Research Article

Concatenation of the Moving Window Technique for Auditory-Perceptual Analysis of Voice Quality

Benjamin Ehrlich,^a Liyu Lin,^a and Jack Jiang^a

Purpose: The purpose of this study is to develop a program to concatenate acoustic vowel segments that were selected with the moving window technique, a previously developed technique used to segment and select the least perturbed segment from a sustained vowel segment. The concatenated acoustic segments were compared with the nonconcatenated, short, individual acoustic segments for their ability to differentiate normal and pathological voices. The concatenation process sometimes created a clicking noise or beat, which was also analyzed to determine any confounding effects.

Method: A program was developed to concatenate the moving window segments. Listeners with no previous rating experience were trained and, then, rated 20 normal and 20 pathological voice segments, both concatenated (2 s) and short (0.2 s) for a total of 80 segments. Listeners evaluated these segments on both the Grade, Roughness, Breathiness, Asthenia, and Strain scale (GRBAS; 8 listeners) and the Consensus Auditory-Perceptual Evaluation of Voice (Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009) scale (7 listeners). The sensitivity and specificity

of these ratings were analyzed using a receiver-operating characteristic curve. To evaluate if there were increases in particular criteria due to the beat, differences between beat and nonbeat ratings were compared using a 2-tailed analysis of variance.

Results: Concatenated segments had a higher sensitivity and specificity for distinguishing pathological and normal voices than short segments. Compared with nonbeat segments, the beat had statistically similar increases for all criteria across Consensus Auditory-Perceptual Evaluation of Voice and GRBAS scales, except pitch and loudness.

Conclusions: The concatenated moving window method showed improved sensitivity and specificity for detecting voice disorders using auditory-perceptual analysis, compared with the short moving window segment. It is a helpful tool for perceptual analytic protocols, allowing for voice evaluation using standardized and automated voice-segmenting procedures.

Supplemental Material: <https://doi.org/10.23641/asha.7100951>

There are an estimated 17.9 million adults diagnosed with dysphonia every year (Bhattacharyya, 2014). This is an issue that leads to difficulties in communication and obtaining jobs and an overall diminished quality of life (Garcia, Laroche, & Barrette, 2002; Klompas & Ross, 2004; Murphy, 2005). Thus, there is a demand for an accurate and swift diagnosis of dysphonia for early and targeted therapy. Auditory-perceptual analysis is a reliable diagnostic technique that is easy to perform and is non-invasive. Speech-language pathologists (SLPs) are trained to rate voice samples of subjects performing the following

tasks: sustained vowel phonation, spontaneous conversation, and reading passages aloud. They then rate voice samples on established scales, such as the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009) or the Grade, Roughness, Breathiness, Asthenia, and Strain scale (GRBAS; Hirano, 1981).

However, this assessment still has issues and limitations that need to be address. It was observed that it was difficult for listeners to observe and rate overlapping parameters for the same segment (Kent, 1996; Kreiman, Gerratt, & Ito, 2007). From changes in listeners' accuracy, it was also shown that the simultaneous task of rating a single perceptual class from a complex voice segment, with many types of variations, was difficult for listeners (Kreiman & Gerratt, 2000). This problem is exacerbated when observing disordered voices, which contain a more diverse selection of variations than

^aSchool of Medicine and Public Health, University of Wisconsin–Madison

Correspondence to Jack Jiang: jjjiang@wisc.edu

Editor-in-Chief: Julie Barkmeier-Kraemer

Editor: M. Preeti Sivasankar

Received July 17, 2017

Revision received December 8, 2017

Accepted May 13, 2018

https://doi.org/10.1044/2018_AJSLP-17-0103

Disclosure: The authors have declared that no competing interests existed at the time of publication.

a normal voice (Watts & Awan, 2015). Methods to isolate components of the voice segment are undeveloped technologies that can be advantageous to listeners.

