Estimating Dysphonia Severity in Continuous Speech: Application of Cepstral Analysis

Shaheen N. Awan¹, Nelson Roy², & Christopher Dromey³

¹Bloomsburg University of PA; ²University of Utah; ³Brigham Young University

Introduction: Previous studies with both sustained vowel and continuous speech samples have indicated that measures of the cepstral peak may be strong indicators of dysphonia and indices of dysphonia severity. In addition, several studies have indicated that cepstral measures may be supplemented by other acoustic measures (such as ratios of low vs. high frequency spectral energy) to account for larger degrees of variance between perceived and acoustically predicted severity ratings. In the clinical domain, voice clinicians need a robust, fairly automatic, measure of dysphonia severity in continuous speech that relates well to listener ratings, and is sensitive to diverse voice qualities as well as to improvements following management. In this regard, the purpose of this study was twofold: (1) to assess the strength of the cepstral peak prominence (CPP) as a correlate of perceived severity using pre and post-treatment samples of continuous speech obtained from a wide variety of voice quality types and severities; and, (2) to determine whether additional measures obtained from frequency-based spectral/cepstral analyses may increase the strength of the association between listener-perceived and acoustically predicted severity ratings. In particular, this study used stepwise linear regression analyses to identify a subset of spectral/cepstral-based methods which could be used to most effectively predict vocal severity as estimated via auditory-perceptual analysis in a set of pre vs. post-treatment voice samples.

Method: Pre- and post treatment continuous speech samples were obtained from 104 females with primary muscle tension dysphonia (208 voice samples in total; mean participant age = 46.4 yrs; SD = 13.7). As part of a standard clinical test battery, each participant was asked to read “The Rainbow Passage” (Fairbanks, 1960) at a comfortable pitch and loudness. Voice samples were recorded using a Shure Prologue Model 14H Dynamic microphone and digitized at 25 kHz and 16 bits of resolution using the Computerized Speech Lab (CSL Model 4300, Kay Elemetrics, Pinebrook, NJ, USA). At a later time, the speech samples were edited to include only the 2nd and 3rd sentences.

All samples were analyzed using a Windows-based computer program developed by the first author. The program included an implementation of spectral and cepstral analysis methods described by Hillenbrand and Houde(1996) and Awan and Roy (2005, 2006, 2009) that have been used to characterize voice quality type and predict dysphonia severity in normal and disordered voice samples. Unlike the computer algorithms employed in previous studies by Awan and Roy (2005, 2006, 2009) which combined spectral/cepstral based acoustic methods with time-based measures such as shimmer and pitch sigma, the computer algorithms used in this study are solely spectral based and do not depend upon accurate identification of cycle boundaries for any of the measurements obtained.

Five master’s degree students in communication disorders rated the pre and post-treatment continuous speech samples (the 2nd and 3rd sentences of “The Rainbow Passage”) via headphones at self-selected comfortable loudness levels. All samples were presented in randomized order for each judge. Judges were asked to rate the severity of voice quality using a custom MATLAB routine which allowed the user to adjust an on screen slider to a particular point on a 100 point visual analog scale (VAS). One end of the scale was labeled as “normal” and the other end as “profoundly abnormal”, with
higher numbers reflecting increased severity of dysphonia. Inter-judge reliability for the dysphonia severity ratings was strong (Average Measures ICC = 0.97; Single Measures ICC = 0.87).

**Results:** Stepwise linear regression produced a three-factor model consisting of the CPP, the mean ratio of low-to-high frequency spectral energy (DFTR), and the standard deviation of the ratio of low-to-high frequency spectral energy (DFTR SD) that was strongly correlated with perceived dysphonia severity ratings ($R = 0.85; R^2 = 0.73$). The resulting stepwise multiple regression equation was used compute a predicted dysphonia severity for each voice sample. The mean predicted dysphonia severity rating across all 208 voice samples was 41.77 (SD = 30.00) vs. the mean perceived severity rating of 41.78 (SD = 35.20). A paired $t$-test indicated no significant difference between mean predicted vs. mean perceived dysphonia severity ratings ($t = -0.01; df = 207; p = 0.97$).

Predicted dysphonia severity ratings were also computed for all 104 pre-treatment voice samples. Results showed a mean pre-treatment predicted dysphonia severity rating of 63.47 (SD = 27.02) vs. the mean pre-treatment perceived rating of 69.27 (SD = 26.83). In a similar fashion, post-treatment severity ratings were computed for the 104 post-treatment samples. Results showed a mean post-treatment predicted dysphonia severity rating of 20.07 (SD = 11.31) vs. the mean post-treatment perceived rating of 14.28 (SD = 15.61). While $t$-tests between predicted vs. perceived ratings were significant for both pre- and post-treatment voices, the mean differences were < 6 pts. on the 100 pt. VAS.

**Conclusions:** The application of spectral/cepstral measures via a multiparameter model as described in this study shows great promise as an objective measure of dysphonia severity in continuous speech samples. The results of this study indicate that strong predictions of listener perceived dysphonia severity can be achieved across the range of vocal quality disruption observed in pre and post-treatment speech samples of primary MTD. While the CPP as a single measure is a very strong correlate of perceived dysphonia severity ($r = -0.81$) accounting for approximately 66% of the shared variance between listener perceived and acoustically predicted dysphonia severity, the addition of other measures obtained via the same spectral/cepstral analysis procedures is able to significantly increase the amount of shared variance to 73%. Since all of these combined measures can be obtained efficiently via a common core of spectral/cepstral analysis procedures, it appears to be advisable to combine measures of the CPP with measures such as DFTR and the DFTR SD to strengthen dysphonia predictions across a wide range of dysphonia severities and types.

**References**


