Landmark-based approach for automatically describing the effect of spasmodic dysphonia on speech production: Preliminary case studies

Keiko Ishikawa, Ph.D., M.M., CCC-SLP,[1]
Joel MacAuslan, Ph.D.[2]

1. Department of Speech and Hearing Sciences, College of Applied Health Sciences, University of Illinois at Urbana-Champaign
2. STAR Analytical Services, Lexington, Massachusetts

INTRODUCTION

Spasmodic dysphonia (SD) is focal laryngeal dystonia, characterized by strained-strangled voice quality and uncontrollable spasms of the vocal folds during speech. The spasms affect the timing of laryngeal articulation, which results in linguistically inappropriate voice breaks.[1] These breaks severely reduce speech intelligibility.

The diagnosis of SD remains a clinical challenge, which has led to an average of 4.5 years between onset and receiving a proper diagnosis.[2] The number of specialized clinicians who are able to render the diagnosis is limited due to the rarity of the disease, likely contributing to the delay in receiving the diagnosis.

The diagnosis of SD heavily depends on the auditory-perceptual analysis.[1] However, reliability of such analysis is vulnerable to intra- and inter-rater variabilities. In particular, differentiating adductory (AD) SD from hyperfunctional muscle tension dysphonia (HFN MTD) is challenging. Implementation of a computer-based speech analysis tool may offer a solution by providing a quantitative, objective measurement of the patient's speech.

Voice evaluation frequently uses acoustic tools. However, currently available tools are designed to measure features relevant to voice quality rather than abnormalities in speech sound production. Accordingly, they may be limited in describing how SD affects speech intelligibility. In this paper, we propose the application of a linguistically-motivated, automatic speech analysis tool based on the landmark (LM) theory for describing speech abnormalities in SD speech.

The Landmark-Based Analysis[3]
Articulatory gestures generate moments of abrupt changes in a speech signal. The LM theory calls these moments "landmarks," and assumes that listeners attend to these moments, and use the acoustic patterns around the LMs to decode speech. An LM-based approach incorporates this linguistic knowledge in the process of speech analysis.

METHODS

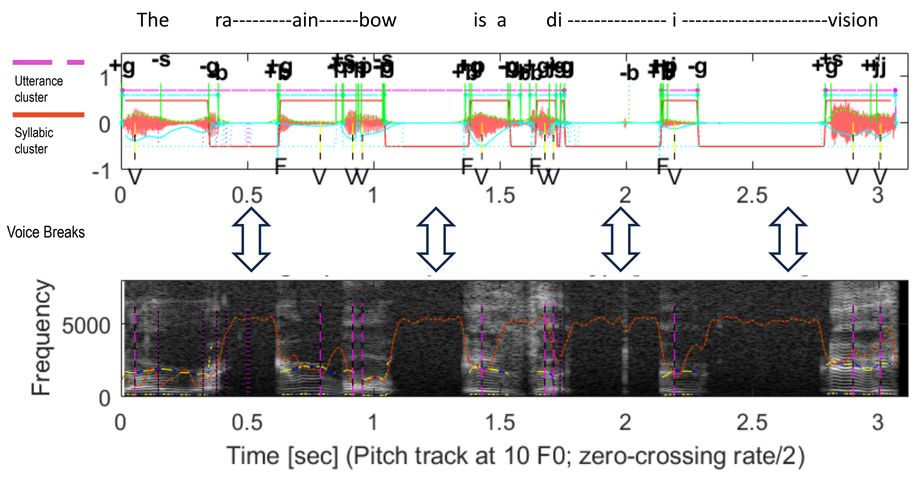
Cases: 10 speakers with healthy voice, 1 speaker with ADSD, 1 speaker with hyperfunctional muscles tension dysphonia (HFN MTD). All selected from Kay-Pentax Disordered Voice Database (Model 4337).

Speech material: A phrase "rainbow is a division," was extracted from the Rainbow Passage.