From previous studies, the moving window technique has been identified as a useful tool for isolating the least perturbed segment in auditory-perceptual analysis. Originally developed by Olszewski, Shen, and Jiang (2011), it was designed to address the nonuniformity in segment selection for acoustic evaluation. The moving window technique divides a recorded voice segment into uniform windows over a period. These windows are then evaluated for perturbation parameters, such as F0, signal-to-noise ratio, percent jitter, percent shimmer, and nonlinear parameters, such as correlation dimension. Based on the collection of calculated parameters, the audio window with the lowest cumulative perturbation and nonlinear dynamics is chosen as a conservative representative of the voice sample. This method is thought to more uniformly select segments for acoustic analysis; therefore, the moving window technique is hypothesized to allow listeners to more easily differentiate pathological and normal voices better than using the whole voice segment and/or segments arbitrarily selected from the middle of the voice segment (Shu, Jiang, & Willey, 2016).

The moving window technique controls for many variables to obtain a more standardized sample of the phonation of the vocal folds. For example, this technique identifies and excludes onset and offset acoustics, varying loudness, and other effects that could vary in the recording. Vocal onset and offset are known to represent periods of changing and unstable phonation. They affect the fundamental frequencies within a voice segment (Parsa & Jamieson, 2001) and affect the biomechanical and aerodynamic properties of the glottis (Choi, Lee, Sprecher, & Jiang, 2012). The duration of the onset and offset periods can also vary greatly and has been noted to change as subjects age (Morris, & Brown, 1994). Varying loudness is another confounding component that can change the subglottic pressure and medial compression of the vocal folds, altering the way the vocal folds vibrate (Glaze, Bless, & Susser, 1990). Even practicing producing a sustained vowel can lead to changes in shimmer and jitter values (Ferrand, 1995). Many of these confounding components are valuable to auditory-perceptual analysis, giving insight into certain pathologies and compensatory states. However, in order for a pathologist to analyze the acoustic characteristics of the voice only, he or she must also be able to use a standardized segment, without these confounding variables, as a supplement to his or her overall analysis of the voice.

With its ability to standardize voice segments for comparison of normal versus disordered voices, find the least perturbed segment, and isolate a segment with less complexity, the moving window will be applicable to auditory-perceptual analysis. Like the moving window technique in acoustic analysis, the moving window technique in auditory-perceptual analysis has the benefit of producing a standardized, comparable segment. Unlike using a midvowel segment for standardization of comparison, the moving window also finds the segment with the least perturbation, thus the segment

with the least number of variables in the voice segment. This could help listeners select criteria instead of being overwhelmed by complex voice segments with many variations.

In order to utilize the possible advantages of the moving window technique in auditory-perceptual analysis, the ratability of the moving window segments first needs to be addressed. The moving window technique produces short segments, which may be difficult to rate. To create a longer, more ratable segment, we propose a program to concatenate, or link, the short 0.2-s moving window segments into a 2-s segment for ease of listening. In this study, the concatenated segment and short segments were evaluated for their ability to differentiate pathological and normal voices.

Because the concatenation process relies upon identifying similar periods to link each segment, this method may be less reliable in Type 3 and 4 pathological voices, which are inherently chaotic and aperiodic (Lin, Calawerts, Dodd, & Jiang, 2015; Zhang, Jiang, Biazzo, & Jorgensen, 2005; Zhang, Sprecher, Zhao, & Jiang, 2011). While our program worked efficaciously to link similar periods in a voice segment, if the segment had irregular periodicity, the repetition of the concatenated segment may have an audible beat or clicking noise as the segment repeats. Post hoc smoothing or removal of these audible beats was not performed to not alter the acoustic characteristics of the segment. Any potential confounding effects of a beat within a concatenated segment should also be assessed to truly evaluate the effectiveness of concatenating the moving window.

Method

Twenty pathological (seven female, 13 male) and 20 normal (11 female, nine male) /a/ sounds were selected from KayPENTAX disordered voice database (Model 4337, Version 1.03, Kay Elemetrics Corp, developed by Massachusetts Eye and Ear Infirmary Voice and Speech Lab). For this initial study, the /a/ sound segment was chosen as it was most easily concatenated, with less variation compared with continuous speech. The normal voices were randomly selected, whereas the pathological segments were chosen to represent common causes of dysphonia, including Reinke's edema, Parkinson's disease, unilateral vocal fold paralysis, and vocal fold lesion.

