H.Ü. Sağlık Bilimleri Fakültesi Dergisi
Cilt:6, Sayı:3, 2019
doi: 10.21020/husbfd.645267

Original Research

Is the Cepstral Analysis Sensitive Enough to Detect Untrained/Trained Resonant Voice in Healthy Subjects? A Preliminary Study

Fatma Esen Aydınlı[1], Esra Özcebe[2], Jackie Gartner-Schmidt[3]

Submission Date: 11.11.2019 **Acceptance Date:** 25.11.2019 **Pub.Date:** 30.12.2019

Abstract

Objectives: In training resonant voice, clinicians rely on listening and feeling sensations that are subjective in nature. In the present study, the aim was to find out if cepstral analysis was sensitive enough to distinguish a resonant voice from an unresonant voice in healthy subjects. It was suggested that the immediate effect of humming exercise could be shown by accompanying higher Cepstral Peak Promince (CPP) and lower CPP standard deviation (CPP SD). **Materials and Methods:** Study included 20 vocally healty subjects aged between 20-30 years. All the participants were trained Humming exercise by teaching airflow technique. The recording protocol was repeated before and after resonant voice training. The recordings were gathered for the three samples of a) Humming b) A sentence weighted with nasal phonemes and c) A sentence weighted with all- voiced phonemes. Cepstral analysis recordings were gathered by the 'Kay Elemetrics Group CSL Model 4300B' equipment via the Analysis of Dysphonia in Speech and Voice software. **Results:** In both sentences, CPP values were found numerically higher and CPP SD values were found to be numerically lower after training. The difference was found statistically significant for the post-training value of voiced- weighted sentence (p=0.004). **Conclusion:** The higher numerical CPP values and lower CPP SD values for both sentences after resonant voice training indicated that cepstral measures were sensitive to detect trained/untrained resonant voice. However, these preliminary findings should be investigated in a larger sample size in order to achieve more reliable data.

Key words: *Resonant voice, humming, voice*

[1] **Fatma Esen Aydınlı (Sorumlu Yazar).** University of Hacettepe, Faculty of Health Sciences, Department of Speech and Language Therapy, Ankara, Turkey, e-mail: fesen04@gmail.com. 0312 305 4387.
[2] **Esra Özcebe.** University of Hacettepe, Faculty of Health Sciences, Department of Speech and Language Therapy, Ankara, Turkey, e-mail: eozcebe@gmail.com
[3] **Jackie Gartner-Schmidt**. University of Pittsburgh, School of Medicine, Department of Otolaryngology, Pittsburgh, United States of America. e-mail: jgs@pitt.edu

H.Ü. Sağlık Bilimleri Fakültesi Dergisi
Cilt:6, Sayı:3, 2019
doi: 10.21020/husbfd.645267

Introduction

Resonant voice exercises are one of the common methods used in voice therapy programs of dysphonic patients. Resonant voice exercises are also used for a healthy population in the music and drama education (Zaliouk, 1963). In terms of production; the resonant voice has been shown to be produced with lower phonation threshold pressure and lesser vocal fold impact stress (Peterson, Verdolini-Marston K, Barkmeier and Hoffman, 1994). Audio-perceptually, 'resonant voice' is perceived with ease of phonation and accompanied vibration sensation on the facial area (Yiu, Lo, Barrett, 2017) (Verdolini, 2000). The vibration areas may include the alveolar ridge, lips, hard palate, soft palate, teeth, and facial bones. Many voice therapy approaches such as Lessac-Madsen Resonant Voice Therapy (Verdolini, 2000), Resonant Therapy (Stemple, 2000), humming exercise (Yiu and Ho, 2002) and Y-Buzz (Barrichelo and Behlau, 2007) therapy approaches are based on the resonant voice production.

In resonant voice-based therapy programs; the focus is on the feelings of vibration sensation on the face as well as the perception of clearer voice quality and ease of phonation. In this therapy method, sensorial information is processed via 'Listening and Feeling' which are subjective tasks (Yiu, Lo, Barrett, 2017). Many factors related to the subjects may affect this process such as concentration, motivation, imagination, auditory attention and personal experiences with the voice as well as the clinician factors (Yiu, Lo, Barrett, 2017) (Verdolini, 2000). As a result, the training of the resonant voice may be a very challenging task, especially during the first therapy session.

