Automated extraction of voice onset time in healthy and pathological speech

Benjamin G. Schultz[1]* & Adam P. Vogel[1,2]

[1]*Centre for Neuroscience of Speech, The University of Melbourne, Australia*

[2]*Redenlab, Australia*

Background: Voice onset time (VOT) is an acoustic measure of speech timing that can be used to signify the onset of neurological disease (Auzou et al., 2000). Conceptually, VOT measures the time difference between the onset time of a burst (i.e., a consonant) and the onset time of the voicing (Lisker & Abramson, 1964). The acoustic features that indicate the onsets of consonants and pitched voices, however, are ill-defined and there is no consensus for how best to annotate VOT (see Auzou et al., 2000).

Aims: We aim to provide guidelines for the best practices when annotating VOT. Our second aim is to develop and test an automated method for extracting VOT.

Methods: We formalize the best practices for annotating VOT based on the seminal method where energy increases in low frequencies (75-500Hz) indicate voicing time and broadband energy increases (75-4000Hz) indicate burst onset times. In line with these practices, the automated VOT extraction method measures summed energy in f0 (75-500Hz) and formant (500-4000Hz) energy bands using continuous wavelet transformations (see Figure 1). We tested the automated VOT method using speech data from a diadochokinetic task where participants repeat the syllables /PA/, /TA/, and /KA/ with manual annotations performed by two naïve assessors. To ensure the method was resilient to speech errors, we compared the speech of healthy controls, frontotemporal dementia patients, and people with Friedrich's ataxia.

Results: Preliminary results show that the automated VOT extraction provides burst and voice onset times that match manual annotation within an error margin of up to 14ms and 19ms, respectively. We will present further data to test the automated VOT extraction method on speech data that contain errors (e.g., from patient with pathological speech).

Conclusion: We provide a fast and reliable means of extracting VOT from acoustic speech data that does not rely on the visual inspection of acoustic parameters. This provides a measure of VOT that is objective and consistent. Applications for using VOT to assess neurodegenerative disease from the diadochokinetic task and natural speech are discussed.

References:

Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: a review. *Clinical linguistics & phonetics*, *14*(2), 131-150.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

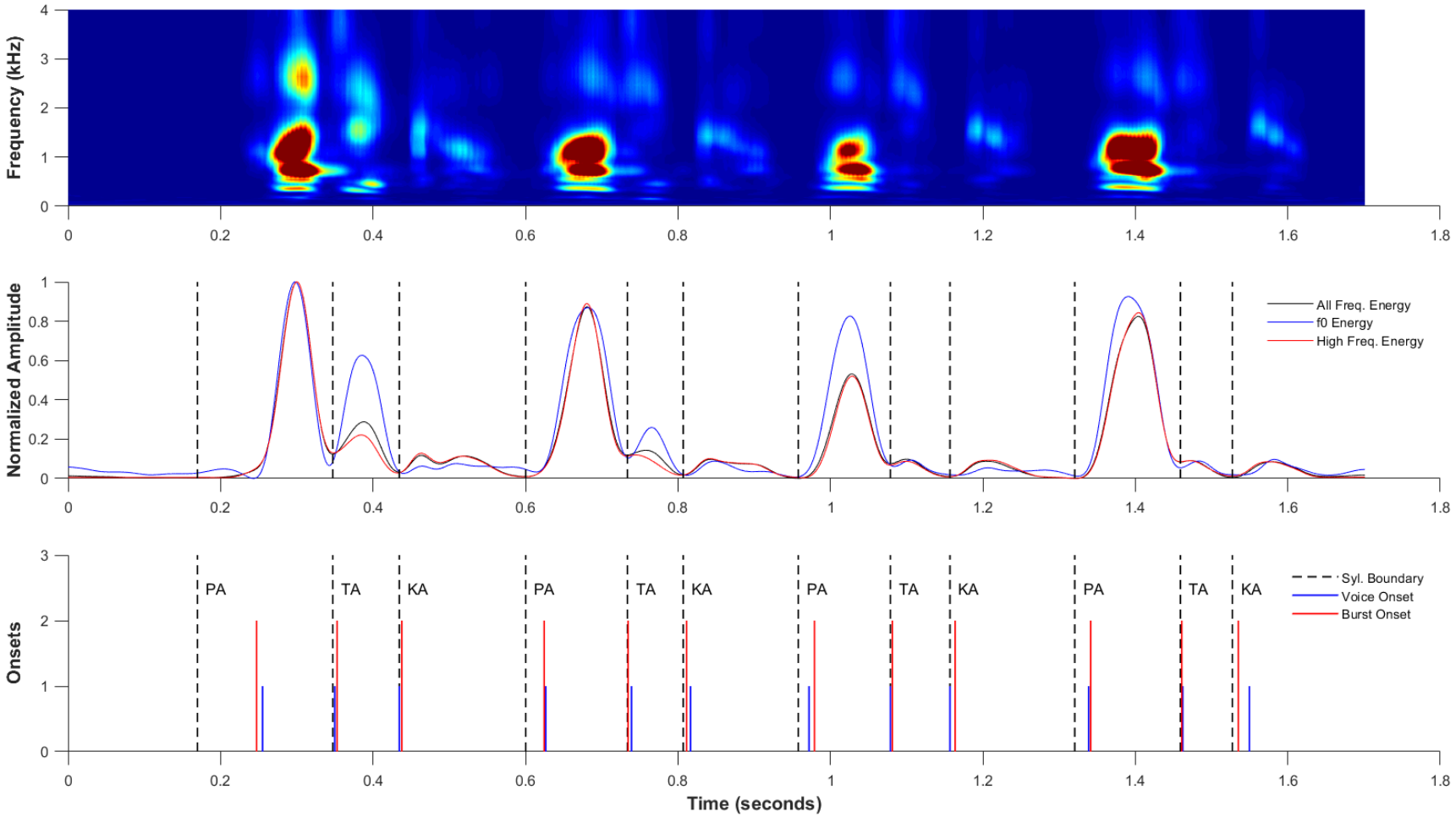

Figure 1. Visualization from the automated voice onset time extraction method. The top panel shows the continuous wavelet transform of a speech signal. The middle panel shows the normalized amplitude of energy across all frequencies (75Hz to 4000Hz; black line), f0 frequencies (75Hz to 500Hz), and high frequencies (500Hz to 4000Hz) with syllable boundaries marked with dotted lines. The bottom panel shows the voice (blue lines) and burst (red lines) onset times.