Smoothed Cepstrum Calculation Using Stressed Vowels in Connected Speech to Objectively Measure Dysphonia

K. M. Murray, R. T. Sataloff and Y. Heman-Ackah
Department of Otolaryngology/Head and Neck Surgery, Drexel University College of Medicine, Philadelphia, PA

Voice disorders are rather common: the NIDCD reports that at least 7.5 million people in the United States have vocal problems [1]. To address this problem, voice scientists have developed various objective vocal health measurements [3-6]. To further study those measurements, vocal health practitioners have collected and organized speech databases [3,4,6]. One standard for evaluation of these measurements is the GRBAS Scale [7]. This scale achieves strong perceptual agreement among medical professionals with respect to a speaker's vocal health. In broad terms, dysphonia represents a marked difficulty in producing voiced phonation. As such, traditional voicing detection algorithms can fail on dysphonic voices [5].

As elaborated in [5], a spectral analysis of a signal records the intensity level of its individual frequency contributions. A cepstral analysis results from applying the Fourier transform again, and taking a logarithm of its output. A time-based "quefrency" is the resulting unit of analysis. As a breathy or distorted voice does not have a prominent cepstral peak, it is useful to create a regression line over the entire cepstrum for comparison. A small difference between the regression line's predicted value, and the cepstral peak's value, can correlate to breathiness and/or distortion. The smoothed cepstrum, or CPPS, is the regression line difference taken over an average, calculated across a number of timeframes and quefrencies.

An analysis of the voiced sentence, "Marvin Williams is only nine," was reported in [3] for 281 speakers. SpeechTool, as described in [5], was used to obtain 87% sensitivity, and 90% specificity, when evaluated with respect to the GRBAS perceptual rating of "Grade." The first author had access to only 277 out of the original 281 samples for analysis, and Table 1 displays the near replication of those results in Line 1. The rest of the entries of Table 1 report results on this database.

Line 2 of Table 1 displays results from using the Praat script [2] published in Appendix 1 of [6], without applying any silence or voicing decisions. The parameters of 0.002, 0.001, and 0.01 seconds correspond to frame size, and quefrency/time smoothing window values, respectively. Line 3 of Table 1 shows results from editing the Praat script [6] to use 0.001, 0.003, and 0.30 seconds for the frame size, and quefrency/time-smoothing window values. [5] advises using those window sizes for analyzing vowels, and they are used in lines 3-5, Table 1.

Description	Sensitivity	Specificity
Nearly replicated results	87%	91%
Original Praat script	91%	87%
Vowel-Based smoothing	89%	93%
Mean CPPs of four stressed vowels	94%	90%
CPPS of four concatenated stressed vowels	91%	92%

Table 1. Connected Speech Results, 277 Samples

The center point of each lexically stressed vowel in the following words, "Marvin," "Williams", "only," and "nine" was hand-labeled in [8]. A 0.18 second window centered at each of those points was created, and the CPPS was calculated for each resulting vocalic interval. Those four CPPS values were averaged to create the results in line 4 in Table 1. Lastly, the four vowels were concatenated together, and a CPPS calculation was made on the resulting file. Line 5 of Table 1 shows those results. It is possible that high frequency periodic "noise" is in the speech of some dysphonic speakers, and can artificially raise the CPPS. Spurious high frequency periodic sounds might not strongly manifest in lexically stressed vowels, even in dysphonic speech. Lexically stressed vowels are of sufficient duration to test for phonetic effects on the CPPS. Further study will continue to examine the relationships among phonetics, CPPS, and dysphonic voices.

REFERENCES

[1] "Statistics on Voice, Speech, and Language." Internet: http://www.nidcd.nih.gov/health/statistics/pages/vsl.aspx, June 7, 2010 [Sept. 30, 2015].

[2] Boersma, Paul & Weenink, David. "Praat: doing phonetics by computer." [Computer program]. Version 5.4.20, http://www.praat.org/ September 26, 2015 [September 28, 2015].

[3] Heman-Ackah, Y.D., Heuer, R.J., Michael, D.D., Ostrowski, R., Horman, M., Baroody, M.M., Hillenbrand, J., and Sataloff, R.T.: 'Cepstral peak prominence: a more reliable measure of dysphonia', The Annals of otology, rhinology, and laryngology, 2003, 112, (4), pp. 324-333.

[4] Heman-Ackah, Y.D., Sataloff, R.T., Laureyns, G., Lurie, D., Michael, D.D., Heuer, R., Rubin, A., Eller, R., Chandran, S., Abaza, M., Lyons, K., Divi, V., Lott, J., Johnson, J., and Hillenbrand, J.: 'Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia', Journal of Voice, 2014, 28, (6), pp. 783-788

[5] Hillenbrand, J., and Houde, R.A.: 'Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech', Journal of Speech and Hearing Research, 1996, 39, (2), pp. 311-321

[6] Maryn, Y., and Weenink, D.: 'Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index', Journal of Voice, 2015, 29, (1), pp. 35-43

[7] Nemr, K., Simões-Zenari, M., Cordeiro, G.F., Tsuji, D., Ogawa, A.I., Ubrig, M.T., and Menezes, M.H.M.: 'GRBAS and Cape-V Scales: High Reliability and Consensus When Applied at Different Times', Journal of Voice, 2012, 26, (6), pp. 812.e817-812.e822

[8] Murray, K., Graff, D., Heman-Ackah, Y.D.,Sataloff,.R.T.: 'Evaluation of a simplified phonetic annotation scheme for disordered speech'. Proc. The 155th Meeting of the Acoustical Society of America (ASA), Paris, France, Jul. 2008.

