A Comparative Study of Variability in Landmark Sequences and Implications for Dysphonic Speech Analysis

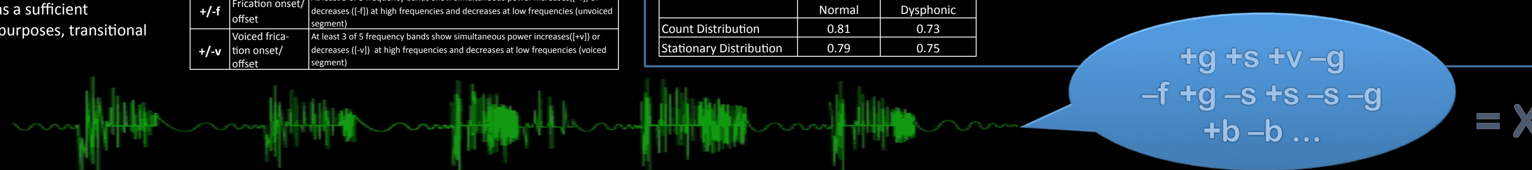
Keiko Ishikawa[1], Marepalli B. Rao[2] and Suzanne Boyce[1]

University of Cincinnati, College of Allied Health, Dpt. of Communication Sciences[1] and Disorders; 2. College of Medicine, Dpt . of Environmental Health, Div. of Biostatistics and Bioinformatics[2]

Introduction

Dysphonia negatively affects speech intelligibility especially in the presence of background noise; however, no clinical tool exists to measure this deficit. Landmark (LM) analysis may serve as the basis of such tool. The analysis identifies characteristic patterns of abrupt changes in the speech signal over time, and assigns them particular "landmarks." Consequently, it describes speech as a sequence of LMs.

Dysphonic speech often contains greater noise and/or less harmonic power, potentially affecting expression of the LMs. Past studies successfully differentiated disordered speech from normal speech based on the number of times each LM occurs. While the count was a sufficient measure for these purposes, transitional patterns in LM sequences may yield more descriptive information on the underlying mechanism of the intelligibility deficits.



Because words and sentences impose a constraint on the order of sounds, we can expect that the LM sequence will have patterns that are particular to the form of the utterance. Information theory and Markov chain model are classic statistical approaches for characterizing a sequence of events.

Information theory is concerned about measuring the extent of chaos present in a random environment. Measuring the degree of chaos present in LM patterns in speech signal may be a way to understand and characterize abnormalities in speech. Entropy is a tool to measure the degree of chaos.

Markov chain models are used to explain time evolution of random phenomenon. The models are appropriate for understanding transition of one LM into another. The transition sequence data of LMs for a subject can be utilized to estimate the 1-step probability transition matrix. Using the transition matrix one can determine the stationary probability distribution of LMs, which summarizes the long-term frequency pattern of LMs.

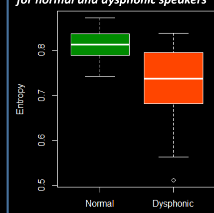
The purpose of this study was to characterize differences in LM sequence between normal and dysphonic speech based on their entropy of count distribution and entropy of stationary distribution.

Ten LMs shown below were examined:

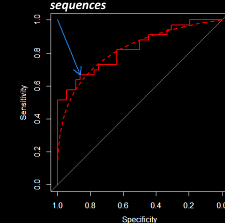
Symbol	Mnemonic	Rule
+/-g	Glottal onset/ offset	[+g]: Beginning of sustained vocal fold vibration, i.e., of periodicity or of power and spectral slope similar to that of a nearby segment of sustained periodicity. [-g]: End of sustained vocal fold vibration
+/-b	Burst onset/ offset	At least 3 of 5 frequency bands show simultaneous power increases([+b]) or decreases ([-b]) of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in an unvoiced segment (not between +g and the next -g)
+/-s	Syllabic onset/ offset	At least 3 of 5 frequency bands show simultaneous power increases([+s]) or decreases ([-s]) of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in a voiced segment (between +g and the next -g)
+/-f	Frication onset/ offset	At least 3 of 5 frequency bands show simultaneous power increases([+f]) or decreases ([-f]) at high frequencies and decreases at low frequencies (unvoiced segment)
+/-v	Voiced frica-tion onset/ offset	At least 3 of 5 frequency bands show simultaneous power increases([+v]) or decreases ([-v]) at high frequencies and decreases at low frequencies (voiced segment)

