Can we detect initiation of tongue internal changes before overt movement onset in ultrasound?

Pertti Palo[1]

[1] *Queen Margaret University, Scotland, UK,* pertti.palo@taurlin.org

Abstract: In order to understand speech articulation, we need to understand not only what movements of the articulators are used to produce a given sound but also how those articulator movements are produced by muscle actions. This paper approaches this problem by analysing ultrasound data with two methods: Pixel Difference accounts for all change apparent in tongue ultrasound data (Palo, 2019; Palo, P. and Moisik, S. R. and Faytak, M., 2020) and Average Nearest Neighbour Distance evaluates the distance between tongue contour splines. Preliminary results show that these methods can be combined to study the time course of articulatory activation and movement and particularly that Pixel Difference does provide a means to detecting tongue internal changes before they become overt movement.

Background: The immediate theoretical interest of this study is in delayed naming (Rastle et al., 2005), speech initiation (Schaeffler et al., 2014) and motor control with specific attention on timing of articulation and acoustic events (Palo, 2019). Previously we have used a dimension reduction method called Pixel Difference to aid in annotating articulatory reaction time (movement onset detection) in Ultrasound Tongue Imaging (UTI) videos (Palo et al., 2015; Palo, 2019). Pixel Difference calculates the Euclidean distance between consecutive raw (uninterpolated) ultrasound images and yields a curve that gives change rate as a function of time over the duration of an ultrasound recording.

Pixel Difference based reaction times are on average shorter than those measured by manually analysing ultrasound videos (Palo, 2019). Given that Pixel Difference takes into account changes at any of the pixels in the images, it seems likely that while human video annotators look for movement of the major structures visible in ultrasound frames, Pixel Difference triggers earlier because it is sensitive to smaller local changes. In this study we seek to determine if this is indeed the case by comparing manual Pixel Difference curve annotation to manual annotation of Average Nearest Neighbour Distance (Zharkova and Hewlett, 2008) curves.

Nearest Neighbour Distance is defined as the distance from a given point in one set (or spline) to the nearest point in the other set (on the other spline). Based on this, Average Nearest Neighbour Distance is calculated by taking the average over the Nearest Neighbour Distances of the sample points on one spline when compared to sample points of another spline. Repeating the procedure for all of the time-adjacent spline pairs in a recording yields a measure of the movement of the tongue spline as a function of time.

Materials: The data we analyse comes from Experiment 3 as reported by Palo (2019). We recorded one 40-year-old native Finnish speaking participant (the first author) in a delayed naming experiment which combined simultaneous acquisitions of audio and UTI and was controlled with AAA (Articulate Instruments Ltd, 2012). Each trial began with the target word being displayed on a computer screen. The participant was instructed read the word internally while remaining at rest until he heard the go-signal (50 ms long 1 kHz beep), which was played out after a random delay of 1.2-1.8 s from the beginning of the UTI recording. He was instructed to produce the target word as soon as possible after he observed the beep.

Preliminary results and discussion: Figure 1 shows Pixel Difference and Average Nearest Neighbour Distance of a token on the same timeline. We can see that the Average Nearest Neighbour Distance curve contains more erratic movement even after smoothing. It is also notable that the curves have the same general shapes and most importantly that at most points the peaks and valleys happen slightly earlier on the Pixel Difference curve. This is particularly evident around the trough around/after 0.5s which is associated with the vowel in the produced

