Application of Laryngeal Landmarks for Characterization of Dysphonic Speech

Keiko Ishikawa[1], Joel MacAuslan[2] and Suzanne Boyce[1]

University of Cincinnati, College of Allied Health, Dpt. of Communication Sciences and Disorders[1]; Speech Technology & Applied Research[2]

Introduction

“People don’t understand me in noisy places” is one of the most commonly reported concerns among individuals with dysphonia.[1-2] Dysphonia is often a result of laryngeal pathology, which elicits greater aperiodicity and instability in a speech signal. These acoustic abnormalities likely contribute to the intelligibility deficit reported by these individuals.

Acoustic analysis is commonly used in dysphonia evaluation. Multiple algorithms are available for characterizing the degree of aperiodicity in speech.[3] Typically, the degree of aperiodicity is measured over a particular length of voicing or speech selected by a user. While such algorithms are effective for describing degree of dysphonic voice quality perceived by listeners, an algorithm that describes timing and frequency of aperiodic moments may provide information more relevant to intelligibility.

The landmark (LM)-based analysis is a linguistically-motivated algorithm based on the landmark theory of speech perception and production.[4] It characterizes speech signals with LMs, which mark moments of acoustic change that are elicited by laryngeal and vocal tract events. SpeechMark® is a semi-automatic, LM-based analysis tool.[5-6] Its repertoire of LM detection algorithms include several that detect offset and onset of voicing moments through cepstral analysis.[7] This study examined the utility of three different laryngeal LM types for differentiating normal and dysphonic speech signals using SpeechMark®. The following markers were examined: [g] (glottal) and [p] (periodicity), each of which detects a sudden change in periodicity (onset and offset) but use different acoustic rules; and [j] (jump), which detects onset and offset of abrupt F0 change. (See speechmrk.com for rule description.)

Hypothesis: Acoustic differences between normal and dysphonic speech could be described by [g]. [p] and [j] LMs.

Results

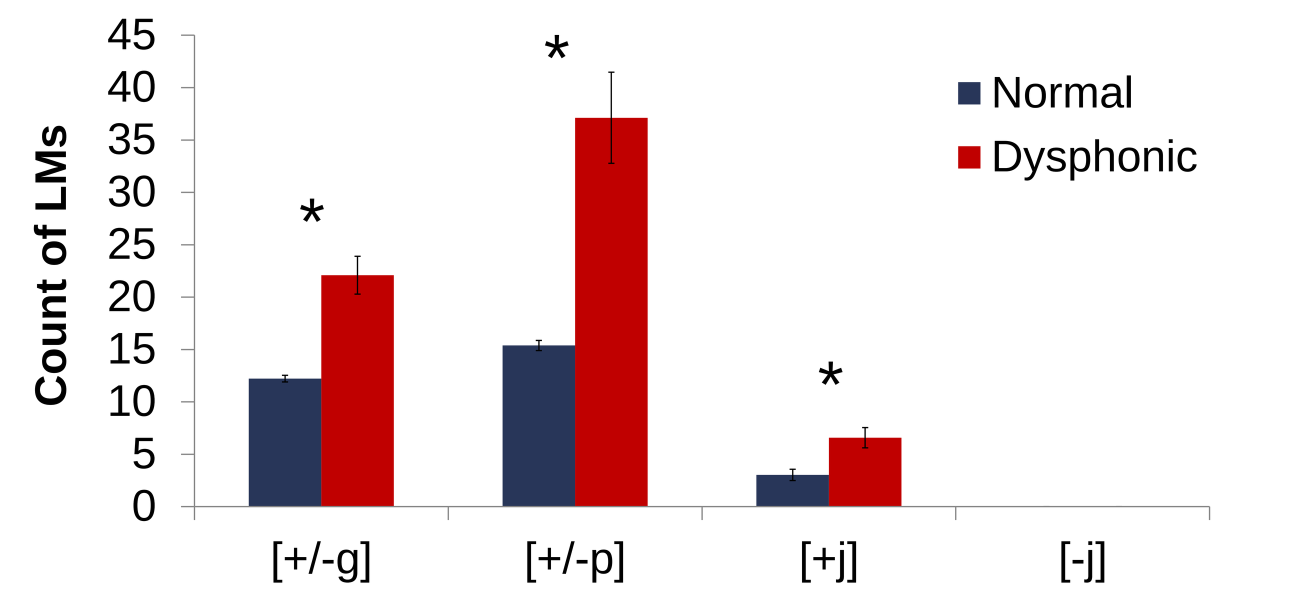

Figure 2. Average number of LMs in normal and dysphonic speech samples. Error bars indicate standard error. Asterisks show significant between-group difference indicated by t-test with Bonferroni correction at 0.008.