Humming is a preferred resonant voice therapy technics in both professional voice users as well as dysphonic patients. Studies have been conducted in both populations to identify the immediate effects of humming. Electroglottography (EGG) (Ogawa et al., 2014) M, fiberoptic laryngoscopy (Ogawa et al. (2013) videolaryngoscopy (VLS) (Verdolini, Druker, Palmer and Samawi, 1998) and facial accelometre (FA) (Chen, Estella, Edwin, 2014) were used in these studies. Ogawa et al. (2013) found that false vocal fold compensation and antero-posterior compression were decreased during the humming in the fiberoptic laryngoscopy. In another study done via EGG (Ogawa et al., 2014) it has been shown that there had been a decreased closure phase ratio and reduction in the variability of perturbation parameter values. Verdolini et al. (1998) used VLS and EGG during the breathy, effortful and resonant voice production in actors and singers. It was concluded that the VLS image in a resonant voice was noticeably different from that of a breathy and effortful voice. The same researchers have found that in EGG, closure phase ratios were similar for the breathy and resonant voice; so that this

doi: 10.21020/husbfd.645267

indicated that EGG could not be used as a distinctive characteristic. Chen et al. (2014) used humming exercise in healthy subjects (n=12) and measured the vibration quantity on their faces via a Piezoelectric accelerometer for the first time. Their study proved that vibrations around the nasal bridge and upper lip were increased during the humming.

As a result, many objective methods have been used in the literature to search for the immediate effects of humming. However, when viewed from the point that if clinicians and patients may get objective feedback on their resonant voice productions; instrumental methods seemed to be ideal to use with this purpose. For example, VLS and fiberoptic laryngoscopy may be used by clinicians to receive feedback, however, the procedures have some difficulties such as requirements of anesthetics and experienced ear nose throat professionals and the discomfort of the patient (Hirano and Bless, 1993). EGG is a relatively easier method that can be used to detect resonant voice, however, in the study of Verdolini et al. (1998), it was concluded that EGG could not be used in the distinction of characteristics of the resonant voice. In the study of Chen et al. (2014) piezoelectric accelerometre to detect associated vibrations in the resonant voice. This instrument is rarely used, and cannot be found in most voice clinics. Besides, the study results could not reveal a cause-and-effect relationship between facial bone vibrations and auditory-perceptual and/or acoustic characteristics of the resonant voice.

In training resonant voice, clinicians rely on listening and feeling sensations that are subjective in nature. Unexperienced speech and language pathologists (SLPies) may find it difficult to discriminate resonant voice from an unresonant voice. An objective instrument that differs resonant voice from unresonant voice may be used as a feedback tool for both the patients and clinicians. In this study, the aim was to find out if cepstral analysis was sensitive enough to distinguish a resonant voice from an unresonant voice in healthy subjects. Since a highly periodic voice signal is considered to contain a well- defined fundemantal frequency (F_0) and harmonic structure, and these values are associated with a high and prominent cepstral peak prominence (CPP) (Heman-Ackah et al., 2003). It was suggested that the immediate effect of humming exercise could be shown by accompanying higher CPP and lower CPP standard deviation (CPP SD) in healthy subjects.

Materials and Methods

This study has been approved by the Hacettepe University Ethical Committee- (G0 14/603-04-26.11.2014).

doi: 10.21020/husbfd.645267

Participants

The study's participants were aged between 20-30 years. Participants were selected based on these criterias; Not having any dysphonia complaints at any time, not having any previous surgeries that might affect their vocal performance, no smoking status, not having any pulmonary or neurological diseases, not using any medication that could affect vocal performance during the day of evaluation, not having an upper respiratory infection on the day of evaluation, and having a Voice Handicap Score under 11 points. Also, the participants were included if two SLPies agreed that their voice was acceptable according to the auditory-perceptual evaluation results. The participants were excluded from the study if they had complaints about their voice, if they were professional voice users such as actors or singers, or if they had any hearing loss in their medical history.

Procedures

VHI-10 completion

All of the participants were asked to fill in the VHI-10 (Kiliç et al. , 2008) to eliminate any possibility of them having a dysphonic voice. Only the subjects whose VHI-10 score were below the 11 points were included.

Auditory-Perceptual Evaluation of Voice

All the participants' voices were evaluated audio- perceptually by two SLPies to confirm that their voices were healthy. The participants were asked to read the sentences in the Turkish version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V/Turkish) (Özcebe et al., 2018). All the data were recorded and saved using Computerizes Speech Labaratory (CSL) (CSL Model 4500 equipment, Kay Elemetrics Group, Lincoln Park, NJ, USA) equipment.

