SpeechMark Acoustic Landmark Tool: Application to Voice Pathology

Suzanne Boyce[1], *Marisha Speights*[1], *Keiko Ishikawa*[1], *Joel MacAuslan*[2]

[1] Department of Comm. Sciences and Disorders, University of Cincinnati, Cincinnati, OH, USA

[2]Speech Technology and Applied Research Corporation, Bedford, MA, USA

boycese@ucmail.uc.edu, speighma@mail.uc.edu, ishikak@mail.uc.edu, joelm@s-t-a-r-corp.com

Abstract

One area of voice research that has historically been understudied is the interaction between voice pathology and acoustic aspects of the speech signal that affect intelligibility. Landmark-based software tools are particularly suited to fast, automatic analysis of small, non-lexical differences in the acoustic signal reflecting the production of speech. We are building a tool set that provides fast, automatic summary statistics for measures of speech acoustics based on Stevens' paradigm of landmarks, points in an utterance around which information about articulatory events can be extracted. This paper explores the use of landmark analysis for evaluation of intelligibility-based measures of vocal pathology.

Index Terms: speech analysis, landmarks, voice pathology

1. Introduction

As part of a three-year U.S. NIH-funded project, we are making our acoustic landmark detection tool, the SpeechMark™ system, available to the broader scientific community. The system will be available in several forms: (1) as a Matlab toolbox for software developers, and (2) as a "plug-in" or "port" designed to augment the capabilities of software already in use by different sectors of the scientific community. At Interspeech 2013, we demonstrated the early version of our SpeechMark plugin for the WaveSurfer acoustical analysis and R statistical analysis packages with illustrative data from Parkinson's patients with and without Deep Brain Stimulation [1]. In Interspeech 2013, we propose to demonstrate a novel use of SpeechMark in identifying aspects of the acoustic signal that reduce intelligibility in cases of vocal pathology. Specifically, we are interested in the interaction of laryngeal pathology with spectral and temporal features that convey the content of sentence-sized messages. We use a form of the landmark analysis system of Liu [2] and Howitt [3] based on Stevens [4-5]. Figure 1 illustrates the basic abrupt landmark detection. Vowel landmarks are based on peak harmonic power. Further details, including relevant articles, may be found at the website www.speechmrk.com.

Patients with vocal pathology frequently complain of being unintelligible to listeners. While some of this effect must be due to the reduced loudness typical of patients with voice pathology, it is also well-known that many aspects of voice pathology involve the speaker's inability to sustain periodic vocal fold vibration and aerodynamic forces in a consistent or predictable way [6]. This difficulty will affect the patterns of abrupt change in frequency bands characteristic of release bursts, fricative onsets, etc. relative to changes in voicing that Stevens and colleagues refer to as landmarks [4-5, 7-8] and which convey acoustic information important for accurate speech perception. Thus, we have found that running speech uttered by a speaker with vocal pathology will show unusual patterns of landmarks. The precise pattern of landmarks can hold interesting clues as to the interaction of the vocal fold pathology with aerodynamic forces in the vocal tract and respiratory support. For instance, we expect different patterns of vowel (V) landmarks, since vocal fold vibration may reach a peak of harmonic power at points in an utterance that differ from the pattern of normal speech. For formant-based measures, we expect that breathiness and lapses in voicing will reduce the ability of the software to detect formants accurately. For instance, Figure 3 shows how increased coupling with what is probably a subglottal resonance around 700 Hz (as a result of breathiness) affects automatic vowel space measures [9], while Figure 2 shows how inconsistent timing between source power (from the larynx) and maximal oral opening for syllabic nuclei affects detection of V landmarks.

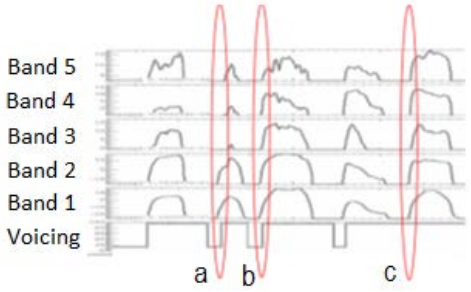

Figure 1. Initial spectral analysis of an utterance: voicing (bottom) and five frequency bands' energy waveforms. (a) Too few bands show large, simultaneous changes in energy. (b) All bands show large, simultaneous energy increases immediately before the onset of voicing, identifying a +b (burst) landmark. (c) All bands show large, simultaneous energy increases during ongoing voicing, identifying a +s (syllabic) landmark.

