Automated Tools for Identifying Syllabic Landmark Clusters that Reflect Changes in Articulation

Suzanne Boyce[1], Harriet Fell [2], Lorin Wilde[3], and Joel MacAuslan[3]

[1]University of Cincinnati, OH, [2]Northeastern University, Boston, MA,
[3]Speech Technology and Applied Research, Bedford MA, USA

Abstract: **We have developed a set of software tools to detect articulatory changes in the production of syllabic units based on acoustic landmark detection and classification. Results from the application of this automatic analysis system to studies of Parkinson's Disease and Sleep Deprivation show the ability to detect subtle change. We are making these tools available as add-ons to systems such as Wavesurfer and R.**

Keywords: **speech-acoustic landmarks, syllabic landmark cluster, automatic vocalization processing.**

I. INTRODUCTION

Acoustic evidence provides information on speech production, but that information is scattered across multiple frequency bands and multiple time scales. Landmark analysis [5,6] is one approach by which acoustic patterns characteristic of particular changes in speech movements are detected. In this paper, we describe an extension of the landmark method to the detection of articulatory complexity in the production of syllables, by using clusters of landmarks as a measure of whether a string of (intended) syllables is produced in its canonical form (dictionary pronunciation), in a less complex (CCVC -> CVC), or more *lenited* form (softened consonants). We refer to this measure as a measure of syllabic complexity, and to our landmark cluster measure as a "syllabic cluster" measure. We have applied this approach successfully to measure speech articulation changes in Parkinson's Disease, in infant speech development, in sleep deprivation, and other studies.

The notion of syllabic complexity is illustrated as follows. A word such as "interesting" can have four syllables in its canonical form, but when uttered as /inrɛstiŋ/ it has three syllables with fewer consonants, and thus reduced articulatory complexity. In landmark systems in general, different types of types and combinations of speech sounds are detected as different patterns of landmarks. In our particular system, a syllabic landmark cluster is a sequence of consecutive landmarks grouped according to specific rules. For example:

1. A syllabic cluster must contain a voiced region of at least 30 ms, corresponding to a syllable nucleus.

2. A noisy sound such as "s" (/s/) must hit a threshold of loudness before being detected.

If uttered in a canonical fashion, the pronunciation of a word will show a characteristic pattern of landmarks for each syllable in that word. As long as the syllables are uttered with the same acoustical characteristics, our measures will detect the same pattern of landmarks. However, if the syllables are uttered less canonically—perhaps with less extreme articulatory movements, less precise timing, or reduced aerodynamic support—--- then fewer landmarks will be detected. Our version of the speech-acoustic landmark system thus can be used to detect two common effects in speech production: (1) simplification of syllable onsets (e.g. "string" /striŋ/ as /sriŋ/), nuclei (e.g. "diamond" /dɑɪmƏnd as /dɑmƏnd/) and rimes (e.g. "pelt" /pɛlt/ as /pɛl/, and (2) fewer uttered syllables.

II. METHODS: LANDMARK SYSTEM

Landmarks and Rules: Our landmark analysis system is based on Stevens *et al.* [6]. especially as developed by Liu [5] and Howitt [4]. The speech signal is automatically partitioned into 5 frequency bands plus a voicing-status contour. Abrupt landmarks are identified as points where abrupt changes in the amplitude of several frequency bands coincide in a specified pattern [5,6]. These landmark patterns are identified by comparison between "coarse" and "fine" temporal resolution.

The system detects the following types of landmarks:

1. g: glottis. Marks a time point at which voicing begins (+g) or ends (-g), based on the harmonic spectrum.
2. s: syllabicity. Marks sonorant consonantal releases (+s) and closures (-s).
3. b: burst. Marks frication onsets or affricate/stop bursts (+b) and points where aspiration or frication ends (–b) due to a stop closure.
4. V: vowel. Marks a time point corresponding to maximum harmonic power.

The +/- b and +/- s landmarks are identified from patterns of rapid change in the amplitude of several frequency bands. The +/-g and V landmarks are identified from the harmonic spectrum.

This system makes no attempt to identify phonemes, but it is sensitive to broad categories of speech sounds and to aspects of metrical structure. The features it detects are those known as "articulator free" [6] because they are independent of the specific articulator used to produce the segment. These features are instead associated with creation and release of constrictions in the vocal tract and with the acoustic consequences of those constrictions and releases.

An example of how abrupt landmarks are determined from patterns across frequency and voicing bands is shown in Fig. 1. An example of landmark location in the speech signal can be found in Fig. 2, which shows a spectrogram of the nonsense word "pataka" repeated 10 times in two breath groups by a native speaker of American English with moderate dysarthria due to Parkinson's Disease.