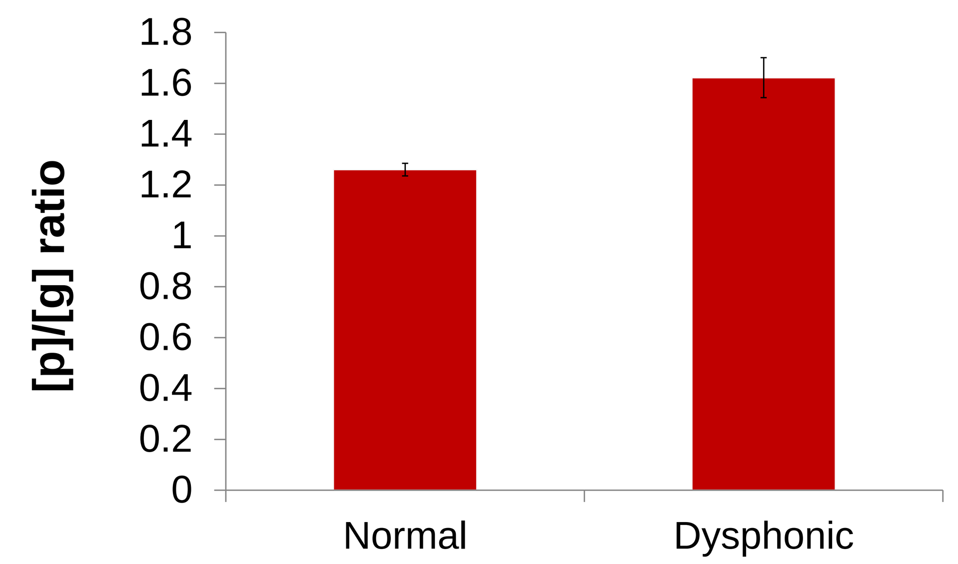

Figure 3. Average ratio between the number of [p] and [g] LMs for normal and dysphonic groups (1.26 and 1.62, respectively). Error bars indicate standard error. The between-group difference was significant, indicated by t-test with significance level of p << 0.001

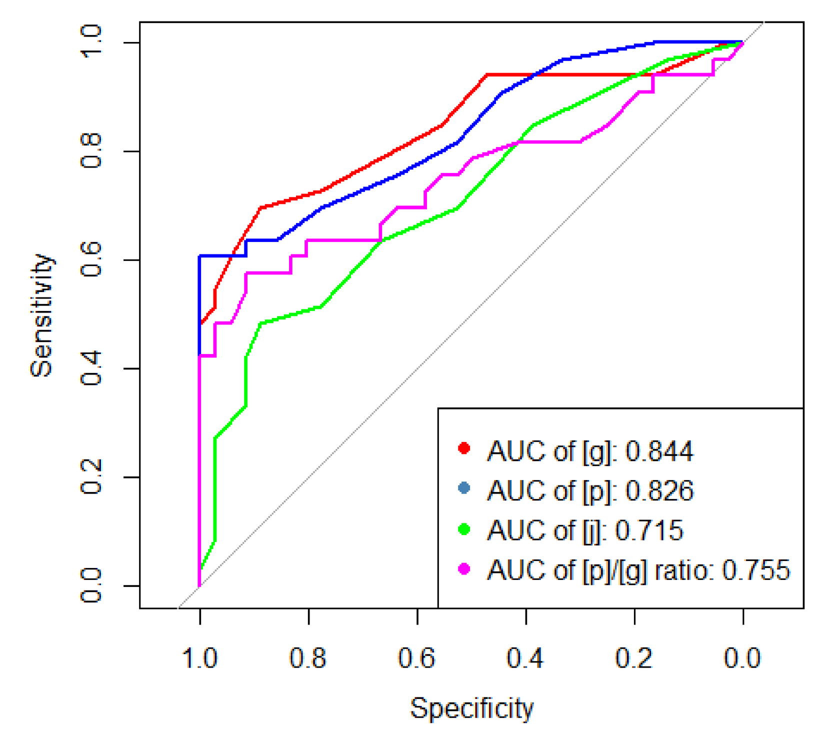

Figure 4. Receiver Operation Characteristics (ROC) curves* of [g], [p], [j] LMs and [p]/[g] ratio. Their areas under the curve (AUC) are greater than 0.5, which rejects the hypothesis, H_0: AUC = 0.5, indicating that they differentiate two groups at greater than a chance.

