Research Article

Smoothened cepstral peak prominence in choral singers, trained singers and non-singers

JAIISH(2014)
Vol 33 pp. 18-23

Madhu Sudharshan Reddy B. [a], Gopi Kishore Pebbili[a] and Vijaitha V Soonan[a]

Affiliations

[a]All India Institute of Speech and Hearing, Mysore

Corresponding Author

Madhu Sudharshan Reddy, B.
All India Institute of Speech and Hearing, Mysore.
madhusudarshan@rocketmail.com

Key Words

Speech tool programme
Smoothened Cepstral peak prominence (Smoothened CPP)
Choral singers
Trained singers

Abstract

The aim of the study was to understand the vocal quality of voice and effect of speech tasks on Smoothened Cepstral peak prominence (Smoothened CPP) across singers. Three groups of participants comprising of 16 choral singers, 13 trained singers and 16 non singers were considered. Speech-Tool program developed by Hillenbrand (1994) was used to analyze the Smoothened CPP. Mixed ANOVA was carried out to verify the statistical significance of difference across the groups and speech tasks. Both Choral singers and trained singers obtained higher smoothened cepstral peak prominence than non-singers. Within singers, trained singers obtained higher smoothened CPP values than Choral singers; however the difference was not statistically significant. With respect to speech tasks, smoothened CPP was found to be higher for sustained phonation than Reading and this difference was found to be statistically significant. Hence, Smoothened CPP was able to differentiate singers and non-singers irrespective of the speech tasks used.

Introduction

A normal voice signal is periodical in nature with a slight cycle- to-cycle variability in both frequency and amplitude. Acoustic measures are developed to measure the voice parameters using digital signal processing (DSP) techniques. These DSP techniques are simple and non invasive in nature. Several algorithms have been developed to calculate the periodicity in both normal and abnormal (dysphonic) voices. One such reliable algorithmic measure is cepstral based estimation. Cepstral based measures gives approximate calculations of noise level in the logarithmic spectrum for speech. It relay on the peak to average measure rather than the accurate calculation of fundamental frequency. The Cepstrum (Noll, 1964) is described as a Fourier transform of the power spectrum of a sound wave. It represents the area to which the spectral harmonics (dominant harmonics accompanying with the fundamental frequency) are individualized and emerge out of the back ground noise level. To generate a cepstrum, spectrum is created from an acoustic signal by the Fourier transformation. This result in the transformation of a signal from one domain (time) to the other domain (frequency) and thus, the intensity of individual frequency is displayed in the spectrum. Another Fourier transformation step of the resulting spectrum produces the cepstrum. Here, the signal is again transformed from the one domain (frequency) to the other domain (quefrency, which equals 1/frequency). This represents the better harmonicity of a signal.

In a periodic signal different spectrums are generated. The first power spectrum indicates energy at harmonically related frequencies, and the second spectrum (i.e., the spectrum of the spectrum) display the regularity of the harmonic peaks. The time ("quefrency") at the cepstral peak represents the fundamental period of a signal and a prominent peak in a spectrum corresponds to a well-defined harmonic structure. Two of such cepstral peak measures include cepstral peak prominence (CPP) and smoothened cepstral peak prominence (smoothened CPP). CPP measure is the difference in amplitude (in dB) between the cepstral peak and the corresponding value on the regression line that is directly below the peak (i.e., the predicted cepstral magnitude for the quefrency at the cepstral peak). The general method that is accepted, involves fitting a linear regression line connecting quefrency to cepstral magnitude. The line is figured between 1 millisecond to the maximum quefrency.

Singing is a unique form of art which frequently requires years of training to attain the utmost form of performance. In a culture of various societies and life events music occupies a major position and among that group singing is a very common form of music in different cultures (Stacy, Brittain & Kerr, 2002). Singing training involves systematic practice in terms of matching the scales, notes, and learning to maintain the rhythm. Although singing involves several years of formal training, some individuals learnt to sing themselves through imitation or trial and error. These are called untrained or amateur singers. Choral singers are one such organized group of *singers* who usually perform vocal music in church choir. These singers usually do not undergo any formal vocal training. Choral singers vary acoustically in style of singing from other group of singers, majorly in terms of spectral features, sound levels and phonation frequency because of the constant matching of their pitch, intensity and timber along with the group (Ternstrom, 1991). While singing, suboptimal choral singing habits like blending (softly/loudly) or straitening of the tone or sometimes beyond the pitch range can be observed in choral singers (Kirsh, Leer, Phero, Xie & Khosla, 2013). Earlier studies on choral singing found that singer's formants occupy less energy regions while singing. (Rossing, sundberg & Ternstro, 1986, 1987; Letowski, Zimak & Lupinowa, 1988). Choral singers might have poor knowledge about the vocal mechanism because of their insufficient formal vocal training.

