Acoustic and Articulatory Speech Reaction Times with Tongue Ultrasound: What Moves First?

Pertti Palo, Sonja Schaeffler and James M. Scobbie

Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University

1 Introduction

We study the effect that phonetic onset has on acoustic and articulatory reaction times. An acoustic study by Rastle et al. (2005) shows that the place and manner of the first consonant in a target affects acoustic RT. An articulatory study by Kawamoto et al. (2008) shows that the same effect is not present in articulatory reaction time of the lips. We have shown in a pilot study with one participant (Palo et al., 2015), that in a replication with Tongue Ultrasound Imaging (UTI), the same acoustic effect is present, but no such effect is apparent in the articulatory reaction time.

In this study we explore inter-individual variation with analysis of further participants. We also seek to identify the articulatory structures that move first in each context and answer the question whether this is constant across individuals or not.

2 Materials and methods

Since the phonetic materials, and recording and segmentation methods of this study are mostly the same as those we used in a previous study (Palo et al., 2015), we will provide only a short overview here. Three native Scottish English speakers (one male and two females) participated in this study. We carried out a partial replication of the Rastle et al. delayed naming experiment Rastle et al. (2005) with the following major changes: Instead of using phonetically transcribed syllables as stimuli, we used lexical monosyllabic words. The use of lexical words makes it possible to have phonetically naive participants in the experiment. In addition, we wanted to test if words with a vowel onset pattern in a systematic way with those with a consonant onset. Thus, the words were of /CCCVC/, /CCVC/, /CVC/, and /VC/ type.

The target words used in the original study were: *at*, *eat*, *ought*, *back*, *beat*, *bought*, *DAT*, *deep*, *dot*, *fat*, *feet*, *fought*, *gap*, *geek*, *got*, *hat*, *heat*, *hot*, *cat*, *keep*, *caught*, *lack*, *leap*, *lot*, *map*, *meet*, *mock*, *Nat*, *neat*, *not*, *pack*, *Pete*, *pop*, *rat*, *reap*, *rock*, *sat*, *seat*, *sought*, *shack*, *sheet*, *shop*, *tap*, *teak*, *talk*, *whack*, *wheat*, and *what*. For this study we added the following words with complex onsets: *black*, *drat*, *flat*, *Greek*, *crap*, *prat*, *shriek*, *steep*, *treat*, and *street*.

The experiment was run with synchronised ultrasound and sound recording controlled with Articulate Assistant Advanced (AAA) software Articulate Instruments Ltd (2012) which was also used for the manual segmentation of ultrasound videos. The participant was fitted with a headset to ensure stabilisation of the ultrasound probe Articulate Instruments Ltd (2008). Ultrasound recordings were obtained at a frame rates of $\sim$83 (for the first session with the male participant) and $\sim$121 (for all subsequent sessions) frames per second with a high speed Ultrasonix system. Sound was recorded with a small Audio Technica AT803b microphone, which was attached to the ultrasound headset. The audio data was sampled at 22,050 Hz.

Each trial consisted of the following sequence: (1) The participant read the next target word from a large font print out. (2) When the participant felt that they were ready to speak the word, they activated the sound and ultrasound recording by pressing a button on a keyboard. (4) After a random delay which varied between 1200 ms and 1800 ms, the computer produced a go-signal – a 50 ms long 1000 Hz pure tone.

The acoustic recordings were segmented with Praat Boersma and Weenink (2010) and the ultrasound recordings were segmented with AAA Articulate Instruments Ltd (2012) as in our previous study.

3 Pixel difference

Regular Pixel Difference (PD) refers simply to the Euclidean distance between two consecutive ultrasound frames. It is based on work by McMillan and Corley (2010), and Drake et al. (2013a,b). Our version of the algorithm is explained in detail by Palo et al. (2014).

Instead of using the usual interpolated ultrasound images in the calculations, we use raw uninterpolated images (Figure 1). The fan image of the ordinary ultrasound data is produced by interpolation between the actual raw data points produced by the ultrasound system. The raw data points are distributed along radial scanlines with the number of scanlines and the number of data points imaged along each scanline depending on the setup of the ultrasound system. In this study we obtained raw data with 63 scanlines covering an angle of about 135 degrees and with 256 pixels along each scanline.

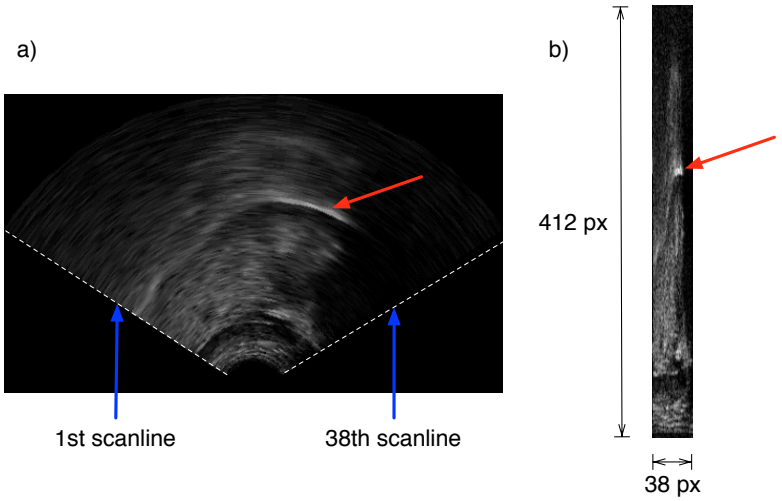

Figure 1: The difference between interpolated and raw ultrasound frames: a) An interpolated ultrasound frame. b) Raw (uninterpolated) version of the same ultrasound frame as in a). The speaker is facing right. Red arrow points to the upper surface of the tip of the tongue.