Results

Boxplots for the entropy of LM sequence for normal and dysphonic speakers



ROC curve for the entropy in LM sequences

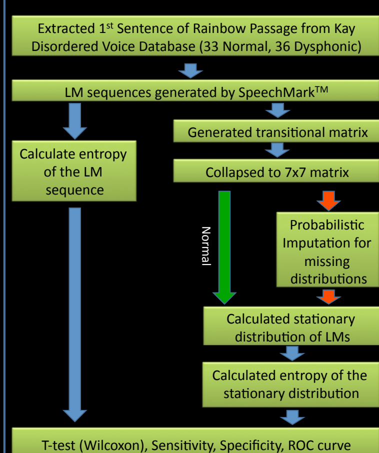


Means of Entropy

	Normal	Dysphonic
Count Distribution	0.81	0.73
Stationary Distribution	0.79	0.75

- Even after the collapse, 11 out of 36 dysphonic speakers were missing the grouped LMs. None of the normal speakers were missing the grouped LMs.
- The entropies calculated directly from the LM sequence and the stationary distribution of the LMs were similar.
- Wilcoxon rank sum test indicated that the difference was statistically significant ($p < 0.01$).
- Area under the ROC curve was 0.83, which rejects the hypothesis, H_0: AUC = 0.5.
- A diagnostic test emerges from the ROC curve analysis:
  - Test is positive (indicating dysphonia) if Entropy ≤ 0.777
  - Test is negative (indicating normal) if Entropy > 0.777
  - Sensitivity of the test = 0.67; specificity of the test = 0.86

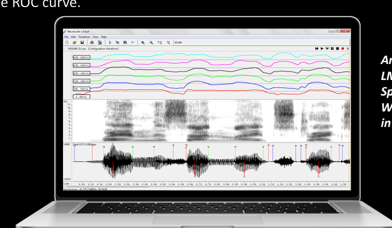
Methods



- **Speakers:** Dysphonic speakers had a diagnosis of vocal fold mass or vocal fold paralysis. Dysphonia severity was greater than moderate, confirmed by 2 speech pathologists.
- **Statistics:** All computations were done with R software. Entropy was calculated as:

$$H = -\sum_i p_i \log_{10} p_i$$

- For the Markov Chain model, the LMs [+f], [-f], [+v], and/or [-v] were eventually grouped and the matrices were collapsed to 7x7 because 27/36 normal and 30/33 dysphonic speakers did not express these LMs.
- The pROC package was used for the ROC curve analysis [5]. Sensitivity and specificity of the entropy was determined by the minimum Euclidean distance between the point (1, 1), which indicates an ideal sensitivity and specificity, and the ROC curve.



An example of LM analysis by SpeechMark WaveSurfer plug-in version

Conclusion

The results of the study indicated that the variability in LM sequence is a useful biomarker for detecting acoustical difference between normal and dysphonic speech. The analyses revealed that patterns in LM sequences for dysphonic speech are significantly less variable (i.e. more predictable) in comparison to normal speech. This finding suggests that intelligibility deficits in dysphonic speech may be due to the masking of segmental information inherent to dysphonic speech. The entropies for count distribution and stationary distribution of LMs were very close, which implies that examining one-step transition does not illustrate the difference in LM sequences better than examining count distribution of LMs alone.

References

[1] K. N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features," *The Journal of the Acoustical Society of America,* vol. 111, pp. 1872-1891, 2002.
[2] E. Yumoto, W. J. Gould, and T. Baer, "Harmonics-to-noise ratio as an index of the degree of hoarseness," *The Journal of the Acoustical Society of America,* vol. 71, pp. 1544-1550, 1982.
[3] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech, Language, and Hearing Research,* vol. 37, pp. 769-778, 1994.
[4] S. Boyce, M. Speights, K. Ishikawa, and J. MacAuslan, "Speechmark acoustic landmark tool: application to voice pathology," in *INTERSPEECH*, 2013, pp. 2672-2674.
[5] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC bioinformatics,* vol. 12, p. 77, 2011.

Acknowledgment

This work was funded by grants R43 and R44 DC010104, R42 AG033523, R21 HL086689, and R42 HD34686 from the United States National Institutes of Health.