Trained Singers on the other hand are assumed to be proficient than non-singers in terms of their vocal range, vocal stability, vocal efficiency, control of tempo and rhythm. They also have better control over pitch and loudness by independently vary frequency and intensity. Singers have greater vocal proficiency than the non-singers (Leonard, Ringel, Horii & Daniloff 1988). Murry (1990) reported that singers are more accurate in matching the frequencies and have less variation than the non-singers. Trained singers have high spectrum envelop peak called "Singer's formant", which is close to 2.8-3.2 kHz and situated between third and fourth formants of vowel sounds (Sundberg, 1974 & Troup, 1981). In terms of formants, F3 and F4 are nearer in professional singing (trained voices) than speaking (untrained voices) (Fant, 1970).Trained and untrained singers differs physiologically in terms of laryngeal and vocal tract functioning. On physiological analyses of different parameters revealed that trained singers performed better due to the effect of training than innate physiologic gift (Brown, Hunt & Williams, 1988). While singing, trained singers have better control over respiratory, articulatory and laryngeal dynamics than untrained singers. But no difference was observed while speaking (Shipp & Izdebski, 1975).

Several studies has been carried out on vocal quality of singers and concluded that speech and professional singing differs in terms of vocal quality (Sundberg, 1970). In physiological point of view, some investigations stated that the position of larynx (vertical) differentiate singing voices from normal voice (Sundberg, 1973, 1975, 1977, 1979),

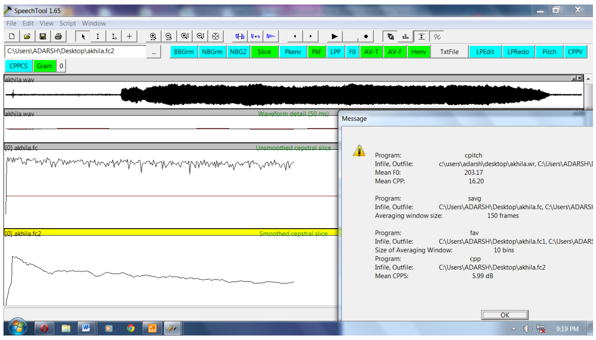

Figure 1: Illustration of the smoothening method used for smoothened CPP (sustained phonation). The top panel (a) shows the original spectrum of a signal (b) the middle panel shows unsmoothened cepstrum (c) bottom panel shows smoothened cepstrum across time and quefrency.

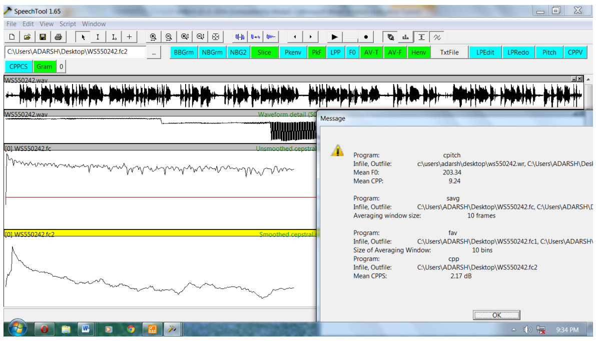

Figure 2: Illustration of the smoothening method used for smoothened CPP (Reading task). The top panel (a) shows the original spectrum of a signal (b) the middle panel shows unsmoothened cepstrum, the (c) bottom panel shows smoothened cepstrum across time and quefrency.

where as others pointed out that position of larynx/height is not responsible to distinguish the voices (Wang, 1983). Acoustic measures like jitter, shimmer, and noise-to harmonic ratio (NHR) might not depict the richness of vocal quality in singer's voice (Burns, 1986). On acoustic analysis of the singing and speaking voices in singing students, Lundy, Roy, Casiano, Xue & Evans (2000)tried to evaluated the singing power ratio (SPR) during singing and speaking voices. Recorded samples were analyzed using computerized speech lab (CSL). Results indicated no statistical difference between singing and spoken samples. The authors also did not find any influence of gender or years of training on SPR. Jannetts and Lowit (2014) investigated the validation of cepstral measure like CPP and smoothened CPP over traditional acoustic measures in hypokinetic and ataxic voices. All the samples acoustically analyzed using Multi dimensional voice program and computerized speech lab. They concluded that CPP and smoothened CPP are best predictors of overall dysphonia. Across speech tasks sustained phonation got higher mean values (CPP-13.1, smoothend CPP-5.62) than continuous speech (CPP-12.37, smoothend CPP-4.73). They concluded that CPP, smoothened CPP are better quantitative measures of vocal quality. Similarly Balasubramanium, Bhat, Fahim and Raju (2010) found CPP measures the overall dysphonia in unilatereal adductor vocal fold palsy.

