Strathprints Institutional Repository

Jannetts, Stephen and Lowit, Anja (2014) Cepstral analysis of hypokinetic and ataxic voices : correlations with perceptual and other acoustic measures. Journal of Voice, 28 (6). pp. 673-680. , http://dx.doi.org/10.1016/j.jvoice.2014.01.013

Cepstral analysis of hypokinetic and ataxic voices: correlations with perceptual and other acoustic measures

Author Details:

Stephen Jannetts, BSc.

Anja Lowit, PhD.

Speech and Language Therapy

School of Psychological Sciences and Health

University of Strathclyde

Graham Hills Building

40 George Street

Glasgow, G1 1QE

Tel. +44 (0)141 548 3102

Office: +44 (0)141 548 2700

Abstract

Objectives: To investigate the validity of cepstral analyses against other conventional acoustic measures of voice quality in determining the perceptual impression in different motor speech disorders - hypokinetic and ataxic dysarthria, and speech tasks - prolonged vowels and connected speech.

Methods: Prolonged vowel productions as well as connected speech samples (reading passages and monologues) from 43 participants with Parkinson’s disease and 10 speakers with ataxia were analysed perceptually by a trained listener using GRBAS. In addition, acoustic measures of cepstral peak prominence (CPP), smoothed CPP (CPPs), harmonics-to-noise ratio (HNR), shimmer %, shimmer dB, amplitude perturbation quotient (APQ), relative average perturbation (RAP), jitter, and pitch perturbation quotient (PPQ) were performed. Statistical analysis involved correlations between perceptual and acoustic measures, as well as determination of differences across speaker groups and elicitation tasks.

Results: CPP and CPPs results showed greater levels of correlation with overall dysphonia, breathiness, and asthenia ratings than the other acoustic measures, except in the case of roughness. Sustained vowel production produced a higher number of significant correlations across all parameters other than connected speech, but task choice did not affect CPP and CPPs results. There were no significant differences in any parameters across the two speaker groups.

Conclusions: The results of this study are consistent with the results of other studies investigating the same measures in speakers with non-motor related voice pathologies. In addition, there was an indication that they performed better in relation to asthenia, which might be particularly relevant for the current speaker group. The results support the clinical and research use of CPP and CPPs as a quantitative measure of voice quality in populations with motor speech disorder.

Introduction

A reliable and valid assessment of voice quality is essential for the management of voice disordered patients, both contributing to differential diagnosis and functioning as an outcome measure.[1-3] Perceptual analysis is considered a central component in this assessment, as Orlikoff *et al*[4] describe that for many patients, the most important factor of their voice disorder is how their voice is perceived by others. There are a number of methods to quantify perceptual analysis, and most make use of a rating scale. The two main approaches are either to describe the voice quality in global terms (e.g. 'severely disordered', 'normal voice') or to define characteristics of the voice such as the degree of breathiness or roughness.[5]

Two auditory-perceptual evaluation tools used worldwide by clinicians are GRBAS and CAPE-V. GRBAS evolved from the work of researchers in the Japanese Association of Logopaedics and Phoniatrics and was described by Hirano in 1981.[6] Four parameters, G – grade (overall dysphonia severity); R – roughness; B – breathiness; A – asthenia (weakness); and S – strain, are rated on a four-point interval scale. A score of 0 represents an absence of impairment. A score of 1, 2 or 3 represents mild, moderate or severe severity of the parameter, respectively. The work of Kreiman *et al*[1] and Gerratt *et al*,[7] in which the authors argue that ordinal interval scales may have limited reliability and that visual analog scales may address limitations of ordinal interval scales, led to a new scaling tool produced by clinicians and voice scientists at the Consensus Conference for Perceptual

Measure of Voice Quality in 2002 called Consensus Auditory Perceptual Evaluation—Voice (CAPE-V).[8] The CAPE-V uses six parameters, Overall Severity, Roughness, Breathiness, Strain, Pitch and Loudness, each rated by a mark placed on a 100mm visual analog scale with predetermined vocal tasks. The far left side represents the least impaired status and the far right represents the most impaired status. Scores are derived from measuring the distance from the leftmost side of the line to the mark in mm.

Given the increased resolution of the measurements obtainable from CAPE-V, this tool is more sensitive to the parameters measured compared with the four-point scale of GRBAS.[9] Additionally, CAPE-V measures two additional parameters – pitch and loudness – and omits asthenia from the evaluation. Despite these differences, studies have shown that there is a high degree of consensus between the two scales, both when used jointly[10] or independently[11] on the same vocal samples. Either thus appears to be an appropriate tool to apply in research and clinical practice.

