VOCAL TRANSIENT ANALYSIS USING PIECEWISE LINEAR APPROXIMATION

Takeshi Ikuma, PhD[1,2], Andrew J. McWhorter, MD[1,2], Robin Samlan[5], Tobias Schraut[4], Michael Doellinger[4], and Melda Kunduk, PhD[1,2,3]

[1]Dept. of Otolaryngology-Head and Neck Surgery, LSU Health Sciences Center, New Orleans, Louisiana, U.S.A.
[2]Voice Center, The Our Lady of The Lake Regional Medical Center, Baton Rouge, Louisiana, U.S.A.
[3]Dept. of Communication Sciences & Disorders, Louisiana State University, Baton Rouge, Louisiana, U.S.A.
[4]Dept. of Otorhinolaryngology Head & Neck Surgery, Medical School Division of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany
[5]Dept. of Speech, Language, & Hearing Sciences, University of Arizona, U.S.A.

Keywords: Voice onset/offset; Vowel-consonant-vowel; Highspeed videoendoscopy

Abstract

Objectives / Introduction:

The mechanics and kinematics of vocal fold vibration during the transients—either onset or offset of sustained phonation or in speech context—potentially contain rich information that can be related to vocal health, dysfunction, and patients specific voice symptoms. Measuring objective features of the transients from real speakers can be challenging, especially in speech context, because the non-transient segments are also dynamic with constantly fluctuating amplitude and frequency. Thus, there is no clear reference point to determine when the signal is out of the transient reaching so-called “steady state”.

In literature, the transient segment of sustained phonations were determined via model fitting[1], by a percentage of the steady-state amplitude[2], by the maximum glottal length[3], or by utilizing a feature such as vocal attack time[4] which does not depend on transient/steady-state boundary. Yet another approach, which has been explored in speech context, is the voiceless-consonant-induced fundamental frequency[5] which is measured from a fixed number of vocal cycles from a consonant-induced cessation. This presentation proposes a method to establish objectively the transient/steady-state boundary by fusing the vibration amplitude envelope and instantaneous fundamental frequency estimates with piecewise linear approximation. Use of a piecewise linear function is a simple and effective way to unify the handling of vocal transitions influenced by both voiced and voiceless consonants.

Methods:

Given a one-dimensional voice signal capturing a vowel-consonant-vowel (VCV) transitions, the goal is to identify when the consonant-influenced segment begins and ends (let t_{ss} to mark the boundary). We assume that an approximate location of the consonant is given. Also, the influence of a consonant is assumed to lower the vibration amplitude envelope and fundamental frequency f_o always.

The proposed process is performed in two steps. First, the presence of vocal cessation is detected to establish the least active reference points (t_{min}). If cessation is present, its starting and ending times are determined. If not, the time markers of the minimum amplitude and f_o are recorded instead. The cessation analysis fuses the information from both envelope and f_o contours. Also, the AC-component of the voice signal is used refine the $t_{\min}$ estimate by detecting the vanishing point of its cyclic behavior.

With the least-active reference points, the second step models the behaviors of the envelope and f_o contours from least active to most with a 2-segment piecewise linear function. The junction between the two segments are expected near the edge of consonant-influenced segment. The vowel-side segment is then extended till it intersects the contour to form the first candidate of the boundary (t_0). Some consonant induces overshooting behavior on the envelope or f_o. This transient characteristic is captured by fitting another 2-segment function is fitted the vowel starting from the nearest peak to t_0 on the vowel side away from the consonant. The first intersection with the contour is marked t_2. If the %overshoot metric is over a preset threshold (e.g., 20%), t_2 is chosen as t_{ss} else t_1 is selected. Boundaries of feature contours are individually processed.

Results:

To demonstrate, Fig. 1 shows analysis outcomes of the glottal area waveforms (GAWs) of nasoendoscopic high-speed videoendoscopic recordings at 4000 frames/second. These are VCV patterns with /i/ and alveolar consonants (/iti/, /idi/, /isi/, /izi/). The algorithm properly handled both with and without cessation.

Conclusions:

Piecewise linear function is an effective tool to describe changes in the behaviors of the amplitude or f_o behavior of vocal signals in consonant-influenced segment from neighboring vowel-only segment. This is achieved by having three simple degrees of freedom to model specific features (the overall slope of each segment and the transition time).