Letowski, Zimak and Lupinowa (1988) studied the choral mode phenomenon in untrained and trained singers. They conclude that individuality plays a major role in trained singers to achieve choral blend using damping technique. This is due to the fact that they try to dampen the vocal quality whereas untrained singers likely to use intense vocal quality in choral context. Among the acoustic measures of voice, cepstral based acoustic measures are reported to be more robust, highly correlated with the perceptual severity of dysphonia (Awan & Roy, 2005) and are sensitive in discriminating normal and pathological voices (Balasubramanium, et al., 2010). Earlier studies evidently showed that singers would develop singer's formant which enhances the overall quality of the voice. Among the cepstral based measures, smoothened cepstral peak prominence (Smoothened CPP) is one such parameter which measures richness of harmonics and quality of voice. These methods are robust enough and have huge clinical applicability to differentiate the vocal periodicity in both singers and non-singers. In this context, the present study assumes that the Smoothened CPP will be able to reflect the differences in harmonicity between classically trained, choral and non-singers. Also it is not clear regarding the influence of speech tasks (vowel versus reading task) on smoothened CPP. Hence, the present study is taken up to document the cepstral based voice characteristics in trained, choral and non-singers. A secondary aim of the present study was to investigate the effect of speech task on the Smoothened CPP.

Method

Participants: Forty five participants were recruited in the study and were divided in to three groups. First group consisted of 16 choral singers in the age range of 17 to 30 years with the mean age of 25 years. Second group consisted of 13 trained singers in the age range of 20-30 years with the mean age of 26 years and third group were 16 non-singers in the age range of 20 to 30 years with the mean age of 23.Choral singers with singing experience of more than 3 years in choir, who had not undergone any formal training in singing, and with no complaints of any upper respiratory tract infections on the day of data collection were recruited for the study. Trained singers had trained singing experience of 5-10 years. All the trained singers had normal speech and language skills and had no complaints of voice related issues or difficulty in hearing. None of the participants had any kind of upper respiratory tract infections at the time of recording. It was also ensure that none of the non-singers had history of vocal misuse, respiratory infection, smoking or exposure to toxic chemicals.

Stimuli and Data Collection: The participants were instructed to sit comfortably in a chair with their back straight and were instructed to phonate /a/ vowel, and to read a first four sentences of the Rainbow Passage. All the voice samples of sustained phonation and Reading were recorded in a quiet room using Olympus digital voice recorder with 44 kHz sampling rate and with a constant mouth to microphone distance maintained at 15cm. Speech tool programme (version 1.65) developed by Hillenbrand, Cleveland and Erickson (1994) was used to analyze the smoothened cepstral peak prominence (smoothened CPP) for both sustained phonation and Reading.

Data Analyses: Recorded samples were subjected to further analysis using speech tool program. Averaging window size was set to 150 frames and the size of the averaging window was 10 bins. The CPP measure for a signal was averaged over all analysis frames. To get the smoothened CPP, the individual cepstra were averaged in a specified number of frames and then peak prominence was calculated. In the present study, there was no attempt was made to set window size as the settings were saved automatically and the measurement of smoothened CPP was fully automatic on speech tool program.

Statistical Analyses: Mixed ANOVA was carried out using statistical packages for social sciences (SPSS, version 17.0) to verify the significant difference, if any, between the groups (Trained singers, choral singers and non-singers) and with in groups (Sustained phonation and reading task) and their interaction. Since the interaction effect was seen between two speech tasks in each group, paired sample t-test was performed. Further, MANOVA was done to check the difference in mean values of smoothened CPP across three groups (Trained singers, choral singers and non-singers). Duncans post-hoc analysis was performed to verify whether three groups are different with respect to speech tasks.

Results and Discussion

Smoothened Cepstral Peak Prominence Among Choral Singers, Trained Sngers and Non-singers: Trained singers exhibited a mean smoothened CPP of 7.11 ($\pm$1.67) for sustained phonation and 3.81 ($\pm$2.28) for reading task.

ht!