LM	AUC	95% CI (low)	95% CI (high)	Sensitivity	Specificity
[g]	0.844	0.749	0.939	0.697	0.889
[p]	0.837	0.743	0.931	0.636	0.917
[j]	0.715	0.593	0.836	0.636	0.667
[p]/[g]	0.755	0.629	0.872	0.636	0.805

Table 1. The 95% confidence intervals of the LMs and their sensitivity and specificity.

LM	z	p-value
[g] vs. [p]	0.274	0.784
[g] vs. [j]	1.886	0.059
[p/g] ratio vs. [j]	0.549	0.583

Table 2: DeLong's test for two ROC curves showed that the performance of [g] and [p] LMs for differentiating dysphonic from normal speech is not significantly different (p = 0.784). The performance of [g] and [j], and [p]/[g] ratio and [j] was also not significantly different (p=0.059, 0.583, respectively).

* The ROC curve was originally designed for analysis of radar signals. The ability of radar receiver operators for correctly detecting an aircraft was characterized with this technique. Today, this technique is widely used in biomedical research for evaluating accuracy of diagnostic tests, which serve as the “receiver” of a signal emitted by a disease.

Methods

- **Speakers:** 33 normal and 36 dysphonic speakers selected from the Kay Disordered Voice Database. Dysphonic speakers. Dysphonic speakers were judged to have moderate to severe dysphonia by two speech-language pathologists who specialize in the care of dysphonia.
- **Speech Material:** First sentence of the Rainbow passage
- **Acoustic Analysis:** SpeechMark® MATLAB Toolbox Version 1.0.3
  - [p] LMs indicate periodic regions, which are determined by the presence of high cepstral peaks.
  - [g] LMs indicate voiced regions, which are determined by the following rules:
    - A region has high HNR (i.e. high periodicity, high Cepstral Peak Prominence)
    - A region is adjacent within 50 ms to a region of high HNR and has similar or higher power and/or similar spectral tilt to the high-HNR region
  - Onset and offset of these events are denoted by + and – signs.
  - [j] LMs are determined by abrupt upward (+) or downward (-) jumps in F0 by at least 0.1 octave.