Although the above perceptual scales have proven to be valid reflections of a speaker's voice quality, their usefulness is limited by poor reliability.[1] Eadie and Baylor[12] have shown that reliability of ratings can be affected by listeners' internal standards, i.e. their exposure to different types of voices, as well as their level of training, and fatigue. In addition, the type of speech task (continuous speech versus sustained vowel) and the type of scale used have been shown to affect perceptual judgement.[12-15] Although Awan and Lawson[16] were able to show that training combined with the use of auditory and textual 'anchors' was able to significantly increase agreement of raters, perceptual analysis is likely to always suffer from at least some degree of reliability issues.

This problem, combined with the increasing focus on evidence-based practice in SLT treatment has led to the development of more objective, quantifiable measures of voice quality. Such alternatives include e.g. aerodynamic, acoustic, or vibratory measures such as electroglottographic wavegrams[17]. Amongst these methods, acoustic analysis has become a popular technique for complementing the perceptual analysis of voice quality, considering the relatively low cost of digital recording hardware, ease of use and non-invasive nature of the procedure.[18, 19]

Buder[20] identifies more than 100 different acoustic analysis algorithms that have been developed during the 20th century, many of which can now be implemented in a clinical setting. However, the validity of many acoustic analysis measures in use by clinicians has been debated since their proliferation in the nineties.[5, 19, 21] For example, shimmer, jitter, and harmonics-to-noise ratio (HNR) are frequently used measures of perturbation and noise, however, many studies show that these parameters do not correlate well with corresponding perceptual measures of dysphonic voices.[22-27] This has been partly attributed to the fact that a single acoustic marker may not be easily associable with perceptions of dysphonia. For this reason, some researchers have considered a multiparameter approach for the measurement of voice quality.[28-33] However, what is not entirely clear to date is which parameters of combination thereof are best suited to reflect voice quality.

Maryn, Roy, De Bodt *et al*[34] conducted a meta-analysis of studies evaluating correlations between acoustic measures and overall voice quality. 25 studies were included in total, with 87 acoustic measures having been performed. The authors concluded that of these 87 acoustic measures four were acceptable as markers of overall voice quality in sustained vowels: Pearson *r* at autocorrelation peak; pitch amplitude; spectral flatness of residue signal; and smoothed cepstral peak prominence. Three measures were acceptable for continuous speech: signal-to-noise ratio as described by Qi *et al*;[35] cepstral peak prominence; and smoothed cepstral peak prominence. The authors also noted that the level of aperiodicity (as represented by CPP and CPPs) in the voice signal appeared to determine overall dysphonia severity.

The reason for the superior performance of CPP/CPPs against HNR, shimmer and jitter may be that the latter are dependent on a time-based, cycle-to-cycle analysis separating the speech signal into discrete pitch periods. The nature of voice disorders means that the speech signals are generally not periodic and extraction of the fundamental frequency (F_0) is poor. In contrast, measures of CPP are not based on F_0 extraction and are based on averaging the voice signal over its entire length, rather than cycle-to-cycle measurements, thus reducing the error in these measurements.

A large number of studies have now established the reliability of CPP measures in predicting overall dysphonia and breathiness for a variety of voice pathologies such as muscle tension dysphonia or nodules.[12, 29, 34, 36-44] Despite these encouraging results, no investigations have been performed on speakers with motor speech disorders to date, despite the prevalence of voice quality changes in this population.[45-47] In a prevalence study, Cooper[48] found that 4.8% of patients in his practice had dysphonia with a neurological basis. Furthermore, Logemann *et al*[49] found that 89% of 200 individuals with PD had voice disorders; more than those with articulation difficulties in this case.

Acoustic analysis of voice quality in PD in particular has received significant attention from researchers. The most commonly reported acoustic markers of dysphonia in PD are reduced maximum phonation time (MPT); alterations in HNR, shimmer and jitter; and reduced F_0 variability.[50-52] As in voice disorders, these findings are not universal. Yücetürk *et al*[53] report that their 30 PD patients had significantly different MPT and HNR from their 20 controls. However, they found no significant difference for jitter or shimmer. Furthermore, Santos *et al*[54] observed no difference in any of their acoustic measures used, including F_0, shimmer, jitter and HNR; although their sample size was small (5 PD and 5 control participants). Where reported, all of these studies agree that perceptual voice quality in PD is characterised by roughness, breathiness and reduced loudness.

Fewer studies are available on the analysis of voice quality in ataxic dysarthria. Perceptually, Gilman and Kluin[55] as well as Hertrich *et al*[56] report harshness and breathiness in their description of ataxic voice quality. A study of the acoustic analysis of ataxic dysarthria by Kent *et al*[57] found excessive variation of F_0 was the most frequent marker, with high shimmer in both gender groups and high jitter in women only.