The moving window technique was completed according to procedures outlined by Shu et al. (2016). Using MATLAB (MATLAB R2015a, The MathWorks, Inc.), two pathological voices and two normal voices were used to determine the ideal segment and step length. A range of 75- to 700-ms long segments at intervals from 25 to 100 ms were evaluated for percent jitter, percent shimmer, and signal-to-noise ratio using TF32 software (Paul Milenkovic, University of Wisconsin-Madison) and correlation dimension using software developed by the Laryngeal Physiology Lab at the University of Wisconsin School of Medicine and Public Health (Shu et al., 2016). A rank sum test was performed to identify the parameters

that had the lowest perturbation scores. This method showed that a 200-ms window at a 75-ms increment produced segments with the lowest perturbations.

Concatenation Program

A classical pitch detection algorithm was used to accurately estimate the fundamental frequency of the voice samples used to synthesize the 2-s signals. The *short-term autocorrelation function* of a limited voice signal was defined as follows (Sakshat Virtual Labs “iitg.vlab.co.in,” 2011):

$$R(m) = \sum_{n=-\infty}^{n=+\infty} x(n)x(n+m) \quad (1)$$

This function describes the similarity between a signal and its m -point-shift copy. The program compares the signal to m -points (points in a certain period) that are shifted, until the m -point-shift copy (corresponding m -point in other periods) matches the signal. This maximal point in the autocorrelation function represents the period. The time duration between two neighboring peaks, which is recorded as T , can be used to calculate the period of the voice signal. To ensure the accuracy of the estimation, we used the average distance between multiple peaks to calculate the period.

Once the segment's period was calculated, the second step was to identify points that aligned in the tail and front periods or the matched connect point pair. The tail end was selected first. To avoid any variability potentially introduced while recording, the 50 points at the tail end were not used. The last period is then recorded as x_{p_last} ; this is the first connect point. The second connect point must then be selected from the beginning of the signal. Period length segments are taken from the beginning of the segment. The first segment is called x_{p_1} , then x_{p_2} , and so on until x_{p_n} . The correlative coefficient is then calculated with the following equation:

$$C_{corr1,2,3\dots n} = - \sum_{m=1}^{m=T} |x_{p_last}(m) - x_{p1,2,3\dots n}(m)| \quad (2)$$

The x_{p_n} with the maximum correlative coefficient was selected as the second connect point. The first connect point was then added to the second connect point. This process was repeated until the segment was 2 s long.

Training of Listeners

Inexperienced listeners were chosen as raters for this study to identify the effect of beats on perceptual analysis ratings. Naive listeners are more likely to be affected by the presence of a beat, compared with experienced SLPs who may have developed tools and habits of analyzing acoustic features to circumnavigate any influence from any produced beat (Helou et al., 2010).

The inexperienced listeners were trained according to the methods detailed by Helou et al. (2010). Each inexperienced listener was provided with an introduction and tutorial

for using the GRBAS and CAPE-V scales and an explanation of how auditory-perceptual evaluation can aid in diagnosis and how various vocal fold pathologies can affect speech (Supplemental Material S1). The listeners then practiced rating on a set of modules from an educational website (<https://csd.wisc.edu/>) designed to teach the audience about variations in voices by providing examples of voices with each of these variations. Finally, the modules included the opportunity to practice rating voices and, then, compare the ratings to those of a trained SLP. Each listener repeated the modules until he or she felt confident with the rating scales. Each listener was also given two control segments, taken from a normal, healthy voice, to use as reference. Each listener required 1 to 3 hrs to complete the training modules and to feel comfortable using the scales.

Segments for Rating

Each listener was given 20 pathological concatenated, 20 pathological short, 20 normal concatenated, and 20 normal short segments in a randomly assigned order. The concatenated and short segments came from the same samples to serve as a control. To calculate intralistener accuracy, two segments in the pathological concatenated and two in normal short categories were included twice, under different names for rating.