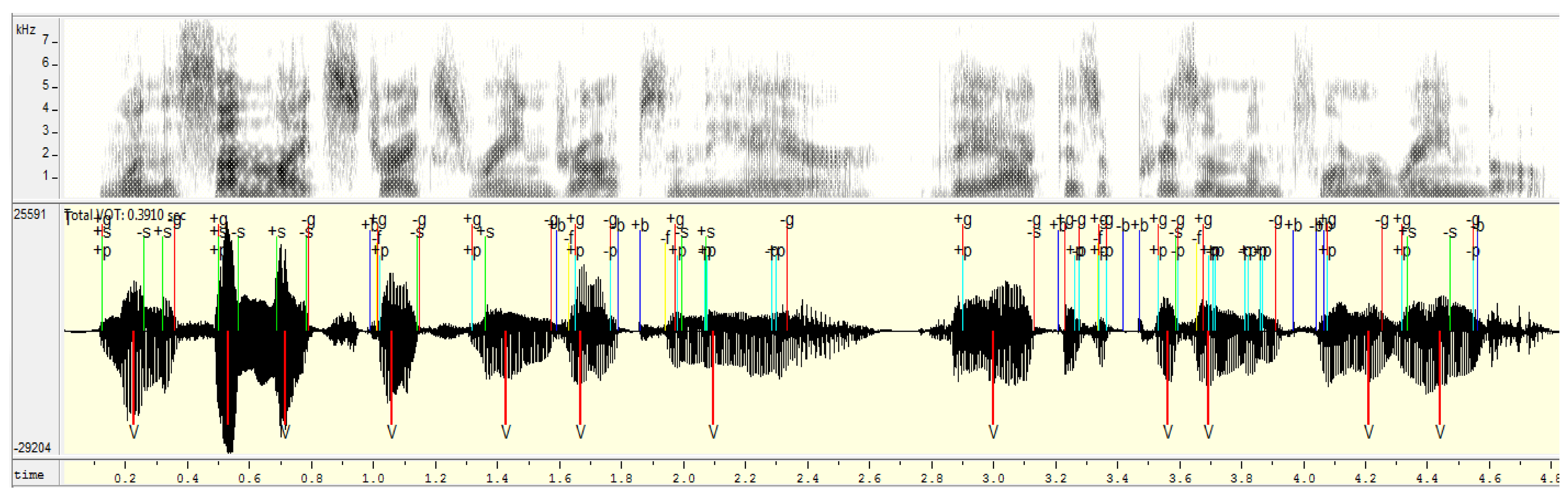

Figure 1: An example of LM analysis by SpeechMark®

Discussions & Conclusions

Results of the study showed that the semi-automatic LM-based analysis tool can reliably differentiate dysphonic speech from normal speech based on the laryngeal landmarks, [g], [p] and [+j]. Although there was no significant difference between these LMs in their ability to identify dysphonic speech, specificity was considerably higher for [g] and [p] than [+j]. For [g] and [p], their high specificity values suggest that these LMs are biomarkers that rarely mistake normal speech for dysphonic speech.

Dysphonic speech generated the greater number of the laryngeal LMs than normal speech, indicating the greater aperiodicity and interruption in voicing. Since periodic regions (between [+p] and [-p] LMs) can only lie within voiced regions (between [+g] and [-g] LMs), the number of [p]s are always equal to or more than the number of [g]s. The [p]/[g] ratio was higher in dysphonic group than in normal group (Figure 3). The ratio of 1 indicates no periodicity breaks occurred in voiced region. The mean number of periodicity breaks was 62% for dysphonic group and 26% for normal group, indicating that ***dysphonic speakers produced over twice as many periodicity breaks per voiced interval as normal speakers did,*** *a result which may extend to much more general speech materials.* This result confirms the greater aperiodicity in dysphonic voice reported in the literature, and these periodicity breaks could result in reduced intelligibility.

Surprisingly, neither normal nor dysphonic speech generated [-j]. This result implies that downward jumps in F0 do not occur even after [+j] at least in the speech material analyzed for this study. As the sample size of this study is relatively small, whether this finding is generalizable to all types of speech needs to be examined further.

CONCLUSION:
Semi-automatic LM-based analysis is a viable option for describing acoustic abnormalities in dysphonic speech.

Future directions:
- Examine correlation between these LMs and intelligibility measurements.
- Compare performance of these LMs with traditional periodicity measures such as HNR.

References

1. Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The voice handicap index (VHI) development and validation. *American Journal of Speech-Language Pathology*, 6(3), 66-70.
2. Ishikawa, K., Boyce, S., Kelchner, L., Golla Powell, M., Schieve, H., de Alarcon, A., Khosla, S. (in press) The Effect of Background Noise on Intelligibility of Dysphonic Speech, *Journal of Speech Language Hearing Research*.
3. Hillenbrand, J. (2011). Acoustic Analysis of Voice: A Tutorial. *SIG 5 Perspectives on Speech Science and Orofacial Disorders, 21*(2), 31-43. doi: 10.1044/ssod21.2.31
4. Boyce, S., Fell, H. J., & MacAuslan, J. (2012). *SpeechMark: Landmark Detection Tool for Speech Analysis.* Paper presented at the INTERSPEECH.
5. Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America, 111*(4), 1872-1891.
6. Boyce, S., Krause, J., Hamilton, S., Smiljanic, R., Bradlow, A. R., Rivera-Campos, A., & MacAuslan, J. (2013). *Using landmark detection to measure effective clear speech.* Paper presented at the Proceedings of Meetings on Acoustics.
7. Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, *37*(4), 769-778.

Acknowledgment

This work was funded by United States National Institutes of Health (NIH)grants R43 and R44 DC010104, R42 AG033523, and R42 HD34686 to S.T.A.R. Corp. For more information and SpeechMark downloads (currently freeware in beta version): www.speechmrk.com