In summary, there is significant evidence that voice quality abnormalities are of concern in the MSD population. In addition, the literature reporting acoustic analyses of these disorders has grown to the point that an acoustic typology of them can be constructed.[58] Yet, despite the reported benefits of CPP analyses over more traditionally employed acoustic parameters for specific voice pathologies, there is currently no information on the suitability of the cepstral measures in the motor speech population.

This study aims to fill this gap and investigate the validity of cepstral analyses in reflecting the perceptual impression of overall dysphonia, roughness, breathiness, and asthenia in speakers with motor speech disorders. As part of this investigation, the study examined (1) how cepstral analyses compare to other conventionally used acoustic measures of voice quality in how well they correlate with perceptual measures in this population; (2) how well any of the acoustic measures were able to discriminate between the different types of motor speech disorders;

and (3) whether there was a difference in measurement reliability and validity for speech and non-speech tasks across the measures.

Method

Participants

Voice samples of 43 speakers with PD and 10 participants with ataxia were used in this study. The PD group consisted of 31 males and 12 females. Ages ranged from 46 to 85 years, with a mean ± SD of 68.02 ± 8.58 years. PD severity was rated with the Hoehn and Yahr scale.[59] Stages ranged from 1 to 5, with a median stage of 2.5. The ataxia group consisted of 5 males and 5 females. Ages ranged from 28 to 72 years, with a mean ± SD of 51.5 ± 13.6 years. Severity of dysarthria in both groups ranged from mild (no articulatory impairment but mild changes to voice quality and volume) to moderate/severe (significant articulatory impairment leading to largely unintelligibile speech, combined with changes to voice quality of various degrees).

Procedure

The data had been recorded as part of other investigations by the second author.[60, 61] Three recordings were used for each participant consisting of: 1. sustained phonation of /a/ for as long as comfortable; 2. a reading passage (the Cherry tree passage[62] for the PD speakers, and the Cinderella Passage[63] for the ataxia participants); and 3. a spontaneous speech sample, consisting of a monologue about a favourite holiday (PD) or story recall of the reading passage (ataxia).

Data had been recorded in the participants' homes or at the university with a DAT recorder (Tascam DA-P1) and a condenser microphone, and digitised using *CSL* (Kay Elemetrics, Model 4300B) at a sampling rate of 20 kHz.

These recordings were edited in preparation for analysis as follows. The sustained vowel samples were trimmed to include only a 3 second portion starting from one second into the sample. All editing was done automatically using *Sony Sound Forge Pro* software (version

10.0). Manual checking of the continuous speech samples excluded any pauses lasting longer than 2 seconds and any unusable parts (e.g. speech from the interviewer, coughing, intrusive external noise). The samples were then automatically trimmed to include a 20 second portion from the middle. Finally, a second version of the connected speech samples was created where the voiced segments were removed using a modified Praat script by Paul Corthals.[64] This resulted in five recordings for each participant: one sustained vowel; one original reading passage; one reading passage with only voiced segments; one original sample of spontaneous speech; and one sample of spontaneous speech with only voiced segments.

Dysphonia ratings

Studies have shown that the two widely used auditory perceptual evaluation tools, GRBAS and CAPE-V, have high reliability with each other, and either can be used in auditory-perceptual evaluation.[10, 11] The omission of asthenia or weakness in CAPE-V and the pertinence of measuring this parameter in MSD, particularly the PD population,[65] meant that GRBAS was selected for use in this study.

GRBAS scores were determined by the first author, who had received voice analysis training during the SLT qualifying course and additional input as part of the research internship. The original versions of the connected speech samples were randomised and analysed in one single sitting. In order to determine intra-rater reliability, 50% of the samples were rated a second time one week later. A further 10% of the samples were rated by an experienced speech and language therapy researcher with GRBAS training and several years of experience judging voice quality to determine inter-rater reliability.

Intra/interrater reliability was explored using the Spearman rank-order correlation coefficient (r_s). Reliability was high for almost all parameters across intra- and inter-rater comparisons. For intra-rater reliability, the analysis yielded: grade ($r_s = 0.863$, $p < 0.001$), roughness ($r_s = 0.818$, $p < 0.001$), breathiness ($r_s = 0.811$, $p < 0.001$), and asthenia ($r_s = 0.749$, $p < 0.001$). For inter-rater reliability values were: grade ($r_s = 0.895$, $p < 0.001$), roughness ($r_s = 0.682$, $p = 0.005$), breathiness ($r_s = 0.690$, $p = 0.004$), and asthenia ($r_s = 0.779$, $p = 0.001$).

On the other hand, both inter- and intra-rater reliability was poor for strain ($r_s = 0.351$, $p = 0.081$; $r_s = 0.271$, $p < 0.261$). The poor reliability for this parameter is a reflection of a methodological issue related to GRBAS scoring, i.e. there is a visual perceptual element in the rating of strain, which renders investigations that are purely based on audio materials difficult. On this basis, only G, R, B, and A were used in this investigation.