Listener Inclusion

Because the raters were inexperienced listeners, it was important to evaluate the consistency of their ratings. To evaluate the intrarater reliability, a t test on the repeated segments was conducted. Listener's ratings were considered to be reliable if they were consistently statistically similar ($\alpha = .05$).

To assess interrater reliability, the intraclass correlation coefficient (ICC) was calculated using R Studio (R Core Team, 2013; Helou et al., 2010). A backward stepwise regression model was used to determine the highest combination of interrater reliability. If the listener consistently lowered the average ICC on over half of the criteria for either of the two tests, his or her ratings were removed. Lowering of the average for over half of the criteria demonstrated a clear lack of understanding or inattention to detail for the ratings.

Statistics

To evaluate the sensitivity, specificity, and overall accuracy of the data, a receiver-operating characteristic (ROC) curve was constructed using SigmaPlot (11.0, Systat Software). The criteria, overall sensitivity in the CAPE-V and grade in the GRBAS scales, were used as they represented a holistic rating of the voice. The area under the curve (AUC) was calculated from the ROC curve. A value of 0.5 indicated that there was no correlation between pathological and normal voices, whereas a value of 1 signified that there was a perfect difference between normal and pathological (Hanley & McNeil, 1982). Methods proposed

by DeLong, DeLong, and Clarke Pearson (1988) were used to determine statistical significance while comparing AUC values.

It was necessary to evaluate any effect of any produced beat on the ratings. Voice segments were divided on whether there was an audible repetition from the concatenation. Nineteen segments were considered nonbeat, whereas 21 were labeled as *containing a beat*. Differences between concatenated and short segments ratings were calculated for both beat and nonbeat ratings. The difference between nonbeat and beat differences was then calculated. A two-tailed analysis of variance (ANOVA) was performed to assess if there were similar trends between differences for the criteria in the CAPE-V and GRBAS scales. Because there were 21 beat segments and 19 nonbeat segments, to calculate the difference for the two extra segments, two nonbeat segments were randomly selected to create equal sample sizes.

Results

Listener Selection

Seven listeners' data were used for the CAPE-V, and eight listeners' data were used for the GRBAS evaluation. After determining intralistener reliability, one listener's CAPE-V data were excluded. Two listeners' data for both the CAPE-V and GRBAS were not included because of interlistener differences. Ninety-five percent confidence interval and average ICC values for the listeners on the CAPE-V scale and GRBAS scale are shown in Tables 1 and 2, respectively.

Accuracy

The ROC curve for grade (GRBAS; see Figure 1) had an AUC value of 0.83 ($SE = 0.02$) for concatenated segment ratings and 0.76 ($SE = 0.02$) for short segment ratings with $p = .030$. The AUC for concatenated segment ratings for the overall sensitivity (CAPE-V; see Figure 2) were 0.84 ($SE = .02$) and 0.77 ($SE = 0.03$) for short segments with $p = .071$.

Beat Evaluation

Differences between concatenated and short segments for beat and nonbeat and the difference between those differences are shown in Figures 3 and 4 for both the GRBAS

and CAPE-V scales. The ANOVA test resulted in a statistically insignificant effect of differences in beat and nonbeat in the GRBAS scale at the $p < .05$ level for the five criteria, $F(3) = 0.3, p = .84$. For the CAPE-V, the ANOVA test revealed a statistically significant effect of differences in beat and nonbeat at the $p < .05$ level for the six conditions, $F(4) = 3.7, p = .0085$. On average, the difference between beat and nonbeat GRBAS scores was 0.14 (± 0.03) higher, whereas on the CAPE-V scale, the difference was 5.538 (± 3.3) higher.

Discussion

This study was performed to address issues outlined in previous studies regarding how the complexity and the multifaceted nature of voice segments can lead to decreases in listeners' accuracy (Kent, 1996; Kreiman & Gerratt, 2000; Kreiman et al., 2007). No studies have focused on isolating components of the voice segment to allow listeners to rate standardized voice segments. Although there have been other studies, such as the concatenation of continuous speech and sustained vowels (Maryn, Dick, Vandenbruaene, Vauterin, & Jacobs, 2009), driven to manipulate the voice segment to improve auditory-perceptual analysis, these studies did not pursue simplifying the rating of complex voices.