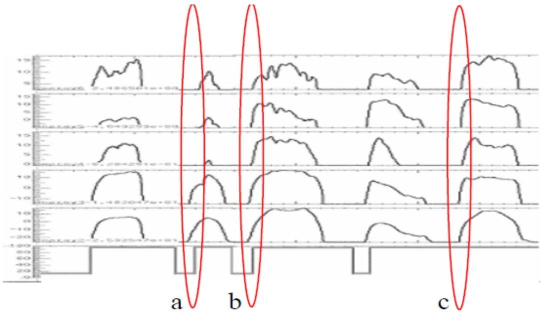

Figure 1. Spectral analysis of an utterance: voicing (bottom) and five frequency bands' energy waveforms. (a) Too few bands show large, simultaneous changes in energy. (b) All bands show large, simultaneous energy increases immediately before the onset of voicing, identifying a +b (burst) landmark. (c)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

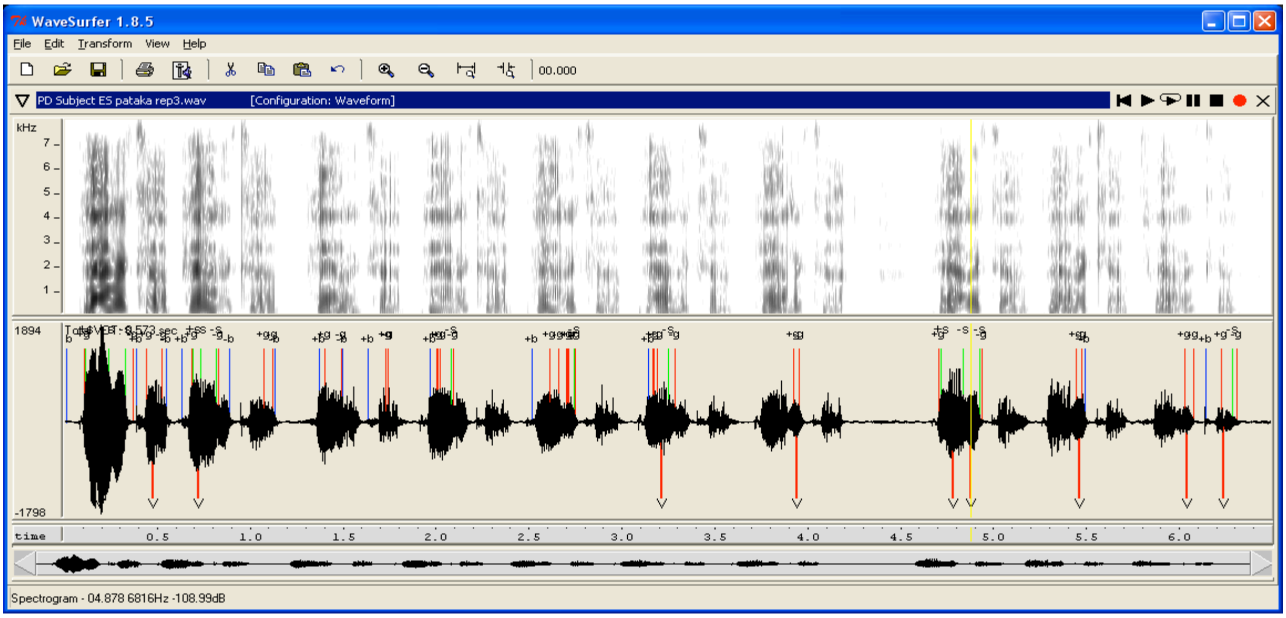

Figure 2. Ten repetitions of the nonsense word /pətəkə/ by an American English speaker with moderate dysarthria due to Parkinson's Disease. Vertical lines above the waveform pane show +/- b, +/-s and +/- g landmarks. Vertical lines below the waveform pane show Vowel landmarks as V. The period of silence shows the pause between breath groups.

Use of the Landmark System to Characterize Differences in Speech Production: The landmark system operates with empirically derived threshold values. As discussed, abrupt landmarks are determined by the patterns of abrupt change across frequency and voicing bands; if the amplitude value of the signal in a particular set of frequency and voicing bands meets the predetermined threshold for abruptness, then a landmark is detected. If the amplitude value of the signal in any of the frequency/voicing bands does not meet this criterion, then no landmark is detected.