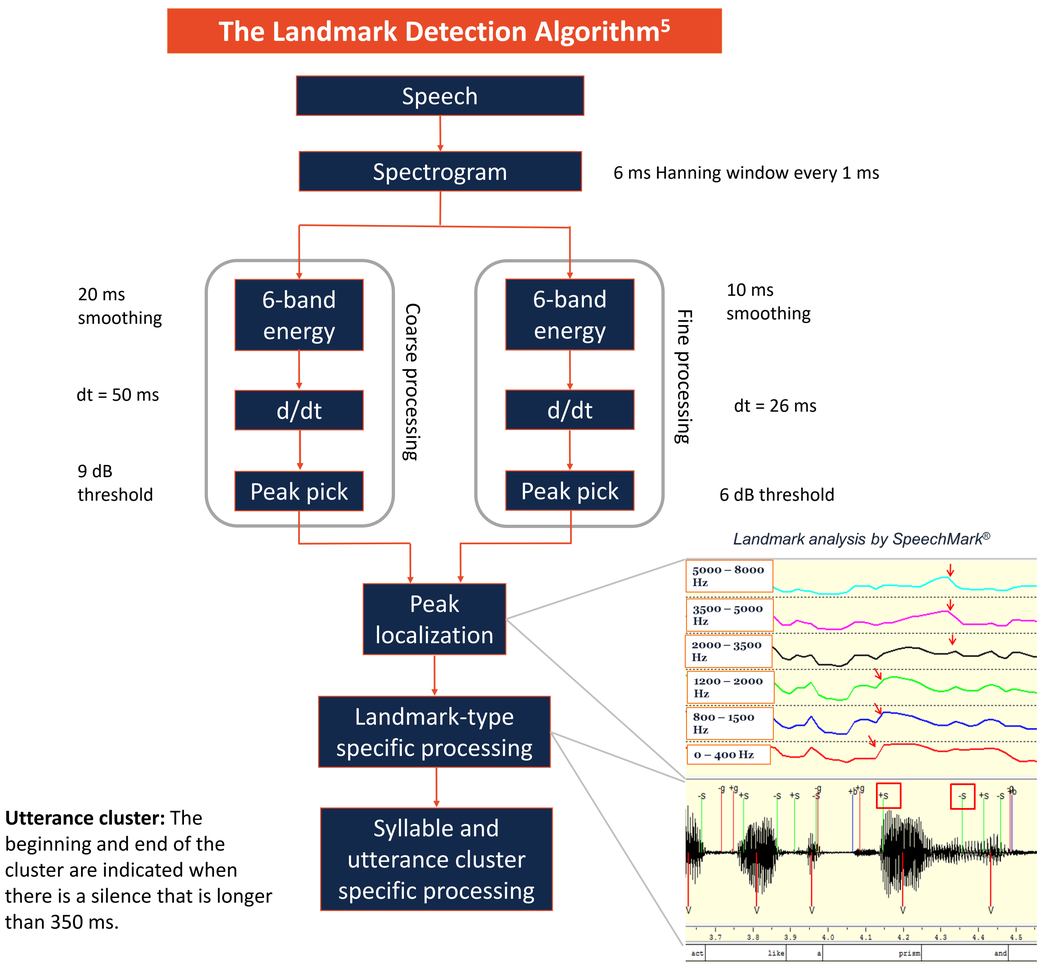
Speech analysis program: SpeechMark® Matlab Toolbox[4] ver 1.1.2

Spasmodic dysphonia causes voice breaks in linguistically inappropriate places in speech.

Landmark-based analysis automatically describes this speech segmentation error.



An example of LM analysis by SpeechMark. Top: Waveform with LMs, syllable clusters, and utterance clusters; Bottom: Spectrogram



Utterance cluster: The beginning and end of the cluster are indicated when there is a silence that is longer than 350 ms.

Syllabic cluster: In general, the start of a syllable cluster in an unvoiced interval is marked by either onset of voicing [+g] or an oral onset event (+b, +f) that occurs up to ~100ms before +g (if there is such an onset event).

If voicing is already present, then the cluster starts at a mouth-opening oral event [+s, -v], provided either:

- that the previous oral event was a mouth-closing event (-s, +v), or
- that it occurs "late", more than ~100 ms after the +g or start of the previous syllable. (This second rule allows us to recognize a word "seven" as having a second syllable after the "v" even if the "v" is not pronounced and the closure is too slow to generate a -s. If the -s or +v is generated, then the first rule applies and there is no complication.)

The ends of syllables follow similar logic, with this complication: If +s or +v occurs "late", then it marks not only the start of a new syllable but also the end of the previous one.