Accuracy of Grade and Overall Severity in Differentiating Normal and Pathological Voices

Our results confirmed the utility of concatenating the moving window. Grade ratings from the GRBAS scale were able to successfully differentiate normal voices from pathological voices with a statistically significant difference in AUC values ($p = .030$). The ability to differentiate normal and pathological voices on the CAPE-V scale was not statistically different for the overall sensitivity criteria on the CAPE-V scale ($p = .071$).

Possible reasons for the lack of statistical difference between normal and pathological voices on the CAPE-V scale could have been the experience of the listeners, loss of listeners, or because of the structure of the CAPE-V scale. In addition, the lack of statistical difference between normal and pathological voices for the CAPE-V scale compared with the GRBAS scale could be because of the increased amount of choices on the CAPE-V scale (100) versus the GRBAS scale (three).

Table 1. Ninety-five percent confidence interval (CI) and average intraclass correlation coefficient (ICC) values for the seven listener ratings on the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) scale.

Parameter	95% CI for concatenated	95% CI for short	ICC value concatenated	ICC value short
Severity	.552 < ICC < .779	.518 < ICC < .761	.666	.639
Roughness	.448 < ICC < .704	.389 < ICC < .664	.572	.519
Breathiness	.447 < ICC < .704	.521 < ICC < .763	.571	.643
Strain	.387 < ICC < .654	.309 < ICC < .593	.513	.439
Pitch	.027 < ICC < .209	.062 < ICC < .287	.0967	.151
Loudness	.049 < ICC < .248	.132 < ICC < .39	.126	.239

Table 2. Ninety-five percent confidence interval (CI) and average intraclass correlation coefficient (ICC) values for the eight listener ratings on the Grade, Roughness, Breathiness, Asthenia, and Strain (GRBAS) scale.

Parameter	95% CI for concatenated	95% CI for short	ICC value concatenated	ICC value short
Grade	.558 < ICC < .778	.529 < ICC < .763	.67	.646
Roughness	.468 < ICC < .714	.44 < ICC < .699	.589	.565
Breathiness	.459 < ICC < .708	.426 < ICC < .688	.58	.552
Asthenia	.379 < ICC < .642	.358 < ICC < .63	.503	.485
Strain	.301 < ICC < .57	.25 < ICC < .525	.424	.372

Evaluation of the Produced Beat Compared With Nonbeat Segments

To evaluate whether the beat affected the listener's ability to rate voice segments, we compared the average ratings of segments with a beat to those without a beat. Without being able to directly compare the same segment with and without a beat, criteria from the scales were observed to see if certain criteria were affected more so than others. Figure 3 shows the effect of a beat on ratings for the GRBAS scale. There is an increase across all criteria, which was shown to be statistically similar ($p = .84$). For ratings performed on the CAPE-V scale, there was an increase for all criteria except for pitch and loudness. These criteria were found to be statistically different ($p = .0085$).

The GRBAS and CAPE-V scales share four of the same criteria: grade/overall sensitivity, breathiness, roughness, and strain. When performing an ANOVA on the

difference of differences for those criteria in the CAPE-V and GRBAS scales, there was statistical similarity for both the CAPE-V, $F(2) = 0.7$, $p = .52$, and GRBAS, $F(2) = 0.4$, $p = .68$. For both the GRBAS scale and for the four similar criteria in the CAPE-V scale, there was likely no influence from the produced beat with increases across all criteria. This was an expected increase as the produced beat was representative of a more chaotic voice segment and, thus, a higher rating. It is possible that loudness and pitch did not have increases because the increased time to rate the segments, provided by the concatenation program, was not needed to effectively rate these criteria.

Limitations

Some listeners were not included in our data due to intrarater or interrater reliability. This resulted in the use of eight listeners for the GRBAS scale and seven listeners

Figure 1. Receiver-operating characteristic curve of the Grade, Roughness, Breathiness, Asthenia, and Strain ratings of normal and pathological voices. The black line represents concatenated segments (2 s), whereas the red represents short segments (200 ms).