SMOOTHED CEPSTRUM CALCULATION USING STRESSED VOWELS IN CONNECTED SPEECH TO OBJECTIVELY MEASURE DYSPHONIA

K. M. Murray, R. T. Sataloff and Y. Heman-Ackah
Department of Otolaryngology/Head and Neck Surgery, Drexel University College of Medicine, Philadelphia, PA

Summary

- Researchers have been attempting to automatically replicate perceptual auditory evaluations by trained clinicians of disordered and healthy speech.
- The most widely used measure is called the Smoothed Cepstrum, or CPPS.
- The CPPS is a robust measure without reliance on voicing and silence decisions.
- Until recently, the calculation of the CPPS has *not* been integrated with a phonetic analysis tool like Praat [6].
- Improved results have been obtained from using
  - Recommended smoothing parameters from [5] when calculating the CPPS.
  - Analyzing just the lexically stressed vowel portions of the utterances
- GRBAS Grade ~ overall dysphonia, and ranges from 1-100
- Higher values correspond to more dysphonia
- 3-10 professional raters rated each speech token, ratings values were averaged to produce a number from 1-100

Classifications:

- 1-33 ~ Mild, 163 total
- 34-67 ~ Moderate, 60 total
- 88-100 ~ Dysphonic, 54 total

"Marvin Williams is only nine", Voiced sentence, Recorded at 25kHz sampling rate

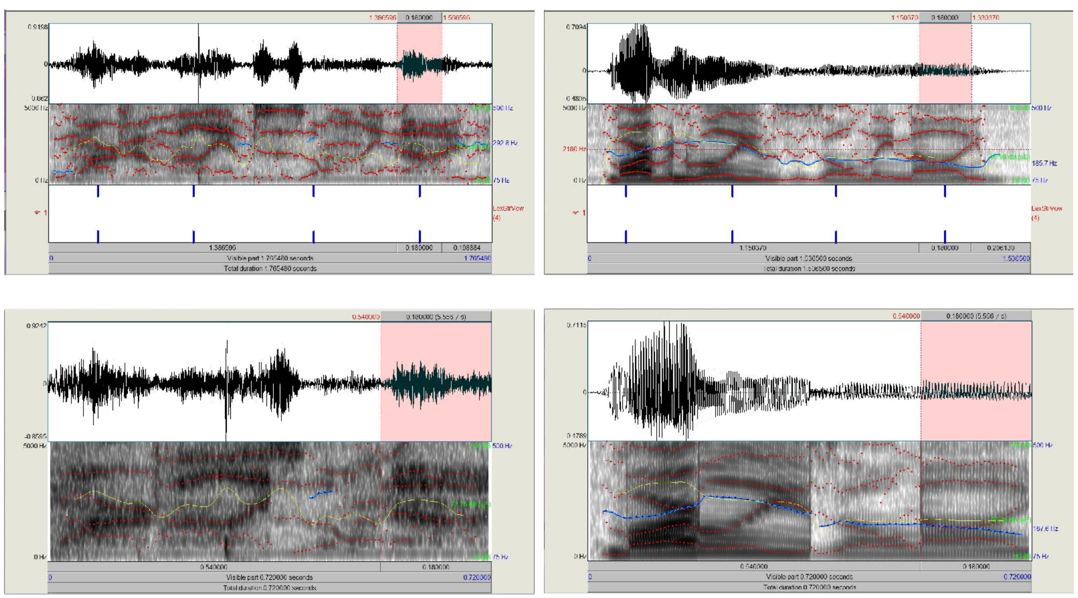

Figure 1: Whole File/4 Vowels, Top/Bottom, Dysphonic Speaker on Left, Non-Dysphonic Speaker on Right

Dysphonia and the CPPS

- Dysphonia represents a marked difficulty in producing voiced phonation.
- A spectral analysis of a signal records the intensity level of its individual frequency contributions.
- A cepstral analysis results from applying the Fourier transform again, and taking a logarithm of its output.
  - A time-based "quefrency" is the resulting unit of analysis.
- Create a regression line over a smoothed Cepstrum for comparison.
- A small difference between the regression line's predicted value, and the cepstral peak's value, can correlate to breathiness and/or distortion.

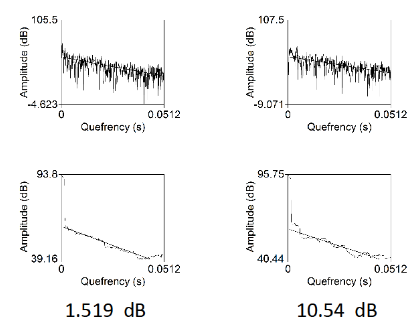

Figure 2: Raw Cepstral Plot/Smoothed Cepstral Plot, Top/Bottom Dysphonic Speaker on Left, Non-Dysphonic Speaker on Right, CPPS calculation is shown with respect to the 4 concatenated stressed vowels

Description	Sensitivty	Specificity	Pospredict	Negpredict
Nearly replicated results	87%	91%	.54	.78
Original Praat script	91%	87%	.79	.80
Vowel-Based smoothing	89%	93%	.86	.77
Mean CPPs of four stressed vowels	94%	90%	.84	.82
CPPS of four concatenated stressed vowels	91%	92%	.86	.82

Table 1: Results