It is important to note here that, as we envision it, the value of the software does not lie in accurately detecting all instances of bursts, or all peaks of harmonic power, or in accurately identifying the shape of the vocal tract by tracking formants. The value of this type of software lies in determining how many times the threshold for identifying abrupt changes is met, or how much the reduced harmonic power affects the accuracy of formant identification. In effect, the software is aimed at detecting how much information is distorted or missing in the acoustic signal as produced by a person with a voice disorder. While at any one time, listeners may be operating with thresholds for detection and measurement different from those chosen for SpeechMark, the degree to which landmarks and measures (such as formants) are misdetected may enable prediction about the ease with which the listener can interpret the meaning of an utterance.

2. Demonstration Plan

Using a laptop computer, we plan to demonstrate the operation of the SpeechMark landmark detection software using speech recordings culled from two sources: the Kay Elemetrics Database of Disordered Voices, Model 4337 (http://www.kayelemetrics.com), and a set of our own recordings for which we have independent data on intelligibility in quiet and in babble noise. Audience members will be able to choose a sentence from a list and a version of the SpeechMark software to test in real time, including real-time display of landmarks. Audience members will also be able to listen to the recording in quiet and in babble noise through headphones at different signal/noise levels.

3. Conclusion

Aspects of disordered voices that affect intelligibility relative to normal voices can be detected by a landmark-based approach.