Cepstral Analysis Recordings

Cepstral analysis recordings were gathered by the 'Kay Elemetrics Group CSL Model 4300B' equipment via the Analysis of Dysphonia in Speech and Voice (ADSV) software (Lincoln Park, NJ, USA). Recordings were gathered in a silent room. The Micromic C520 headset microphone with a distance of 5 cm and a 45- degree angle to the mouth was used (Titze, 1995). Voice recordings were made using the sampling rate of 25.000 Hertz (Hz) (Awan, Roy, Jetté, Meltzner and Hillman, 2010). The recording protocol was repeated before and after resonant voice training. The recordings were gathered for the three samples of a) humming b) A sentence weighted with nasal phonemes: /mʌjmuŋ mʌmʌdʌŋ mɛmŋuŋdu/ (Maymun

doi: 10.21020/husbfd.645267

mamadan memnundu) and c) A sentence weighted with all- voiced phonemes sentence: /vʌlɪ vʌlɪzɪnɪ vɛrdɪ/ (Vali valizini verdi).

Training

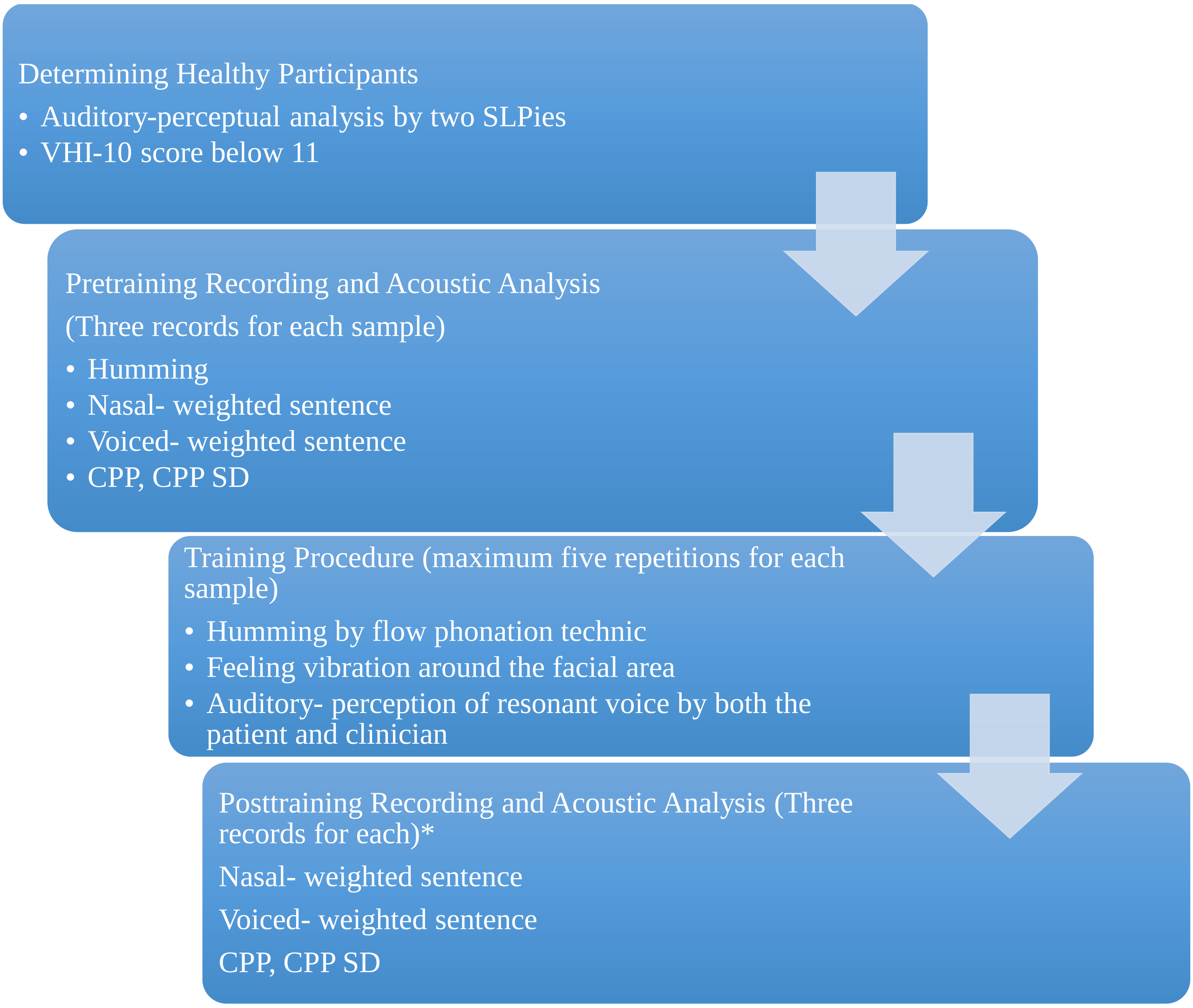
Each participant was thought the humming exercise via the flow phonation method (Behrman and Haskell, 2008) by an experienced SLP. In this exercise, a strip of tissue held between the index and third finger approximately two inches from the nose. The tissue is expected to move with the airflow; this is repeated until consistent results are achieved. Than a ‘hummm’ sound is produced with forward oral resonance. (Behrman and Haskell, 2008). The participants were told to provide the SLPies with two variables as they trained in resonant voice production: a) a vivid description of the subject’s senses of vibration around the facial area b) auditory- perception of resonant voice by the SLP. For each participant, a maximum of 5 repetitions were obtained and if one of the variables was not obtained, the participant was excluded from the study. This training procedure took about 15-30 minutes for each participant. Since the loudness level is directly proportional with the CPP it may affect other spectral measures (Awan, Giovinco and Owens, 2012). During the recordings, the amplitude was checked by the Sound Level Meter (SLM) (RadioShack Digital SLM) at real-time and was kept in the range of habitual amplitude decibel (dB): ±5(70-78 dB).