Figure 2. Receiver-operating characteristic curve of the Consensus Auditory-Perceptual Evaluation of Voice ratings of normal and pathological voices. The black line represents concatenated segments (2 s), whereas the red represents short segments (200 ms).

Figure 3. Differences in short and concatenated beat and normal segments on the Grade, Roughness, Breathiness, Asthenia, and Strain (GRBAS) scale. The difference between the beat and normal segments is shown with the green line. The Grade, Roughness, Breathiness, Asthenia, and Strain scale is out of 3 points.

for the CAPE-V scale. This lower number of listeners may have affected the significance of our results; however, despite the exclusion of these data, the difference for the GRBAS scale was large enough to have statistical significance. In addition, inexperienced listeners were tested to observe how a naive audience would perceive a produced beat; however, as Helou et al. (2010) reported, this can lead to larger variations in rating scores. Table 1 displays the ICC values for listeners on the CAPE-V scale. These values are lower than other previous studies that calculated ICC for SLPs rating on the CAPE-V scale (Helou et al., 2010; Kelchner et al., 2010). Further testing using experienced listeners would help

elucidate whether there is a statistical difference between normal and pathological voices.

In addition, the vowel stimulus duration of 2 s needs to be evaluated. Other studies have used different amounts of time for listeners to rate sustained vowels. Kempster et al. (2009) made great progress in standardizing auditory-perceptual analysis by creating the CAPE-V scale and stating that sustained vowels should be 3–5 s. Our goal in this study was to evaluate if the concatenation of the moving window technique for use in auditory-perceptual analysis was possible and if it could further standardize perceptual evaluation. In future studies, we will explore the optimal amount of time

Figure 4. Differences in concatenated and normal segments of nonbeat and beat segments rated on the Consensus Auditory-Perceptual Evaluation of Voice scale. Differences between beat and nonbeat are described with the green line. The Consensus Auditory-Perceptual Evaluation of Voice scale is out of 100 points.

for a listener to have to most accurately rate voice segments. The concatenation program has the ability to extend a voice segment for any amount of time, which will be advantageous in finding the optimal segment length for listeners.

Summary and Future Aims

With the developed methodology and observed results in this study, there are future applications of our technique to pursue. Our segment selection was at 200 ms, which is near the mean time for vowel production (Jacewicz & Fox, 2008). With the 200-ms segment being efficaciously used with the moving window, we would like to utilize the moving window with the concatenation program in extracting vowels from continuous speech. It has been previously observed that continuous speech is more representative of natural phonation than a sustained vowel and would be better for auditory-perceptual analysis (Eadie & Doyle, 2005). Isolating and concatenating vowel segments from continuous speech will be a study pursued in the future.

In addition, in future studies, we would like to test the utility of the moving window technique as a supplement for auditory-perceptual analysis. We will test if listeners better differentiate normal and pathological voices with the concatenated moving window technique, sustained vowel, and continuous speech segments versus just the sustained vowel and continuous speech segments.

The utility of the concatenation program with the moving window technique for the auditory-perceptual analysis will be a potential asset for SLPs and for future applications. We have developed a program that makes it easier for the listener to extract a segment with the lowest perturbation, devoid of many distracting variations, to help listeners focus on the quality of the voice. While we acknowledge the utility of some of these perturbations in auditory-perceptual analysis, we also see the necessity to expand the tools available for SLPs and give them the ability to isolate the voice quality and rate less complicated voice segments.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant 2 R01 DC006019-06A1 awarded to Dr. Jack Jiang.