Acoustic Analysis

In order to identify the best match between perceptual and acoustic measures, a wide variety of acoustic methods were initially investigated, including percent jitter, relative average perturbation (RAP), pitch perturbation quotient (PPQ), percent shimmer, shimmer in dB, amplitude perturbation quotient (APQ), and harmonics-to-noise-ratio (HNR). All measures were performed twice on two commonly used analysis systems, *Multi-Dimensional Voice Program*™ (MDVP) and Praat, with the exception of HNR, which was only available from Praat. In addition, cepstral peak prominence (CPP) and smoothed CPP (CPPs) were measured for all samples. A detailed description of the CPP and CPPs algorithm is provided in Hillenbrand *et al*[33] and Hillenbrand and Houde.[66] Current data were extracted using Hillenbrand's script (available online at: http://homepages.wmich.edu/~hillenbr/cpps.exe).

In total, 15 measures were thus gathered for each sample, resulting in around 3600 data points.

Statistical Analysis

Statistical analyses were completed using *SPSS* (version 19.0). The correlation between perceptual ratings of G, R, B, and A, and the 15 acoustic measures was determined using the Spearman rank-order correlation coefficient (r_s). The Mann-Whitney U test was used to compare differences between the participants with PD and ataxia and the Wilcoxon singed-rank test was used to compare the measurement values across the two analysis systems. An alpha level of 0.05 was used to determine statistical significance.

Results

Data Reduction

In order to reduce the 3600 data points, these were compared with each other to establish which measures from which analysis system corresponded best to the perceptual evaluation. The Wilcoxon signed-rank test and Spearman rank order correlation were used to compare the data. The results suggested that MDVP data correlated more often with perceptual results than Praat findings. Only the HNR measure was therefore retained from Praat. There was no significant difference between reading and spontaneous speech data, which made one of these data sets redundant. Only the reading data were therefore included as these were more comparable across participants. Significant differences were found between measures of the original connected speech samples and the samples including only voiced segments. Results from the latter correlated better with overall dysphonia ratings (G). This is in line with the convention to use voiced only segments when analysing connected speech (thus emulating sustained vowel production), and only the filtered connected speech data were therefore used for subsequent analysis.

As a result of this data reduction exercise, the following report of results will only concentrate on the measures taken from the sustained vowel and voiced segment only reading passage data, and restrict itself to CPP, CPPs (from Hillenbrand's script), jitter (%), RAP, PPQ, shimmer (%), shimmer (dB), APQ (from MDVP) and HNR (from Praat).

Auditory-perceptual ratings

Figures 1 and 2 show the participants' perceptual scores G, R, B, and A for sustained vowel and reading tasks. In the sustained vowel task, median values for G, R, B, and A were 1.5, 0, 1, and 0 respectively. In the reading task, median values for G, R, B, and A were 1, 0, 0, 1 respectively.

--- insert figures 1 & 2 around here ---

Acoustic Measures

Summary statistics for each of the acoustic measures are shown in Table 1, and correlations between perceptual ratings and the acoustic measures are presented in Table 2. In three of the four perceptual parameters (grade, breathiness, and asthenia) CPP and CPPs correlated better than the other acoustic measures. This was the case for both the sustained vowel and reading.

--- Insert table 1 around here ---

--- Insert table 2 around here ---

For both sustained vowels and reading, CPP and CPPs correlated more consistently across the perceptual categories than all other acoustic measures, and in the case of B in the reading task were the only measures to show a significant relationship. The strongest correlation identified is for CPP / CPPs and overall dysphonia (G). Jitter (%), RAP and PPQ also showed a high number of significant correlations across the perceptual ratings, but their r-values are never as strong as for CPP/CPPs.

Despite the positive results for G, B and A, CPP and CPPs performed poorly in relation to roughness (R). The other parameters showed better correlations in this case, however, with the exception of PPQ, this was only true in one of the speech tasks. In addition, the correlation was relatively weak compared to some of the other relationships.

In relation to task differences, there was a higher number of significant correlations for the sustained vowel task (n = 20) than for reading (n = 15).

Differences between PD and Ataxia

A Mann-Whitney *U* test revealed no significant differences between participants with PD or ataxia in any of the acoustic or perceptual measures for both connected speech and sustained vowel samples (see Table 3).

---insert table 3 around here ---

Discussion

The purpose of this study was to apply cepstral measures and other more conventional acoustic measures of voice quality to motor speech disorders in order to investigate whether cepstral measures are more valid in determining the perceptual impression of overall dysphonia, roughness, breathiness, and asthenia patients with motor speech disorders, or more sensitive to differences between different types of disorder. In addition, the suitability of sustained vowels versus connected speech data for this analysis was examined.