Determining Appropriate Recordings

By listening to trained/untrained sample couples, the samples that were concluded to be described as a ‘resonant voice’ by two SLPies and which showed pitch values in the range between ±15 Hz were included for further cepstral analysis. In the cepstral analysis, pretraining-posttraining CPP and CPP SD values of nasal- weighted and voiced- weighted sentences were compared.

H.Ü. Sağlık Bilimleri Fakültesi Dergisi
Cilt:6, Sayı:3, 2019
doi: 10.21020/husbfd.645267

Figure 1: Representation of materials and methods.



*Only appropriate samples in terms of F_0 (F_0 values in the range between ±15 Hz and auditory-perceptually 'resonant voice' samples were included for analysis, mean values averaged of three records).

Statistical Analysis

Statistical analyses were done by using the IBM SPSS Statistics 21 programme. Since the pre and post training Cepstral analysis numeric variables were found to be nonparametric, the 'Wilcoxon two- sample paired signed rank test' was used for the difference between the pre/post training values. As for the significance level, p value was accepted < 0.05. All of the descriptive statistics were reported using the mean and median values, as well as the interquartile range (IQR).

H.Ü. Sağlık Bilimleri Fakültesi Dergisi
Cilt:6, Sayı:3, 2019
doi: 10.21020/husbfd.645267

Results

Participants

A total of 20 healthy participants aged between 20-30 years were used in the results. The subjects were composed of 10 males and 10 females.

Findings

Descriptive statistics revealed median CPP value as 8.352 in pre- training evaluation for the nasal- weighted sentence where as this value was increased to 8.861 after training. Although a numerically higher CPP value was obtained, the difference was not significant (p=0.255). The CPP SD value which was equal to 2.712 decreased to 2.191 after training. Similar to the CPP, p value belonging the pre- post training CPP SD difference was not found statistically significant (p=1.790)

Table 1: Descriptive statistics and comparison of cepstral analysis parameteres for nasal- weighted sentence.

	PreTraining Values			**PostTraining Values**			
Variable	**Median**	**Mean**	**Range**	**Median**	**Mean**	**Range**	**p**
CPP(dB)	8.352	8.415	3.03	8.861	8.653	3.79	0.255
CPPSD(dB)	2.712	2.680	2.24	2.191	2.380	1.79	0.179

Abbreviations: CPP: Cepstral Peak Prominence, CPP SD: Cepstral Peak Prominence Standard deviation.

Descriptive statistics and comparison of cepstral anaylsis parameters for voiced- weighted sentence were shown in Table 2. According to the Table 2, it is seen that median value of the pre-training CPP value for the voiced- weighted sentence was increased to 8.816 from 8.202 after training. This difference was also found statistically significant (p=0.004). When the pre- post training values of CPP SD were compared; it was found that pretraining value decreased after training. That is the CPP SD value which was equal to 3.199 had decreased to the 3.013 dB however this difference was not significant (p=0.184).