References

- Bhattacharyya, N. (2014). The prevalence of voice problems among adults in the United States. *The Laryngoscope*, 124(10), 2359–2362.
- Choi, S. H., Lee, J., Sprecher, A. J., & Jiang, J. J. (2012). The effect of segment selection on acoustic analysis. *Journal of Voice*, 26(1), 1–7.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44, 837–845.
- Eadie, T. L., & Doyle, P. C. (2005). Classification of pathological voice: Acoustic and auditory-perceptual measures. *Journal of Voice*, 19(1), 1–14.
- Ferrand, C. T. (1995). Effects of practice with and without knowledge of results on jitter and shimmer levels in normally speaking women. *Journal of Voice*, 9(4), 419–423.
- Garcia, L. J., Laroche, C., & Barrette, J. (2002). Work integration issues go beyond the nature of the communication disorder. *Journal of Communication Disorders*, 35(2), 187–211.
- Glaze, L. E., Bless, D. M., & Susser, R. D. (1990). Acoustic analysis of vowel and loudness differences in children's voice. *Journal of Voice*, 4(1), 37–44.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36.
- Helou, L. B., Solomon, N. P., Henry, L. R., Coppit, G. L., Howard, R. S., & Stojadinovic, A. (2010). The role of listener experience on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ratings of postthyroidectomy voice. *American Journal of Speech-Language Pathology*, 19(3), 248–258.
- Hirano, M. (1981). *Clinical examination of voice*. New York, NY: Springer Verlag.
- Jacewicz, E., & Fox, R. A. (2008). Amplitude variations in coarticulated vowels. *The Journal of the Acoustical Society of America*, 123(5), 2750–2765.
- Kelchner, L. N., Brehm, S. B., Weinrich, B., Middendorf, J., deAlarcon, A., Levin, L., & Elluru, R. (2010). Perceptual evaluation of severe pediatric voice disorders: Rater reliability using the consensus auditory perceptual evaluation of voice. *Journal of Voice*, 24(4), 441–449.
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18(2), 124–132.
- Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, 5(3), 7–23.
- Klompas, M., & Ross, E. (2004). Life experience of people who stutter, and the perceived impact of stuttering on quality of life: Personal accounts of South African individuals. *Journal of Fluency Disorders*, 29(4), 275–305.
- Kreiman, J., & Gerratt, B. R. (2000). Sources of listener disagreement in voice quality assessment. *American Journal of Acoustic Society*, 108(4), 1867–1876.
- Kreiman, J., Gerratt, B. R., & Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *The Journal of the Acoustical Society of America*, 122(4), 2354–2364.
- Lin, L., Calawerts, W., Dodd, K., & Jiang, J. J. (2015). An objective parameter for quantifying the turbulent noise portion of voice signals. *Journal of Voice*, 30(6), 1–6.
- Maryn, Y., Dick, C., Vandenbruaene, C., Vauterin, T., & Jacobs, T. (2009). Spectral, cepstral, and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels. *Laryngoscope*, 119(12), 2384–2394.
- Morris, R. J., & Brown, W. S., Jr. (1994). Age-related differences in speech variability among women. *Journal of Communication Disorders*, 27, 49–64.
- Murphy, J. (2005). Perceptions of communication between people with communication disability and general practice staff. *Health Expect*, 9(1), 49–59.
- Olszewski, A. E., Shen, L., & Jiang, J. J. (2011). Objective methods of sample selection in acoustic analysis of voice. *Annals of Otolaryngology, Rhinology & Laryngology*, 120(3), 155–161.
- Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. *Journal of Speech, Language, and Hearing Research*, 44(2), 327–339.

-
- R Core Team.** (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Sakshat Virtual Lab.** (2011). *Short term time domain processing of speech*. Retrieved July 10, 2017 from <http://iitg.vlab.co.in/>
- Shu, M., Jiang, J. J., & Willey, M.** (2016). The effect of moving window on acoustic analysis. *Journal of Voice*, 30(1), 5–10.
- Watts, C. R., & Awan, S. N.** (2015). An examination of variations in the cepstral spectral index of dysphonia across a single breath group in connected speech. *Journal of Voice*, 29(1), 26–34.
- Zhang, Y., Jiang, J. J., Biazzo, L., & Jorgensen, M.** (2005). Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *Journal of Voice*, 19(4), 519–528.
- Zhang, Y., Sprecher, A. J., Zhao, Z., & Jiang, J. J.** (2011). Non-linear detection of disordered voice productions from short time series based on a Volterra–Wiener–Korenberg model. *Chaos, Solitons & Fractals*, 44(9), 751–758.