This study found that cepstral measures (i.e. CPP and CPPs) correlated most consistently and most strongly with overall dysphonia, breathiness and asthenia in both sustained vowel and reading samples. Jitter (%), RAP, and PPQ were also good predictors of overall dysphonia and breathiness in sustained vowels only; elsewhere their validity was poor. The other measures, although at times statistically significant, were poor at predicting any perceptual ratings.

These findings are consistent with the results of other studies on patients with voice pathology unrelated to neurological conditions. Dejonckere and Wieneke[42] found that CPP predicted hoarseness better than other perturbation and spectral measures. Eadie and Baylor[12] found a strong correlation between CPPs and overall dysphonia on sustained vowel samples. Maryn *et al*[67] included connected speech as well as sustained vowels and found that CPPs correlated most strongly amongst their measures with overall dysphonia (G). Heman-Ackah *et al*[40] also used samples of continuous speech and sustained vowels, and expanded their list of perceptual parameters to roughness and breathiness as well as overall dysphonia. Their study also found CPPs to be the best predictor of dysphonia and breathiness, but not for roughness.

No other studies have included asthenia in their evaluation to date and no conclusions can thus be drawn about how representative the current results are in this respect. Given the strength of the correlations of CPP and CPPs with this parameter, it can be hypothesised that the two measures may be clinically useful predictors of asthenia.

The observation that none of the measures in this study were particularly well correlated with roughness is also consistent with previous studies. There is conflicting evidence as to which acoustic measure is the ideal predictor of roughness with studies showing jitter to be more strongly related than shimmer, and vice versa.[23, 25, 68, 69] The current study found that shimmer measures (shimmer % and dB, and APQ) correlated with perceptions of roughness better than jitter measures (jitter %, RAP, and PPQ) on sustained vowels, and the opposite result in reading samples.

No difference was found between the two MSDs investigated in this study, PD and ataxia. Given that perceptual studies describe differences in terms of voice quality in these populations, this may have been due to limitations of study such as the small sample size, and the relatively mild severity of voice problems in both groups. It is thus an area that would benefit from further investigation.

Finally, although the sustained vowel task rendered a higher number of significant correlations across all parameters, there was no particular difference in terms of CPP measures. They thus appear to be equally applicable to both types of speech sample. One consideration that should be taken in this respect is clinical applicability though. This study, similar to Maryn *et al*,[6766] used connected speech samples with the voiceless segments removed. This required post-recording processing, which may not be feasible in a clinical setting. However, Moers *et al*[70] evaluated the use of cepstral analyses on both sustained vowels and connected speech with all pauses and voiceless segments remaining in patients with hoarse voices. The authors found that, while correlations with perceptual measures were reduced for connected speech compared with sustained vowels, cepstral measures outperformed other perturbation measures (including those used in the current study) in both sustained vowels and continuous speech and were applicable to both.

Another clinical consideration is the applicability of the measurement procedure. This study used a standalone programme freely available on the internet, however, many clinicians will not have the technical skills to perform this analysis by these means. However, there are commercially available alternatives. Although some of the earlier systems such as the

Computerized Speech Laboratory (CSL; KayPENTAX, Prine Brook, New Jersey) were associated with problems,[39] the newly released voice analysis program (Analysis of Dysphonia in Speech and Voice (ADSV model 5109; KayPENTAX, Montvale, NJ)) successfully addressed these and thus represents a more easy to use yet equally reliable clinical interface for CPP analysis.[71]

Conclusion

This study has shown that cepstral measures, namely CPP and CPPs, remain a robust and valid predictor of overall dysphonia and breathiness in pathological voice arising from a MSD, irrespective of speech task. In addition, this study suggests cepstral measures may be clinically useful predictors of asthenia in this patient group. This is an important finding as no other studies have included asthenia in their correlations with cepstral measures and it is of particular importance with motor speech disorders such as hypokinetic dysarthria, in which a weak voice is of diagnostic importance. This finding thus merits further investigation.

Although these results are in agreement with other studies of cepstral analyses in voice disordered populations in general, improvements could be made in several areas of this study. Limitations of sample size and severity profile were already alluded to above. Ideally, groups should have been matched in size, and the severity of the voice problem should have been more wide ranging from normal through to severely disordered, which would have allowed the evaluation of the specificity of cepstral measures in differentiating severity of the voice disorder. In addition, despite the strong inter- and intrarater reliability, more raters would have given greater consensus of the perceptual parameters with which to correlate the acoustic parameters.

Despite these limitations, our study has provided an indication that cepstral analysis appears to be a valid tool in the acoustic analysis of hypokinetic and ataxic dysarthria, and should be applied in conjunction with the other measures typically used by clinicians at present. Future studies are necessary to investigate this area further, with particular emphasis on establishing a normative database and looking further into the discriminatory diagnostic value of the measures by including participants with a wider range of severity levels and different underlying pathologies than was the case in this exploratory study.