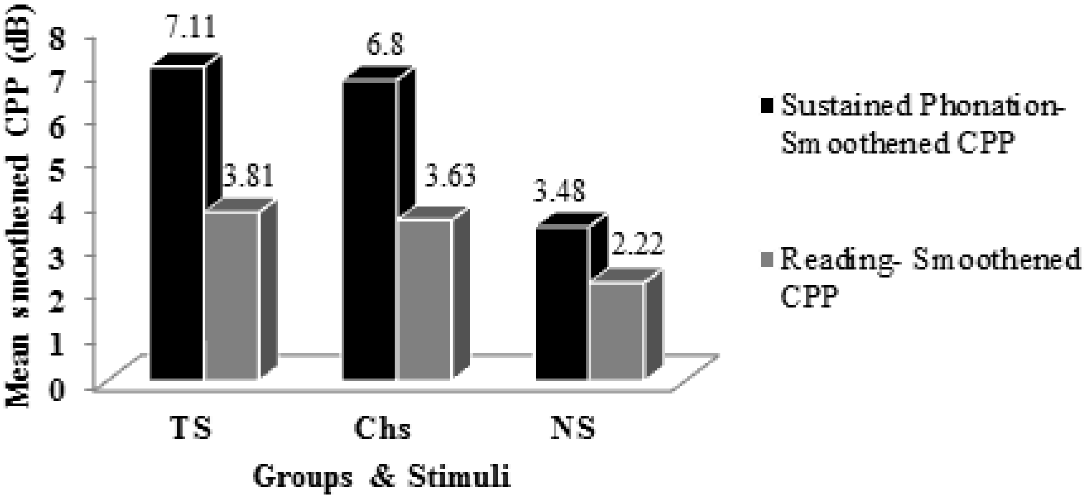

Figure 3: Mean scores of smoothened CPP across Groups and speech tasks. (TS- trained singers, Chs- Choral singers, NS- Non-singers)

Choral singers obtained a mean of 6.80 ($\pm$1.53) for sustained phonation and 3.63 ($\pm$0.48) for reading task. It is apparent from the Table 1, that the mean smoothened CPP values of non-singers are lower than that of trained and choral singers, with the mean smoothened CPP of 3.48 ($\pm$2.31) for sustained phonation and a mean of 2.22 ($\pm$0.46) for reading task. Mixed ANOVA revealed a significant difference [$F(2, 42) = 24.05$, $p<0.05$] across three groups (choral singers, trained singers and non-singers) and also across [$F(1,42) = 57.90$, $p<0.05$] (sustained phonation and reading task). Results of post-hoc Duncan's test indicated that the two singer groups (trained and choral) differed significantly from non-singer group with respect to sustained phonation as well as reading task. However, both trained and choral singers were similar.

In the current study, non-singer group presented with the lower mean smoothened CPP values when compared to choral and trained singers. This can be attributed to the differences in vocal periodicity between singers and non-singers. It might be possible that singers have superior vocal periodicity owing to their regular, systematic practice and maintenance of vocal resonance. Superior vocal quality in trained singers indicates greater periodicity and constant maintenance of sub-glottic pressure along with vocal fold tension, and morphology of closure (across glottal cycles). Since the cepstral peak represents the energy which emerge out of the background noise levels, the peak mainly signifies the harmonic structure of a signal. Higher mean smoothened CPP values in choral singers and trained singers could be due to well defined harmonic structures than non-singers.

Heman-Ackah, Michael and Goding (2002) reported that lower CPP values correspond to higher noise levels in the signal and vice versa. The presence of superior vocal periodicity in singers is also confirmed by the earlier fMRI studies (Zarate & Zatorre, 2008). A similar study by Prakup (2011) reported an elevated perturbation in non-singers when compared to singers. But, interestingly trained singers and choral singers did not vary

Table 1: Mean and standard deviation of smoothened CPP for both sustained phonation and reading task

	Groups	Mean	Std. Deviation
Smoothened CPP- Sustained phonation	Trained Singers	7.11	1.67
	Choral Singers	6.80	1.53
	Non-Singers	3.48	2.31
	Overall	5.71	2.49
Smoothened CPP- Reading task	Trained Singers	3.81	2.28
	Choral Singers	3.63	.48
	Non-Singers	2.22	.46
	Overall	3.18	1.45

in terms of smoothened CPP in the current study. Hence, both trained and untrained singers have rich in harmonics in their voice. Although, choral singers did not undergo any formal training, their learning and style of singing based on self-correction and feedback from senior singers in choir could have contributed to their richness in harmonics. This is supported by the earlier findings that the training results in enhancement of voice with rich harmonic structure. For instance, Vijaitha & Gopikishore (2014) reported that trained singers obtained lower amplitude and frequency perturbations across their pitch range. This can be attributed to the formal training and consistent practice at different notes across the phonation range through sensory motor learning and feedback.