Table 2: Descriptive statistics and comparison of cepstral anaylsis parameteres for voiced- weighted sentence.

	PreTraining Values			**PostTraining Values**			
Variable	**Median**	**Mean**	**Range**	**Median**	**Mean**	**Range**	**p**
CPP (dB)	8.202	8.03	3.81	8.816	8.67	3.79	0.004*
CPPSD (dB)	3.199	3.31	1.55	3.013	3.21	1.65	0.184
CPPF0 (Hz)	231	223	136	241	231	127	0.006*

Abbreviations: CPP: Cepstral Peak Prominence, CPP SD: Cepstral Peak Prominence Standard deviation. * indicates that p < 0.05.

doi: 10.21020/husbfd.645267

Discussion

Cepstrum measures are frequency-based measures that are gathered by applying twice the Fourier analysis on the logarithm power spectrum. Measures of the cepstrum are proved to indicate dysphonia severity in both sustained vowel and continuous speech contexts (Awan et al., 2010), (Awan et al., 2012). Many studies reported that increased vocal abnormality is associated with a decrease in the amplitude of the CPP (i.e., lower harmonic energy) and an increase in value of CPP SD. In other words, a highly periodic voice signal is expected to contain a well- defined F_0 and harmonic structure, and this corresponds with a high and prominent CPP (Heman-Ackah et al. ,2003) (Awan et al., 2012).

In the present study, it was aimed to search if CPP and CPP SD were sensitive enough to detect trained/untrained resonant voice in healthy subjects. For this purpose; two samples including nasal- weighted and all- voiced sentences were used. First, when the CPP findings were examined; it was seen that CPP values after training were found to be numerically higher for the two sentences. These results support the hypothesis of resonant voice being associated with a higher level of CPP values. It is thought that a highly periodic signal of the resonant voice caused higher CPP amplitudes. However it was also found that these increments were only significant only for the CPP value of voiced- weighted sample. It is thought that the differences between the results of different sample types may arise from the different spectral contents of two samples. It was expected to have a higher CPP value for the nasal- weighted sentence because research indicated that vibration sensation on the facial area was highest for the /m/ phoneme (Chen et al., 2014). However the spectral content of /m/ phoneme may have induced a relatively lesser CPP amplitude compared to the voiced- weighted sentence. In cepstral measures, there had been different values were reported between different speech samples (Sauder, Bretl and Eadie, 2017). To conclude, it was thought that there had been a tendency of higher CPP values associating with the resonant voice however, it must be validated by including a higher number of participants.

According to Awan and Roy (2006), variability in the CPP SD was increased when the dysphonia severity was higher. In the present study, all the statistics regarding to the CPP SD values were found to be lowered after the resonant voice training; however, these findings were not found statistically significant. Two reasons may be the possible explanation for these findings: first; the CPP parameter is known to be a stronger predictor of voice quality (Awan et al., 2010) rather than CPP SD. Similarly, CPP SD may not be a sensitive predictor for showing

doi: 10.21020/husbfd.645267

a resonant voice. Also, it is thought that a larger sample size is needed to get a reliable comment on this variable.

In addition, the authors supposed that more prominent findings on cepstral measures would be revealed if dysphonic voice samples had been included since the irregularity changing of vocal folds would be more prominent comparing to the healthy voice. However, this study design would also have been some limitations at controlling the F0 and amplitude of the resonant voice sample. Resonant voice is known to have a higher amplitude and includes harmonics at higher frequencies. It is also known that cepstral measures are depended upon the pitch and loudness levels (Awan et al., 2012) (Sauder et al., 2017). So that in the present study, a very strict pitch range (range between ±15 Hz) and loudness criteria (dB) :±5 (70-78 dB) implemented through the values reported in the study of Awan et al. (2012). It is believed that these criterias were the other factors that made it difficult to reveal difference in cepstral measures after training.

In the interpretation of the present study's, it must be emphasized that the training period for humming was relatively short. The training period which is 15-30 minutes duration found to be insufficient for some subjects (Ogawa et al, 2013) and those subjects were not included in the study. In future studies, the cepstral analysis values of the resonant voice may be searched after longer training periods.