In addition to the overall frame-to-frame PD and more importantly for the current study, we also calculate the PD for individual scanlines as a function of time. This makes it possible to identify the tongue regions that initiate movement in a given token. Figure 2 shows sample analysis results. The lighter band in the middle panels around scanlines 53-63 is caused by the mandible, which is visible in ultrasound only as a practically black area with a black shadow extending behind it. This means that there is less change to be seen in most frame pairs in these scanlines than there is in scanlines which only image the tongue and its internal tissues.

As can be seen for the token on left ('caught'), the tongue starts moving more or less as a whole. In contrast the token on the right ('sheet') shows an early movement in the pharyngeal region before activation spreads to the rest of the tongue. This interpretation should be taken with (at least) one caveat: The PD does not measure tongue contour movement. This means that a part of the tongue contour might be the first to move even if the scanline based PD shows

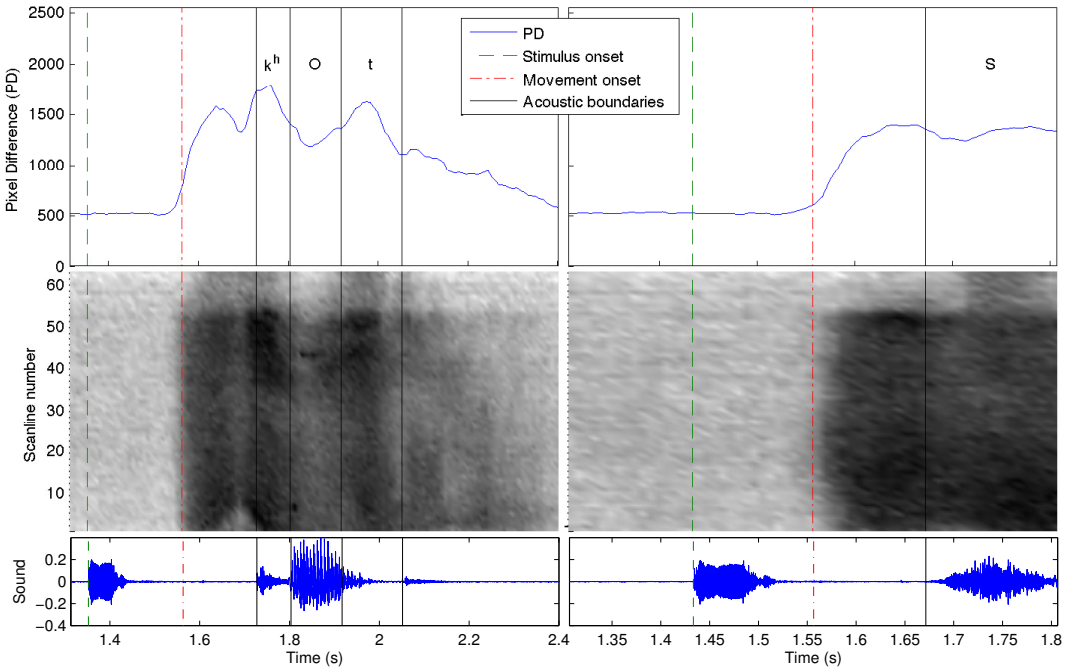

Figure 2: Two examples of regular PD and scanline based PD. The left column has a repetition word 'caught' ([kɔːt]) and the right column has the beginning of the word 'sheet' ([ʃiːt]). The panels are from top to bottom: Regular PD with annotations from acoustic segmentation, scanline based PD with the back most scanline at the bottom and the front most on top with darker shading corresponding to more change, and the acoustic waveform.

activation everywhere. This is because the PD as such measures change from frame to frame (whether on scanlines or on the whole frame). More detailed analysis will be available at the time of the conference.

References

Articulate Instruments Ltd (2008). *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd.

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

Boersma, P. and Weenink, D. (2010). Praat: doing phonetics by computer [computer program]. Version 5.1.44, retrieved 4 October 2010 from `http://www.praat.org/`.

Drake, E., Schaeffler, S., and Corley, M. (2013a). Articulatory evidence for the involvement of the speech production system in the generation of predictions during comprehension. In *Architectures and Mechanisms for Language Processing (AMLaP)*, Marseille.

Drake, E., Schaeffler, S., and Corley, M. (2013b). Does prediction in comprehension involve articulation? evidence from speech imaging. In *11th Symposium of Psycholinguistics (SCOPE)*, Tenerife.

Kawamoto, A. H., Liu, Q., Mura, K., and Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2):347 – 365.

McMillan, C. T. and Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117(3):243 – 260.

Palo, P., Schaeffler, S., and Scobbie, J. M. (2014). Pre-speech tongue movements recorded with ultrasound. In *10th International Seminar on Speech Production (ISSP 2014)*, pages 304 – 307.

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 *Proceedings of ICPhS 2015*, Glasgow, UK.

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.