Acknowledgments

The original data used for this study had been collected as part of research grants funded by Parkinsons UK and Ataxia UK. The current analysis was supported by a Strathclyde University internship award. We are indebted to all the speakers who took part in this study, as well as to Dr Wendy Cohen, who acted as a second examiner for the perceptual voice analysis.

References

1. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J. Speech Hear. Res.* 1993;36:21-40.

2. Eadie TL, Sroka A, Wright DR, Merati A. Does Knowledge of Medical Diagnosis Bias Auditory-Perceptual Judgments of Dysphonia? *J. Voice.* 2011;25:420-429.

3. Oates JM. Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatr. Logop.* 2009;61:49-56.

4. Orlikoff R, Dejonckere P, Dembowksi J, et al. The perceived role of voice perception in clinical practice. *Phonoscope.* 1999;2:89-108.

5. Kreiman J, Gerratt BR. Measuring vocal quality. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement*. San Diego: Singular Pub. Group; 2000.

6. Hirano M. Psycho-acoustic Evaluation of Voice: GRBAS Scale for Evaluating the Hoarse Voice. *Clinical Examination of Voice*. New York: Springer-Verlag; 1981:81-84.

7. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judments. *J. Speech Hear. Res.* 1993;36:14-20.

8. Kempster GB, Gerratt BR, Abbot KV, Barkmeier-Kraemer J, Hillman RE. Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. *Am. J. Speech Lang. Pathol.* 2009;18:124-132.

9. Wuyts FL, De Bodt M, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J. Voice.* 1999;13:508-517.

10. Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA, Hoffman HT. Reliability of Clinician-Based (GRBAS and CAPE-V) and Patient-Based (V-RQOL and IPVI) Documentation of Voice Disorders. *J. Voice.* 2007;21:576-590.

11. Nemr K, Simões-Zenari M, Cordeiro GF, et al. GRBAS and CAPE-V Scaled: High Reliability and Consensus When Applied at Different Times. *J. Voice.* 2012;26:812.

12. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J. Voice.* 2006;20:527-544.

13. Bele IV. Reliability in perceptual analysis of voice quality. *J. Voice.* 2005;19:555-573.

14. Eadie TL, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. *J. Voice.* 2005;19:1-14.

15. Zraick RI, Wendel K, Smith-Olinde L. The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *J. Voice.* 2005;19:574-581.

16. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. *J. Voice.* 2009;23:341-352.

17. Herbst CT, Flitch WT, Švec JG. Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively. *J. Acoust. Soc. Am.* 2010;128:3070-3078.

18. Awan SN, Roy N. Outcomes measurement in voice disorders: application of an acoustic index of dysphonia severity. *J. Speech. Lang. Hear. Res.* 2009;52:482-499.

19. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J. Speech. Lang. Hear. Res.* 2001;44:327-339.

20. Buder EH. Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. In: Kent RD, Ball MJ, eds. *Voice quality measurement*. San Diego: Singular Publishing; 2000:119-244.

21. Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement*. Iowa City: National Center for Voice and Speech; 1995.

22. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. *J. Speech Hear. Res.* 1996;39:126-134.

23. Deal RE, Emanuel FW. Some waveform and spectral features of vowel roughness. *J. Speech Hear. Res.* 1978;21:250-264.

24. Hillenbrand J. A methodological study of perturbation and additive noise in synthetically generated voice signals. *J. Speech Hear. Res.* 1987;30:448-461.

25. Nichols AC. Jitter and shimmer related to vocal roughness: a comment on the Deal and Emanuel study. *J. Speech Hear. Res.* 1979;22:670-671.

26. Wolfe V, Fitch J, Cornell R. Acoustic prediction of severity in commonly occurring voice problems. *J. Speech Hear. Res.* 1995;38:273-279.

27. Yumoto E, Sasaki Y, Okamura H. Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J. Speech Hear. Res.* 1984;27:2-6.

28. Maryn Y, Dick C, Vandenbruaene C, Vauterin T, Jacobs T. Spectral, cepstral, and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels. *The Laryngoscope.* 2009;119:2384-2394.

29. Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin. Linguist. Phon.* 2006;20:35-49.

30. Ma EP, Yiu EM. Multiparametric evaluation of dysphonic severity. *J. Voice.* 2006;20:380-390.

31. Wuyts FL, De Bodt MS, Molenberghs G, et al. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J. Speech. Lang. Hear. Res.* 2000;43:796-809.

32. Yu P, Ouaknine M, Revis J, Giovanni A. Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements. *J. Voice.* 2001;15:529-542.

33. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J. Speech Hear. Res.* 1994;37:769-778.

34. Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P. Acoustic measurement of overall voice quality: a meta-analysis. *J. Acoust. Soc. Am.* 2009;126:2619-2634.