The operation of this system is shown by the pattern of V landmarks in Fig. 2. As noted above, the speaker produced /pətəkə/ in two breath groups; the first seven repetitions belong to the first breath group, and the following three repetitions belong to the second breath group. This speaker showed a tendency to dysphonia typical of Parkinson's patients, characterized subjectively as causing a harsh and breathy voice, and the dysphonic phonation was more marked in the late portions of a breath group—presumably because reduced breath support made it more difficult to sustain normal periodic vocal fold vibration. Because the V landmarks are computed on the basis of harmonic power, and dysphonic vowels are produced with less harmonic power, fewer V landmarks will be detected on dysphonic voices. This effect is shown in Fig. 2, where the first few repetitions in the first breath group are marked with V landmarks on the stressed syllable, while the last few repetitions in the same breath group show that no such landmarks have been detected. Note that these repetitions were produced with vowels—this is evident in the spectrogram--but the vowels had too little harmonic power to be registered as V landmarks.

Grouping Landmarks to Characterize "Syllabic Clusters": Fell & MacAuslan originally developed the "syllabic cluster" measure to detect the increasing syllabic complexity of utterances by young children [2, 3]. More recently, we have applied this method, termed the Syllabic Cluster analysis, to speech uttered under normal and sleep-deprived conditions, and to speech by Parkinson's Disease patients undergoing Deep Brain Stimulation (DBS) therapy.

Cluster Rules: The Syllabic Cluster analysis works by grouping sequences of detected landmarks into clusters that roughly correspond to syllabic units in the acoustic speech signal. The grouping rules include categorical dependencies as well as dependencies of timing, and were empirically determined from datasets of speech.

For example, one such rule states that a gap of 30 ms in voicing, with whatever ±b's immediately follow it, identifies a type of syllable cluster endpoint. In contrast, burst-like noise that does not occur within 120 ms before a voiced region, or 80 ms after, is not part of a cluster. Indeed, we have found it useful to designate these types of isolated bursts as non-speech noise. The syllabic grouping procedures are described in more detail in Fell et al. [2,3] and Boyce et al. [1] The following is a list of examples of some common types of syllabic cluster that occur in speech:

- (+g,-g)- singleton V [vowel] or CV [consonant-vowel] syllables, where C is voiced;
- (+g,-s) - V or voiced-CV syllables followed by a sonorant consonant and syllabic cluster;
- (+s, -g) - V or voiced-CV syllables, preceded by a syllabic cluster;
- (+g,-s,-g) - VS syllable, where S is a sonorant consonant or voiced obstruent adjacent to the +g or -g;
- (+b,+g,-g) - syllable beginning with fricative: (+b) marks the presence of frication;
- (+b,-b,+g,-g) - syllables with an initial plosives: (+b , -b) mark the beginning and end of the release.

III. METHODS: APPLICATION

Parkinson's Disease Study: In one study using the Syllabic Cluster measure, we contrasted speech as produced by Parkinson's Disease (PD) patients who were receiving Deep Brain Stimulation (DBS). In the typical progression of Parkinson's Disease, patients show clinically significant levels of unintelligible speech later than they show gross motor symptoms. Thus, patients in DBS programs may not be showing clinically overt signs of dysarthric speech. However, the application of DBS therapy can sometimes cause their speech intelligibility to worsen, and this is both a matter of clinical concern and scientific interest. The data described in Fig. 3 come from a study of 15 Control vs 15 PD patients who had undergone surgery for Deep Brain Stimulation (DBS) repeating the syllable /ka/. The aim of the study was to detect subtle and/or overt changes in speech production when DBS stimulus was OFF vs. ON.

Sleep Deprivation: In another study, we used the Syllabic Cluster analysis to test whether speech articulation changes as a result of sleep deprivation. Studies of both speech articulation per se, and listener perceptions of change, have shown conflicting results to date [1]. In our study, the speech of 17 speakers of American English (9 female, 8 male) was recorded at 8 hour intervals over 32-40 hours without sleep. (Not all subjects completed the final session.) Subjects read aloud the Rainbow Passage each time. To control for the possible effect of familiarity with the speech materials, another set of 15 subjects (7 male and 8 female) read aloud the Rainbow Passage at 8-hour intervals while maintaining a normal sleep schedule.

III. RESULTS AND DISCUSSION

Parkinson's Disease Study: The mean cluster rate in rapid repetitions of the syllable /kɑ/ decreases (a) between Control vs. PD speakers, and (b) as a result of DBS. The differences were significant at the .01 level.

Sleep Deprivation: The first two sessions were combined as the Early, or Rested, condition. The last two sessions were combined as the Late, or Sleep Deprived condition. As Fig. 4 shows, Syllabic Cluster rate decreased between the Early and Late sessions. This difference was significant at the $p < .05$ level by a binomial (sign) test. In contrast, the Early vs. Late sessions were not significantly different for speakers performing the identical task while following their normal sleep schedule ($p > .10$ by a binomial (sign) test.