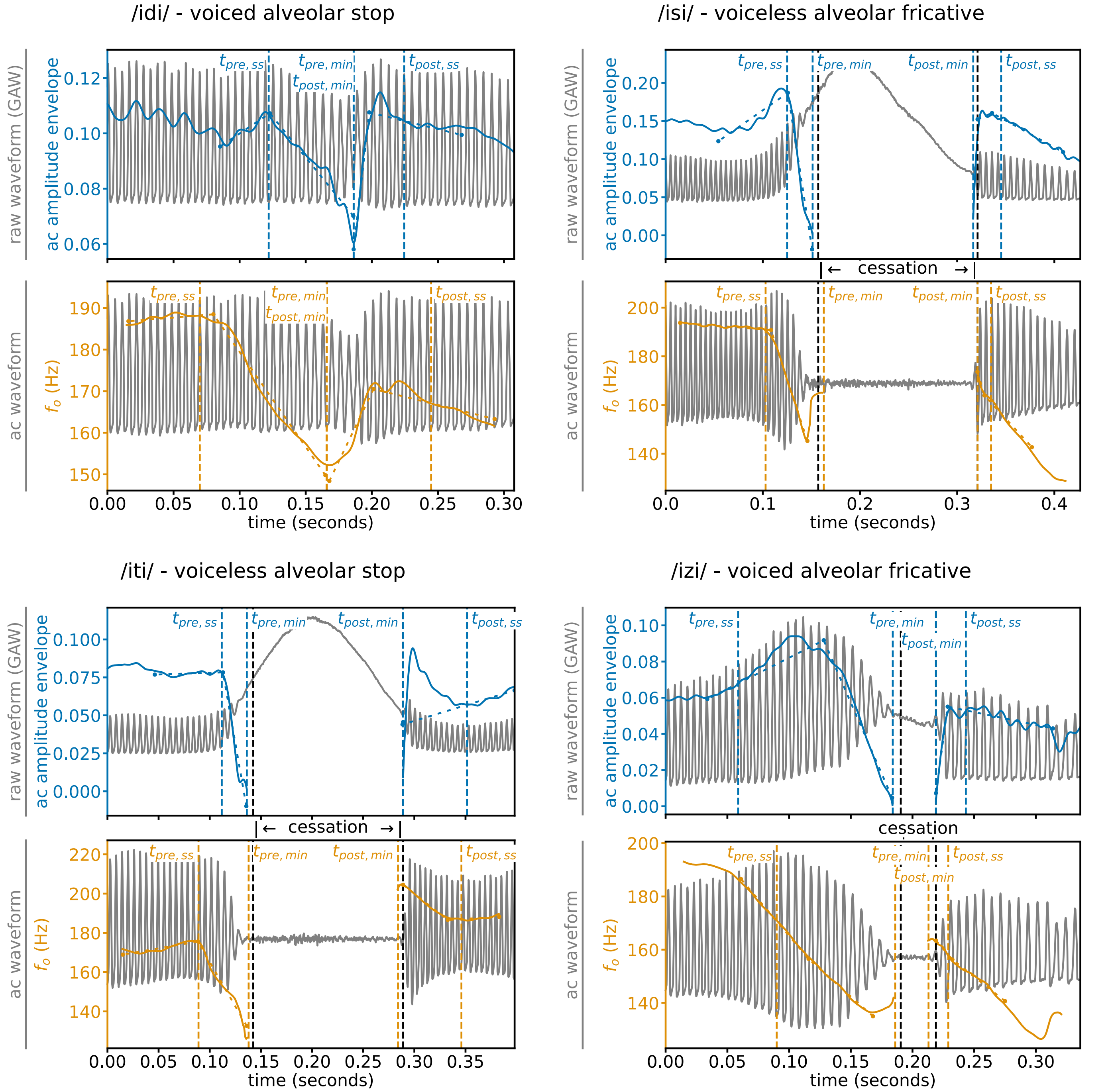

Fig. 1. Analysis outcome examples: Constant-influenced transient segments in alveolar VCV patterns of GAW signals.

References

1. Mergell P, Herzel H, Wittenberg T, Tigges M, Eysholdt U. Phonation onset: Vocal fold modeling and high-speed glottography. *J Acoust Soc Am*. 1998;104(1):464-470. doi:10.1121/1.423250
2. Kunduk M, Ikuma T, Blouin DC, McWhorter AJ. Effects of volume, pitch, and phonation type on oscillation initiation and termination phases investigated with high-speed videoendoscopy. *J Voice*. 2017;31(3):313-322. doi:10.1016/j.jvoice.2016.08.016
3. Ikuma T, Melda K, Fink D, McWhorter AJ. A spatiotemporal approach to the objective analysis of initiation and termination of vocal-fold oscillation with high-speed videoendoscopy. *J Voice*. 2016;30(6):756.e21-756.e30. doi:10.1016/j.jvoice.2015.09.007
4. Orlikoff RF, Deliyski DD, Baken RJ, Watson BC. Validation of a glottographic measure of vocal attack. *J Voice*. 2009;23(2):164-168. doi:10.1016/j.jvoice.2007.08.004
5. Löfqvist A, Baer T, McGarr NS, Story RS. The cricothyroid muscle in voicing control. *J Acoust Soc Am*. 1989;85(3):1314-1321. doi:10.1121/1.397462