35. Qi Y, Hillman RE, Milstein C. The estimation of signal-to-noise ratio in continuous speech for disordered voies. *J. Acoust. Soc. Am.* 1999;20:105-117.

36. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. *J. Commun. Disord.* 1997;30:403-415; quiz 415-406.

37. Wolfe VI, Martin DP, Palmer CI. Perception of dysphonic voice quality by naive listeners. *J. Speech. Lang. Hear. Res.* 2000;43:697-705.

38. Halberstam B. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL J. Otorhinolaryngol. Relat. Spec.* 2004;66:70-73.

39. Heman-Ackah YD, Michael DD, Goding GS, Jr. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J. Voice.* 2002;16:20-27.

40. Heman-Ackah YD, Heuer RJ, Michael DD, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann. Otol. Rhinol. Laryngol.* 2003;112:324-333.

41. de Krom G. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech Hear. Res.* 1993;36:254-266.

42. Dejonckere P, Wieneke G. Cepstra of normal and pathological voices: Correlation with acoustic, aerodynamic and perceptual data. In: Ball MJ, Duckworth M, eds. *Advances in Clinical Phonetics.* Vol 6. Amsterdam: John Benjamins; 1996:217.

43. Dejonckere PH. Cepstral voice analysis: link with perception and stroboscopy. *Rev. Laryngol. Otol. Rhinol. (Bord.).* 1998;119:245-246.

44. Balasubramanium RK, Bhat JS, Fahim S, 3rd, Raju R, 3rd. Cepstral analysis of voice in unilateral adductor vocal fold palsy. *J. Voice.* 2011;25:326-329.

45. Aronson AE, Bless DM. *Clinical voice disorders.* 4th ed. New York: Thieme; 2009.

46. Ramig LO, Scherer RC, Klasner ER, Titze IR, Horii Y. Acoustic analysis of voice in amyotrophic lateral sclerosis: a longitudinal case study. *Journal of Speech and Hearing Disorders.* 1990;55:2-14.

47. Smith ME, Ramig LO, Dromey C, Perez KS, Samandari R. Intensive voice treatment in Parkinson disease: laryngostroboscopic findings. *J. Voice.* 1995;9:453-459.

48. Cooper M. *Modern techniques of vocal rehabilitation*. Springfield, Ill.,: Thomas; 1973.

49. Logemann JA, Fisher HB, Boshes B, Blonsky ER. Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *J. Speech Hear. Disord.* 1978;43:47-57.

50. Carrillo L, Ortiz KZ. [Vocal analysis (auditory - perceptual and acoustic) in dysarthrias]. *Pro-fono : revista de atualizacao cientifica.* 2007;19:381-386.

51. Holmes RJ, Oates JM, Phyland DJ, Hughes AJ. Voice characteristics in the progression of Parkinson's disease. *Int. J. Lang. Commun. Disord.* 2000;35:407-418.

52. Jimenez-Jimenez FJ, Gamboa J, Nieto A, et al. Acoustic voice analysis in untreated patients with Parkinson's disease. *Parkinsonism Relat. Disord.* 1997;3:111-116.

53. Yücetürk AV, Yilmaz H, Eğrilmez M, Karaca S. Voice analysis and videolaryngostroboscopy in patients with Parkinson's disease. *Eur. Arch. Otorhinolaryngol.* 2002;259:290-293.

54. Santos LL, Reis LO, Bassi I, et al. Acoustic and hearing-perceptual voice analysis in individuals with idiopathic Parkinson's disease in "on" and "off" stages. *Arquivos de neuro-psiquiatria.* 2010;68:706-711.

55. Gilman S, Kluin K. Speech disorders in cerebellar degeneration studied with positron emission tomography. In: Blitzer A, Brin MF, Sasaki CT, Fahn S, Harris KS, eds. *Neurologic Disorders of the Larynx*. New York: Thieme; 1992.

56. Hertrich I, Ackermann H. Auditory perceptual evaluation of rhythm-manipulated and resynthesized sentence utterances obtained from cerebellar patients and normal speakers: A preliminary report. *Clin. Linguist. Phon.* 1998;12:427-437.

57. Kent RD, Kent JF, Duffy JR, Thomas JE, Weismer G, Stuntebeck S. Ataxic dysarthria. *J. Speech. Lang. Hear. Res.* 2000;43:1275-1289.

58. Kent RD, Kim YJ. Toward an acoustic typology of motor speech disorders. *Clin. Linguist. Phon.* 2003;17:427-445.

59. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology.* 1967;17:427-442.

60. Lowit A, Kuschmann A, MacLeod JM, Shaeffler F, Mennen I. Sentence Stress in Ataxic Dysarthria - A Perceptual and Acoustic Study. *Journal of Medical Speech - Language Pathology.* 2010;18:77-82.