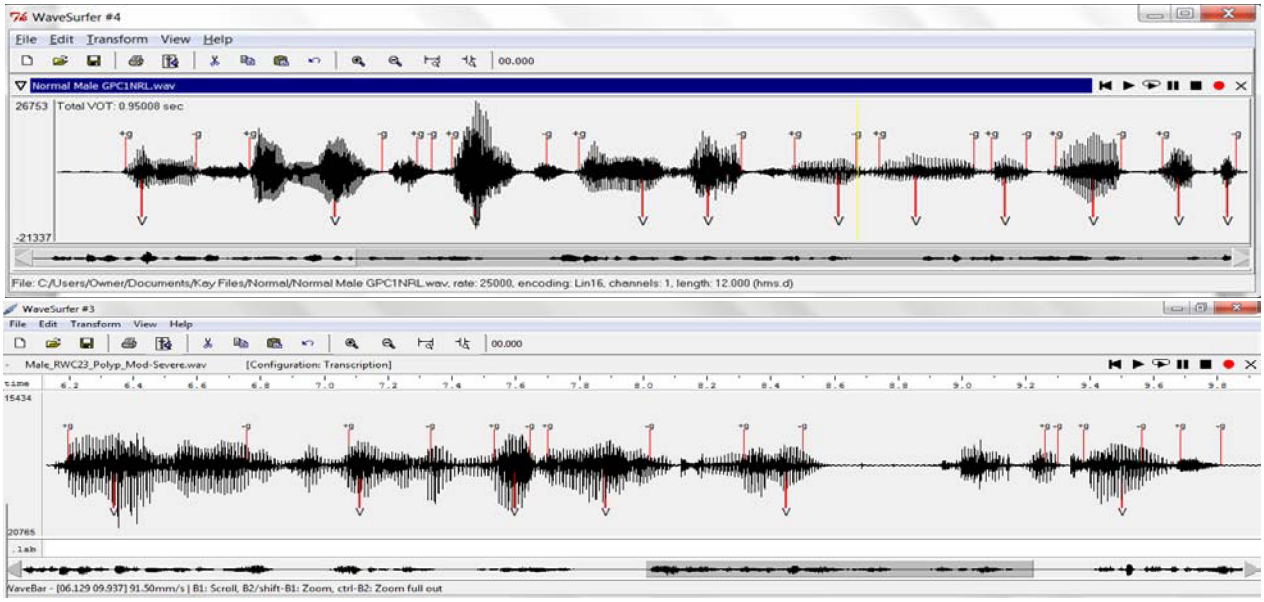

Figure 2. Results of the SpeechMark wavesurfer plugin for the 2nd phrase of the Rainbow Passage as produced by 2 male speakers of American English of similar age, with normal (top panel) and disordered (bottom panel) voices. The disordered voice comes from a speaker with a diagnosis of moderate-severe dysphonia as a result of vocal fold polyps. The landmarks are shown as vertical lines with labels. Note that fewer V landmarks are detected for the disordered voice.

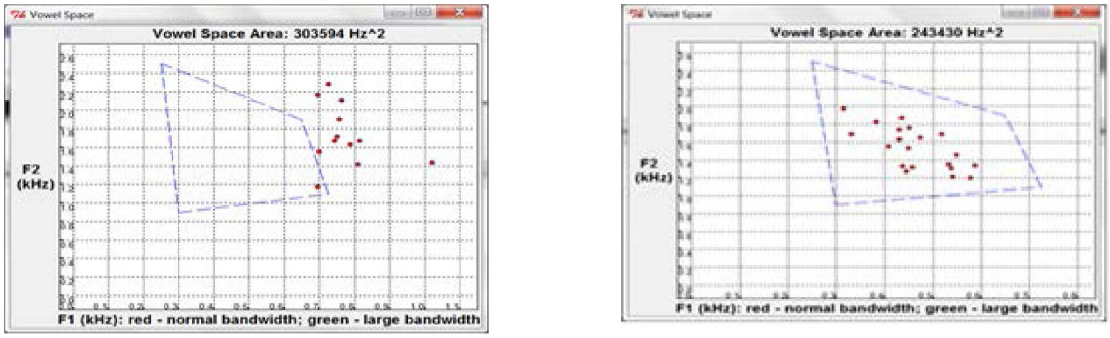

Figure 3. Vowel space measure from disordered male voice with diagnosis of severe vocal fold paralysis (left) and normal male voice (right) producing 2nd sentence of Rainbow Passage. Note that detected F1 clusters around 700-800 H, which does not reflect the different vowels of the passage, but is a typical subglottal resonance for male speakers [9].

4. References

[1] Boyce, S., Fell, H., & J. MacAuslan. "SpeechMark: Landmark Detection Tool for Speech Analysis", in 13th Annual Conference of the International Speech Communication Association Proc.,, September 9-13, 2012.

[2] Liu, S., "Landmark detection in distinctive feature-based speech recognition", Ph.D. Dissertation, Massachusetts Institute of Technology: Cambridge, Massachusetts, 1995.

[3] Howitt, A.W., "Automatic Syllable Detection for Vowel Landmarks", Ph.D. Dissertation, Massachusetts Institute of Technology: Cambridge, Massachusetts, 2000.

[4] Stevens, K.N., et al. "Implementation of a Model for Lexical Access based on Features", in International Conference on Spoken Language Processing (ICSLP) Proc., 1992.

[5] Stevens, K.N., "Toward a model for lexical access based on acoustic landmarks and distinctive features". Journal of the Acoustical Society of America, 111 (4), 1872-91, 2002.

[6] Jiang, J., Zhang, Y., and C. McGilligan, Journal of Voice, Vol. 20 (1), 2-16, 2006.

[7] Juneja, A. and C.Y. Espy-Wilson. "Speech Segmentation Using Probabilistic Phonetic Feature Hierarchy and Support Vector Machines", in International Joint Conference on Neural Networks Proc., 2003.

[8] Slifka, J.S., et al. "A Landmark-Based Model of Speech Perception: History and Recent Developments", in From Sound to Sense: Fifty Years of Speech Research, 2004.

[9] Lulich, S., Morton, J., Arsikere, H., Somers, M. &. Leung, G. & Alwan, A. "Subglottal resonances of adult male and female native speakers of American English", Journal of the Acoustical Society of America, 132 (4), 2592–2602, 2012.