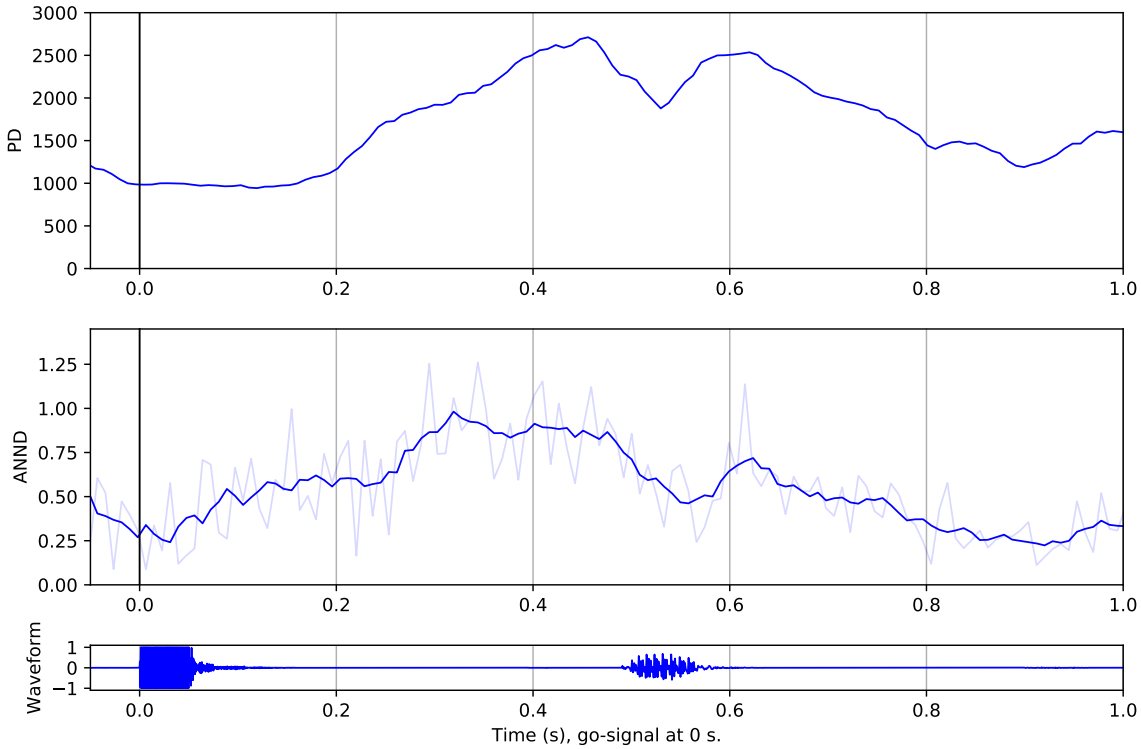

Figure 1: Example of a Pixel Difference (PD) contour, a smoothed Average Nearest Neighbour Distance (ANND) contour (original contour in lighter colour, smoothing with an non-causal moving average window of length 7), and the corresponding acoustic waveform. The participant says the syllable 'po'. Stimulus onset marks the go signal (a 50 ms long 1 kHz beep visible at the beginning of the waveform).

syllable 'po'. However, the beginning of the movement is less easily analysable with a raise in the level of the Average Nearest Neighbour Distance curve that is surprisingly not reflected by Pixel Difference. To provide reliable results at the time of the conference, we will analyse the whole corpus of 479 tokens and manually annotate movement onset on both curves.

References:

Articulate Instruments Ltd (2012). *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

Palo, P. (2019). *Measuring Pre-Speech Articulation*. PhD thesis, Queen Margaret University, Edinburgh.

Palo, P., Schaeffler, S., and Scobbie, J. M. (2015). Effect of phonetic onset on acoustic and articulatory speech reaction times studied with tongue ultrasound. In *International Conference of Phonetic Sciences (ICPhS 2015)*, Glasgow, Scotland, UK.

Palo, P. and Moisik, S. R. and Faytak, M. (2020). SATKIT: Speech Articulation ToolKIT [Python software package]. Available in a public software repository, accessed 28 May 2020. `https://github.com/giuthas/satkit`.

Rastle, K., Harrington, J. M., Croot, K. P., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083 – 1095.

Schaeffler, S., Scobbie, J., and Schaeffler, F. (2014). Measuring reaction times: Vocalisation vs. articulation. In *Proceedings of 10th ISSP*, pages 383 – 386.

Zharkova, N. and Hewlett, N. (2008). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using english /t/ and /ÉŚ/. *Journal of Phonetics*, 37(2):248 – 256.