Effect of speech tasks on smoothened Cepstral peak prominence: Results indicated that smoothened CPP values are higher for sustained phonation compared to reading task. This difference was found to be statistically significant on MANOVA for both sustained phonation [F (2, 42) = 17.46, $p<0.05$] and reading task [F (2, 42) = 7.03, $p<0.05$]. Since the interaction effects were seen in mean smoothened CPP across speech tasks in each singer group, paired sample t-test was administered. The results of paired sample t-test revealed highly significant difference for trained singers [t (12) = 3.72, $p<0.05$], choral singers [t (15) = 8.62, $p<0.05$] and non-singers [t (15) = 2.52, $p<0.05$] across two speech tasks. Mean and standard deviation values of the smoothened cepstral peak prominence for sustained phonation and reading task are depicted in Table 1.

Earlier investigations on formant structure pointed out that speech and professional singing differs in terms of vowel quality because of the glottal waveform differences (Sundberg, 1970; Carlson & Sundberg, 1992). Therefore the present study considered the null hypothesis that smoothened CPP value does not differ with respect to speech tasks used for analysis. Results indicated significant difference in smoothened CPP with respect to speech tasks, thus rejecting the null hypothesis. This result is in consistence with literature that CPP and smoothened CPP measures are the strong correlates of vocal quality in both sustained phonation and reading task. Similar study by Awan et al. (2012) reported that vocal intensity and type of vowel have an effect on cepstral measures. They also opined that, either glottal or sub-glottal differences will have an effect on the acoustic signals. In the present study, sustained phonation based smoothened CPP revealed higher amplitudes compared to reading task indicating well defined harmonic structure for sustained phonation. This can be attributed to the inherent variations in periodicity in speech signal due to factors such as transitions from vowel to consonant or consonant to vowel, changes in vowel spectrum and also due to variation of intonation patterns (Watts & Awan, 2011).

Sundberg and Huppmann (2007) opined that vowels are primary musical elements and have longer duration than consonants. Even in the present study CPP with vowels obtained higher amplitudes. However, both CPP obtained using both sustained phonation and reading task is able to differentiate singers from non-singers at acceptable levels of statistical significance, indicating that both yield comparable results. Therefore, considering the fact that reading samples may provide a better representation of naturalistic voice context than sustained phonation (Heman-Ackah, Heuer, & Michael, 2003), this might be considered superior stimuli and preferred stimulus over vowels with sustained phonation for obtaining smoothened CPP.

Conclusions

The results of the current study indicated that choral singers and trained singers had better harmonic structure in voice when compared to non-singers. Thus, indicating that singer's voice is superior to non-singers. Although trained singers obtained slightly higher smoothened CPP values than choral singers, the difference is not statistically significant indicating that they are comparable, irrespective of the training, at least in terms of their harmonic structure. Further, smoothened CPP is able to differentiate singers from non-singers irrespective of the speech tasks used, thus indicating both speech tasks are equivalent. Therefore, considering the naturalness, reading task could be preferred to sustained phonation for measuring smoothened CPP.

Acknowledgements

The authors would like to thank the Director, All India Institute of Speech & Hearing, Mysore, for the support and permitting to carry out this study. The authors also thank all the participants who are involved in the study.

References

Awan, S. N., & Roy, N. (2005). Acoustic prediction of voice type in adult females with functional dysphonia. *Journal of voice, 19,* 268-282.

Awan, S. N., Giovinco, A., & Owen, J. (2012). Effects of vocal intensity and vowel type cepstral analysis of voice. *Journal of voice, 26*(5), 670 e, 15-20.

Balasubramanium, R. S., Bhat, R. S., Fahim., S., & Raju, R. (2011). Cepstral Analysis of Voice in Unilateral Adductor Vocal Fold Palsy. *Journal of Voice, 25*(3), 326-329.

Brown, W. S., Hunt, E., & Williams, W. N. (1988). Physiological differences between the trained and untrained speaking and singing voice. *Journal of Voice, 2*(2), 102-110.