Take a picture to download the poster

Take a picture to learn more about SpeechMark®

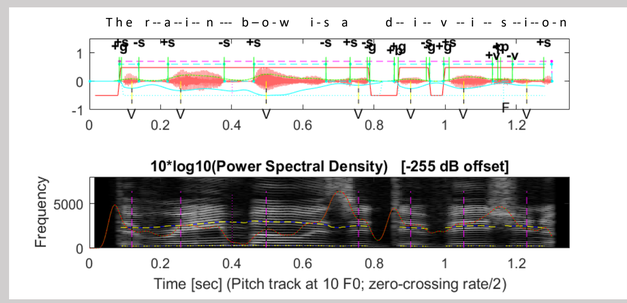
STAR ANALYTICAL SERVICES

Voice and Speech Rehabilitation Research Lab

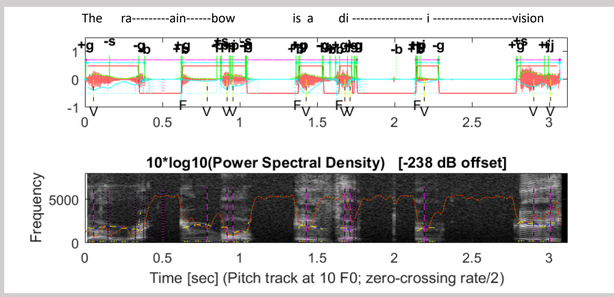
ILLINOIS

RESULTS & DISCUSSION

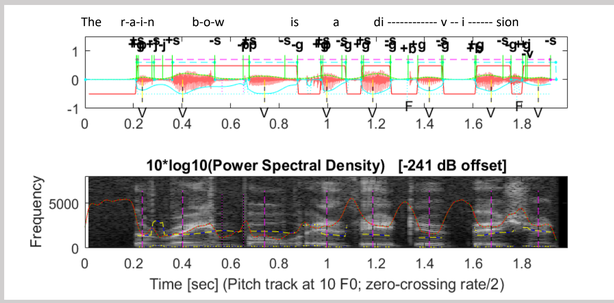
Healthy speaker:
1 utterance, 3 syllabic clusters



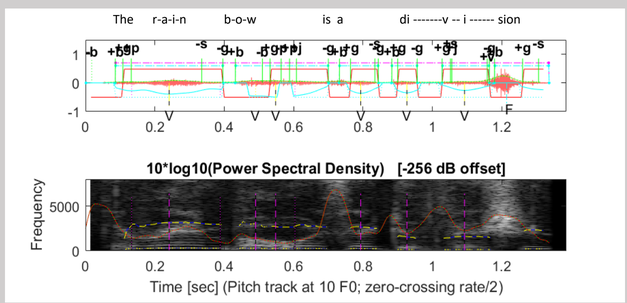
ADSD Pre-Tx:
3 utterance clusters, 6 syllabic clusters



ADSD Post-Tx:
1 utterance, 6 syllabic clusters



HFN MTD:
1 utterance, 6 syllabic clusters



- The LM-based algorithm detected a greater number of utterance clusters in pre-treatment ADSD speech than normal and HFN MTD speech.
- The number of utterance clusters in ADSD speech was reduced to the level of healthy speech after the treatment.
- The number of syllabic cluster in ADSD speech did not change after the treatment; however, the absence of inappropriate voice break in the word "rainbow" was captured.
- In average, the number of syllabic cluster in normal speech was 2.9 clusters (SD =1.1). The number of syllabic clusters was twice as many in ADSD and HFN MTD samples.
- These observations may indicate clinical utility of the syllabic and utterance clusters.
- The lack of samples that represent SD and HFN MTD limited further analysis.
- Future studies
  - Validate the method with a larger number of speakers
  - Evaluate the relationship between number of utterance clusters and intelligibility in SD speech.

References

1. Roy, N., Whitchurch, M., Merrill, R. M., Houtz, D., & Smith, M. E. (2008). Differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia using phonatory break analysis. *The Laryngoscope*, *118*(12), 2245-2253.
2. Creighton, F. X., Hapner, E., Klein, A., Rosen, A., Jinnah, H. A., & Johns, M. M. (2015). Diagnostic delays in spasmodic dysphonia: a call for clinician education. *Journal of Voice*, *29*(5), 592-594.
3. 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.
4. Boyce, S., Fell, H., & MacAuslan, J. (2012). SpeechMark: Landmark detection tool for speech analysis. In *Thirteenth Annual Conference of the International Speech Communication Association*.
5. Liu, S. A. (1996). Landmark detection for distinctive feature-based speech recognition. *The Journal of the Acoustical Society of America*, *100*(5), 3417-3430.