Conclusion

The higher numerical CPP values and lower CPP SD values for both sentences after resonant voice training indicated that cepstral measures were sensitive to detect trained/untrained resonant voice. In addition, present findings showed that voiced- weighted sentence was more sensitive to detect resonant voice by cepstral analysis compared to the nasal-weighted sentence. However, these preliminary findings should be investigated in a larger sample size in order to achieve more reliable data.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Disclosure Statement

The authors have no conflicts of interest to declare.

doi: 10.21020/husbfd.645267

References

Awan, S.N., Roy, N. (2006). Toward the development of an objective index of dysphonia severity: A four-factor model. *Clinical linguistics and phonetics*, 20(1), 35–49.

Awan, S. N., Roy, N., Jetté, M. E., Meltzner, G. S., Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical linguistics and phonetics*, *24*(9), 742-758.

Awan, S. N., Giovinco, A., Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of voice*, *26*(5), 670-e15.

Barrichelo, V.M, Behlau, M. (2007). Perceptual identification and acoustic measures of the resonant voice based on "Lessac's Y-Buzz" a preliminary study with actors. *Journal of voice*, 21(1), 46-53.

Chen, F.C., Estella, M.A., Yiu P.M., Edwin, M.L. (2014). Facial bone vibration in resonant voice production. *Journal of voice*, 28(5), 596-602.

Gartner-Schmidt, Jackie (2008). Flow phonation. In: Behrman, A., Haskell, J. (Eds). The Practice of Voice. Exercises for Voice Therapy (pp.24-25) San Diego: Plural Publishing Inc.

Heman-Ackah, Y.D., Michael, D.D., Baroody, M.M., Ostrowski, R., Hillenbrand, J., Heuer J.R. et al. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Annals of otology rhinology laryngology*, 112 (4), 324–333.

Hirano, M., Bless, D.M. (Ed.). (1993). *Videostroboscopic Examination of the Larynx*. San Diego: Singular Publishing.

Ogawa, M., Hosokawa, K., Yoshida, M., Yoshii, T., Shiromoto,O.,Inohara,H.(2013). Immediate effectiveness of humming on the supraglottic compression in subjects with muscle tension dysphonia. *Folia phoniatrica at logopedica*, 65(3), 123-128.

Ogawa, M., Hosokawa, K., Yoshida, M., Iwahashi, T., Hashimoto, M., Inohara, H. (2014). Immediate effects of humming on computed electroglottographic parameters in patients with muscle tension dysphonia. *Journal of voice*, *28*(6), 733-741.

Peterson, K.L., Verdolini-Marston, K., Barkmeier, J.M., Hoffman, H. (1994). Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. *Annals of Otology Rhinology Laryngology*, *103*(5), 335-346.

Sauder, C., Bretl, M., Eadie, T. (2017). 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*, *31*(5), 557-566.

Stemple, J.C. (2000). *Resonant Therapy*. In: Stemple, J.C., Glaze, L.E., Klaben, B.G. (eds): The Diagnostic Voice Evaluation. In: *Clinical voice pathology: Theory and management* (pp. 332-393). San Diego: Singular Publishing Group.

Titze, I.R. (1995) *Workshop on Acoustic Voice Analysis. Summary Statement*, Denver, CO: National Center for Voice and Speech.

Özcebe, E., Aydinli, F.E., Tiğrak, T. K., İncebay, Ö., Yilmaz, T. (2018) Reliability and validity of the Turkish version of the Consensus auditory-perceptual evaluation of voice (CAPE-V). *Journal of voice*, 33(3), 382.e1-382.e10.

Verdolini, K., Druker, D.G., Palmer, PM, Samawi H. (1998). Laryngeal adduction in resonant voice. *Journal of voice*,*12*(3), 315-327.

Verdolini, K. (2000). *Resonant Voice Therapy*. In: Stemple J (ed.). Voice therapy: Clinical studies (pp.45–46). San Diego: Singular Publishing Group.

Vlot, C., Ogawa, M., Hosokawa, K., Iwahashi, T., Kato, C., Inohara, H. (2017). Investigation of the immediate effects of humming on vocal fold vibration irregularity using electroglottography and high-speed laryngoscopy in patients with organic voice disorders. *Journal of voice*, 31(1), 48-56.

Yiu, E.M., Ho, E.Y. (2002) Short-term effect of humming on vocal quality. *Asia Pacific Journal of Speech Languange and Hearing*, 7(3), 123-137.

Yiu, E.M.L., Lo, M.C., Barrett, E.A. (2017). A systematic review of resonant voice therapy. *International journal of speech-langauage pathology*, 19(1), 17-29.

Zaliouk, A. (1963). The tactile approach in voice placement. *Folia Phoniatrica at Logopedia*, 15 (2-3), 147-154.