61. Lowit A, Dobinson C, Timmins C, Howell P, Kröger B. The effectiveness of traditional methods and altered auditory feedback in improving speech rate and intelligibility in speakers with Parkinson's disease. *International Journal of Speech-Language Pathology.* 2010;12:426-436.

62. Lowit A, Brendel B, Dobinson C, Howell P. An investigation into the influences of age, pathology and cognition on speech production. *Journal of Medical Speech - Language Pathology.* 2006;14:253-262.

63. Grabe E, Post B, Nolan F. The IViE Corpus: Department of Linguistics, University of Cambridge; 2001.

64. Maryn Y, Corthals P, De Bodt M, Van Cauwenberge P, Deliyski D. Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat. *Folia Phoniatr. Logop.* 2009;61:217-226.

65. Midi L, Dogan M, Koseoglu M, Can G, Sehitoglu MA, Gunal DL. Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. *Acta Neurologica Scandinavica.* 2008;117:26-34.

66. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J. Speech Hear. Res.* 1996;39:311-321.

67. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J. Voice.* 2010;24:540-555.

68. Klingholz F, Martin F. Quantitative spectral evaluation of shimmer and jitter. *J. Speech Hear. Res.* 1985;28:169-174.

69. Wolfe VI, Steinfatt TM. Prediction of vocal severity within and across voice types. *J. Speech Hear. Res.* 1987;30:230-240.

70. Moers C, Mobius B, Rosanowski F, Noth E, Eysholdt U, Haderlein T. Vowel- and text-based cepstral analysis of chronic hoarseness. *J. Voice.* 2012;26:416-424.

71. Peterson EA, Roy N, Awan SN, Merrill RM, Banks R, Tanner K. Toward Validation of the Cepstral Spectral Index of Dysphonia (CSID) as an Objective Treatment Outcomes Measure. *J. Voice.* 2013;27:401-410.

Table 1 – Descriptive Statistics for each acoustic measure for both sustained vowels and reading passages.

	Sustained Vowel				Reading			
	Mean	**Std Dev**	**Max**	**Min**	**Mean**	**Std Dev**	**Max**	**Min**
CPP	13.1	2.88	20.98	9.45	12.37	1.66	15.29	9.52
CPPs	5.62	2.24	9.26	1.68	4.73	1.31	6.95	2.36
Jitter (%)	3.39	3.05	21.33	0.63	5.38	1.13	8.96	3.06
RAP	2.00	1.79	12.75	0.37	3.00	0.64	5.06	1.77
PPQ	2.16	1.96	13.59	0.37	3.55	0.76	6.16	2.02
Shimmer (%)	8.28	4.55	31.56	2.88	14.42	3.07	23.29	8.59
Shimmer (dB)	0.93	1.18	8.78	0.26	1.53	0.32	2.46	0.79
APQ	6.65	3.25	21.27	2.25	15.86	3.41	26.10	9.27
HNR	14.75	4.27	24.48	6.73	12.04	2.56	17.65	7.72

Table 2. Correlations between acoustic measures and perceptual ratings.

	Sustained Vowel				Reading			
	G	**R**	**B**	**A**	**G**	**R**	**B**	**A**
CPP	-0.86**	NS	-0.77**	-0.52**	-0.54**	NS	-0.42**	-0.47**
CPPs	-0.88**	NS	-0.74**	-0.54**	-0.53**	-0.35*	-0.38*	-0.47**
Jitter (%)	0.72**	NS	0.60**	0.38**	0.31*	0.34*	NS	0.29*
RAP	0.73**	NS	0.60**	0.38**	0.31*	0.37**	NS	0.29*
PPQ	0.67**	0.35*	0.52**	0.36**	0.28*	0.31*	NS	NS
Shimmer (%)	0.30*	0.35*	NS	NS	NS	NS	NS	NS
Shimmer (dB)	NS	0.43**	NS	NS	NS	NS	NS	NS
APQ	NS	0.36**	NS	NS	NS	NS	NS	NS
HNR	-0.40**	NS	-0.32*	NS	NS	NS	NS	NS

* $p < 0.05$

** $p < 0.01$

NS Not significant (i.e. $p > 0.05$)

Table 3. p-values for Mann-Whitney *U* test comparing PD and ataxia participants across all measures.

Acoustic Measures	Sustained Vowel	Reading	Perceptual Measures	Sustained Vowel	Reading
CPP	0.574	0.574	G	0.442	0.065
CPPs	0.916	0.798	R	0.721	0.161
Jitter (%)	0.959	0.083	B	0.279	0.645
RAP	0.916	0.130	A	0.721	0.442
PPQ	0.721	0.234			
Shimmer (%)	0.279	0.798			
Shimmer (dB)	0.208	0.753			
APQ	0.195	0.529			
HNR	0.345	0.382			