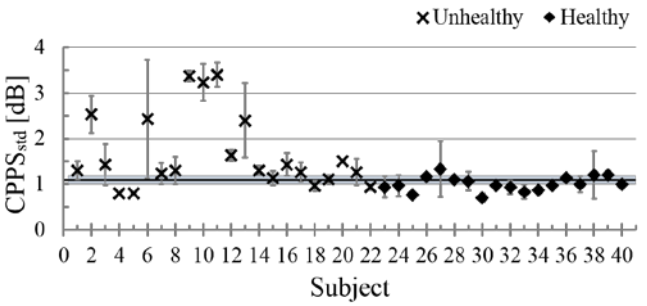

Fig. 6. Averaged values of $CPPS_{std}$ in the three repetitions of the vowel for each subject, acquired with the contact microphone ECM AE38. Cross points indicate the patient group; diamond points represent the control group. Bars indicate the experimental standard deviation for each subject. The bold line indicates the threshold value (1.1 dB) and the gray area corresponds to its confidence interval obtained with a coverage factor $k = 2$.

The best-fitted distributions of the parameter $CPPS_{std}$ for unhealthy and healthy voices acquired with ECM are bimodal and lognormal, respectively. Their probability density functions have been used for the implementation of the Monte Carlo method. The distribution of threshold-values, which has been obtained from 1000 trials, has a standard deviation of 0.04 dB that represents the standard uncertainty estimation of the threshold value of $CPPS_{std}$ between healthy and unhealthy voices. The gray area around the $CPPS_{std}$ threshold in the Fig. 6 represents the confidence interval obtained with a coverage factor $k = 2$.

IV. CONCLUSION

In this work, the descriptive statistics from individual distribution of Cepstral Peak Prominence Smoothed (CPPS) have been investigated as possible indicators of vocal health status. CPPS has been computed for sustained vowels /a/ acquired with a microphone in air and a contact sensor (ECM) from a patient group and a control group. The fifth percentile ($CPPS_{5prc}$) of individual CPPS distributions resulted the best descriptive statistic that discriminates healthy and unhealthy

voices for the vocal samples acquired with the microphone in air, showing a strong discrimination power (AUC = 0.96). Its threshold value was equal to 15.1 dB, with lower values indicating unhealthy status of voice. The standard deviation ($CPPS_{std}$) was instead the best CPPS parameter that separates the two groups for the vocal samples acquired with ECM. It has a moderate discrimination power, with AUC of 0.83. Differently from the results by Mehta et al. [30], the proposed method is able to classify healthy and unhealthy voice from both the microphone in air and ECM. Further investigations are needed to identify the reasons of the lower discrimination power found for ECM, which could be due to the different frequency behavior of the two microphones (flatness and/or bandwidth), as highlighted in [26].

The intra-speaker variability of the two CPPS parameters was larger in the patients group than in the control one, as expected; its respective values were 1.0 dB and 0.4 dB for $CPPS_{5prc}$ and 0.2 dB and 0.3 dB for $CPPS_{std}$.

Preliminary results showed that the standard uncertainty of the threshold values between healthy and unhealthy voices is negligible for both the $CPPS_{5prc}$ and $CPPS_{std}$, which is equal of 0.18 dB and 0.04 dB, respectively.

Future works will extend the investigation to the descriptive statistics from individual distribution of CPPS in continuous speech acquired with the two types of sensor.

REFERENCES

[1] V. Parsa, D.G. Jamieson, "Acoustic discrimination of pathological voice: sustained vowels versus continuous speech," J. Speech Lang. Hear. Res., vol. 44, pp. 327–339, 2001.

[2] L.G. Portney, M.P. Watkins, Foundations of Clinical Research: Applications to Practice, 2 Ed., Upper Saddle River, Prentice-Hall, 2000.

[3] E.H. Buder, Acoustic analysis of voice quality: a tabulation of algorithms 1902–1990. In: R.D. Kent, M.J. Ball, eds. Voice Quality Measurement, San Diego, CA: Singular Publishing Group, 2000, pp.119–244.

[4] C.R. Rabinov, J. Kreiman, B. Gerratt and S. Bielamowicz, "Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter," Journal of Speech and Hearing Research, vol. 38, pp. 26–32, 1995.

[5] S. Bielamowicz, J. Kreiman, B. R. Gerratt, M. S. Dauer and G. S. Berke, "Comparison of voice analysis systems for perturbation measurement," Journal of Speech and Hearing Research, vol. 39, 126–134, 1996.

[6] Y. Zhang and J.J. Jiang, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," J Voice, vol. 22, 1–9, 2008.

[7] S.Y. Lowell, R.H. Colton, R.T. Kelley, Y.C. Hahn, "Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker", J Voice, vol. 25, pp.223–232, 2011.

[8] J. Hillenbrand, R.A. Houde, "Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech," J. Speech Hear Res., vol. 39(2), pp. 311-21, 1996.

[9] J. Hillenbrand, R.A. Cleveland, R.L. Erickson, "Acoustic correlates of breathy vocal quality," J Speech Hear Res, vol. 37, pp. 769–778,1994.

[10] Y. Maryn, N. Roy, M. De Bodt, P. Van Cauwenberge, P. Corthals, "Acoustic measurement of overall voice quality: a meta-analysis," J. Acoust. Soc. Am., vol. 126, pp. 2619–2634, 2009.