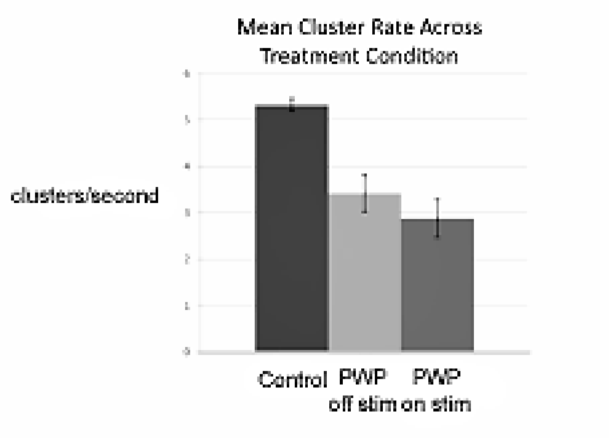

Figure 3. The mean rate of Syllabic Cluster occurrence for 15 age and gender-matched control subjects (Control) vs 15 speakers of American English with Parkinson's Disease (PWP) across Stimulus ON, and Stimulus OFF conditions.

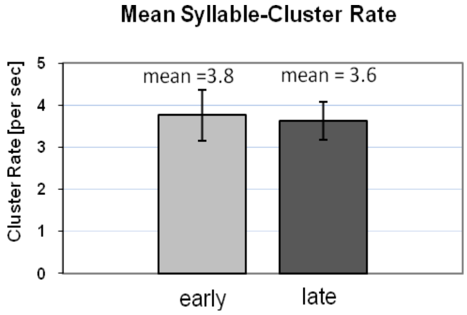

Figure 4. The mean rate of Syllabic Cluster occurrence for 17 speakers of American English reading the Rainbow Passage aloud in Early vs. Late sessions of a 30-40 hour period without sleep.

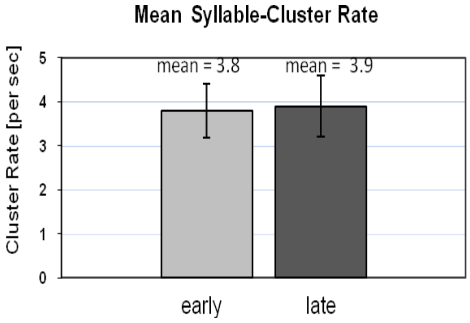

Figure 5. The mean rate of Syllabic Cluster occurrence for 15 speakers who read the Rainbow Passage aloud in Early vs. Late sessions while following their normal sleep patterns.

VI. CONCLUSION

The Syllabic Cluster analysis based on acoustic landmark detection appears to be sensitive to articulatory differences in speech production scattered across multiple frequency bands and multiple time scales. The Parkinson's Disease results suggest this analysis provides a rough measure of a speaker's ability to repeat speech materials with a certain level of articulatory precision at a particular speech rate. The Sleep Deprivation results suggest that speech articulation does indeed change with sleep deficit in a way that reduces the rate at which well-formed syllabic clusters are produced and that this change is not due to familiarity with the speech materials. Both sets of results suggest the analysis is sensitive to very subtle changes that listeners do not always detect. The automatic nature of the analysis facilitates evaluation of large amounts of data.

We are currently developing a set of software tools for automatic landmark detection, and classification into syllabic cluster patterns, to be available as add-ons to systems such as Wavesurfer and R.

ACKNOWLEDGEMENT

Support: R43 DC010104, R42 AG033523, R21 HL086689, R42 HD34686 from the U. S. National Institutes of Health.

REFERENCES

[1] S. Boyce and J. MacAuslan, "Effects of Sleep Deprivation on Speech Articulation and Intelligibility in Noise," *Proc. Of ASMA* (Aerospace Medical Association), Anchorage, AK. 2011.

[2] H. J. Fell, J. MacAuslan, L. J. Ferrier, K. Chenausky, "Automatic Babble Recognition for Early Detection of Speech Related Disorders," *Journal of Behaviour and Information Technology*, 1999, 18, no. 1, 56-63.

[3] H. J. Fell, J. MacAuslan, L. J. Ferrier, S. Worst, & K. Chenausky, "Vocalization Age as a Clinical Tool," *Proc. of ICSLP* (International Conference on Speech Processing), Denver, USA, 2002.

[4] A. W. Howitt, *Automatic Syllable Detection for Vowel Landmarks*, doctoral thesis M.I.T., Cambridge, MA. 2000.

[5] S. Liu, S. *Landmark detection in distinctive feature-based speech recognition*. doctoral thesis M.I.T., Cambridge, MA. 1995.

[6] K. N. Stevens, S. Manuel, S. Shattuck-Hufnegel, and S. Liu, "Implementation of a model for lexical access based on features," *Proc. Int. Conf. Spoken Language Processing*, Banff, Alberta, 1, 499-502. 1992.