Carlson, G., & Sundberg, J. (1992). Formant frequency tuning in singing. *Journal of Voice, 6*(3), 256-260.

Fant, C. G. M. (1970). *Acoustic Theory of Speech Production.* Mouton & Co., N. V, Printers and publishers, The Hague, Netherland.

Heman-Ackah, Y., & Michael, D., & Goding, G. (2002). Relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16,* 20-27.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech language and Hearing Research, 37,* 769-778.

Jannetts, S., & Lowit, A. (2014). Cepstral Analysis of Hypokinetic and Ataxic Voices: Correlations with Perceptual and Other Acoustic Measures. *Journal of Voice, 28*(6), 673-680.

Kirsh, E. R., Van Leer, E., Phero, H. J., Xie, C., & Khosla, S. (2013). Factors associated with singer's perception of choral singing well-being. *Journal of voice, 27*(6),786 e, 25-32.

Leonard, R. J., Ringel, R., Horii, Y., & Daniloff, R. (1988). Vocal shadowing in singers and non-singers. *Journal of Speech and Hearing Research, 31*(1), 54-61.

Letowski, T., Zimak, L., & Lupinowa, C. H. (1988).Timbre differences of an individual voice in solo and choral singing. *Archtone Acoustics, 13*, 55-65.

Lundy, D. S., Roy, S., Casiano, R. R., Xue, J. W., & Evans, J. (2000). Acoustic analysis of the singing and Speaking Voice in Singing Students. *Journal of Voice, 14*(4), 490-493.

Noll, A. M. (1964). Short-term spectrum and "cepstrum" techniques for vocal pitch detection. *Journal of the Acoustical Society of America, 41,* 293-309.

Prakup, B. (2011). Acoustic measure of the voices of older singers and non-singers. *Journal of voice, 20*(3), 341-350.

Rossing, T. D., Sundberg, J., & Ternstrom, S. (1986). Acoustic comparison of voice use in solo and choir singing. *Journal of the Acoustical Society of America, 76*, 1975-1981.

Rossing, T.D., Sundberg, J., & Ternstrom, S. (1987). Acoustic comparison of soprano solo and choir singing. *Journal of the Acoustical Society of America, 82,* 830-836.

Shipp, T, & Izdebski, K. (1975). Vocal frequency and vertical larynx positioning by singers and non-singers. *Journal of Acoustic Society America, 58,* 1104-1106.

Stacy, R., Brittain, K & Kerr, S. (2002). "Singing for health": an exploration of the issues, *Health education, 102*(4), 156-162.

Sundberg, J. (1973). The source spectrum in professional singing. *Folia Phoniatrica, 25,* 71-90.

Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica, 22,* 28-48.

Sundberg, J. (1974). Articulatory interpretation of the 'singing formant'. *Journal of the Acoustical Society of America, 55,* 838-844.

Sundberg, J. (1975). Formant technique in a professional female singer. *Acoustica, 32,* 89-96.

Sundberg, J. (1977). Singing and timbre. In *Music room acoustics.* Royal Swedish Academy of Music, 17, 57-81.

Sundberg, J. (1979). Formants and fundamental frequency control in singing-an experimental study of coupling between vocal tract and voice source. *Progress and Status Report- Speech Transmission Laboratory, 1*, 65-78.

Sundberg, J., & Bauer-Huppmann, J. (2007). When does a sung tone start?. *Journal of Voice, 21*(3), 285-293.

Sundberg, J. (1979). Perception of singing. *Progress and Status Report- Speech Transmission Laboratory, 1*, 1-48.

Ternstrom, S. (1991). Physical and acoustic factors that interact with the singers to produce choral sound. *Journal of voice, 5*(2), 128-143.

Troup, G. (1981). The physics of the singing voice. *Physics Reports, 74*, 379-401.

Vijaitha, V. S., & Gopikishore, P. (2014). Vocal stability in trained carnatic singers and non singers, *Proceedings of frontiers of research in speech and music (FRSM),* AIISH, Mysore.

Wang, S. (1983). Singing voice: bright timbre, singer's formants and larynx positions. *Proceedings of Stockholm Music Acoustics Conference, Stockholm, 1*, 313-322.

Watts, C. R., & Awan, S. N. (2011). Use of spectral/cepstral analysis for differentiating normal from hypofuntional voices in sustained vowel and continuous speech context. *Journal of Speech Language and Hearing Research, 54*(6), 1525-1537.

Zarate, J. M., & Zatorre, R. J. (2008). Experience dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage, 40*(4), 1871-1887.