[11] V. Wolfe, D. Martin, "Acoustic correlates of dysphonia: type and severity," J. Commun. Disord., vol. 30, pp. 403–415, 1997.

[12] Y.D. Heman-Ackah, D.D. Michael, G.S. Goding, "The relationship between cepstral peak prominence and selected parameters of dysphonia", J Voice,vol.16, pp. 20–27, 2002.

[13] Y.D. Heman-Ackah, R.J. Heuer, D.D. Michael, R. Ostrowski, M. Horman, M.M. Baroody, J. Hillenbrand, R.T. Sataloff, "Cepstral peak prominence: a more reliable measure of dysphonia," Ann. Otol. Rhinol. Laryngol., vol.112, pp. 324–333, 2003.

[14] B. Halberstam, "Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels," J. Otorhinol. Relat. Spec., vol. 66, pp.70–73, 2004.

[15] C. Moers, B. Möbius, F. Rosanowski, E. Nöth, U. Eysholdt, T. Haderlein, "Vowel- and text-based cepstral analysis of chronic hoarseness," J. Voice, vol.26, pp.416–424, 2012.

[16] R.A. Samlan, B.H. Story, K. Bunton, "Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling," J. Speech Lang. Hear. Res., vol.56, pp. 1209–1223, 2013.

[17] L.F. Brinca, P.F. Batista, A.I. Tavares, I.C. Goncalves, M.L. Moreno, "Use of Cepstral Analyses for Differentiating Normal From Dysphonic Voices: A Comparative Study of Connected Speech Versus Sustained Vowel in European Portuguese Female Speakers," J Voice, vol. 28 (3), pp. 282-286, 2014.

[18] Y. Maryn, M. De Bodt, N. Roy, "The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders," J. Commun. Disord.,vol. 43, pp.161–174, 2010.

[19] P. Boersma and D. Weenink, Institute of Phonetic Sciences, University of Amsterdam, The Netherlands, http:// www.praat.org/. (last view: 31/10/2016).

[20] J.M. Hillenbrand, James M. Hillenbrand Homepage, [Online] Available at: http://homepages.wmich.edu/~hillenbr/ (last view: 31/10/2016).

[21] S.N. Awan, N. Roy, M.E. Jette, G.S. Meltzner, R.E. Hillman, "Quantifying dysphonia severity using a sepctral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V," Clin. Linguist. Phon., vol.24, pp.742–758, 2010.

[22] M. Ghassemi, J.H. Van Stan, D.D. Mehta, M. Zanartu, H.A. Cheyne, R.E. Hillman, J.V. Guttag, "Learning to Detect Vocal Hyperfunction From Ambulatory Neck-Surface Acceleration Features: Initial Results for Vocal Fold Nodules," IEEE Tr. on Biomedical Engineering, vol. 61(6), pp. 1668-1675, 2014.

[23] P.S. Popolo, J.G. Svec, and I.R. Titze, "Adaptation of a Pocket PC for Use as a Wearable Voice Dosimeter," Journal of Speech Language and Hearing Research, vol. 48, pp. 780-791, 2005.

[24] VoxLog portable voice meter [online] available: http://www.sonvox.com/index.html. (last view: 31/10/2016).

[25] H.A. Cheyne, H.M. Hanson, R.P. Genereux, K.N. Stevens and R.E. Hillman, "Development and Testing of a Portable Vocal Accumulator," J. of Speech Lang. and Hear Research, vol. 46, pp. 1457-1467, 2003.

[26] A. Carullo, A. Vallan, and A. Astolfi, "Design Issues for a Portable Vocal Analyzer", IEEE Tr. on IM, vol. 62(5), pp. 1084-1093, 2013.

[27] A. Carullo, A. Vallan, and A. Astolfi, "A Low-Cost Platform for Voice Monitoring", in Proceedings of I2MTC Conference, Minneapolis, MN (USA), May 6-9 2013.

[28] A. Carullo, A. Vallan, A. Astolfi, G.E. Puglisi, L. Pavese, "Validation of calibration procedures and uncertainty estimation of contact-microphone based vocal analyzers," Measurement, vol. 74, pp. 130-142, 2015.

[29] D.D. Mehta, M. Zaartu, S.W. Feng, H.A Cheyne, R.E. Hillman, "Mobile Voice Health Monitoring Using a Wearable Accelerometer Sensor and a Smartphone Platform," IEEE Tr. on Biomedical Engineering, vol. 59(11), pp. 3090-3096, 2012.

[30] D.D. Mehta, J.H. Van Stan, R.E.Hillman, "Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer," IEEE/ACM Trans Audio Speech Lang Process., vol. 24(4), pp. 659-668, 2016.

[31] G. de Krom, "A cepstrum-based technique for determing a harmonics to noise ratio in speech signals," J. S. Hear. Res., vol. 36, pp. 254-266, 1993.

[32] P. Gramming, "Vocal loudness and frequency capabilities of the voice", J. Voice, vol. 5, pp. 144-157, 1991.

[33] R.F. Coleman, "Sources of variation in phonetograms," J. Voice, vol. 7, pp. 1-14, 1993.

[34] J.D. Gibbons and S. Chakraborti, "Nonparametric Statistical Inference," Taylor & Francis, 2003, pp. 215-223.

[35] D. Hosmer, S. Lemeshow, R. Sturdivant, "Applied Logistic Regression," third edition, Wiley, 2013.