# EFFECT OF PHONETIC ONSET ON ACOUSTIC AND ARTICULATORY SPEECH REACTION TIMES STUDIED WITH TONGUE ULTRASOUND

Pertti Palo*, Sonja Schaeffler, and James M. Scobbie

CASL, Queen Margaret University, Edinburgh
* ppalo@qmu.ac.uk

## ABSTRACT

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 hypothesise, therefore, that in a replication with articulatory instrumentation for the tongue, we should find the same acoustic effect, but no effect in the articulatory reaction time. As a proof of concept of articulatory measurement from ultrasound images, we report results from a pilot experiment which also extends the dataset to include onset-less syllables. The hypothesis is essentially confirmed with statistical analysis and we explore and discuss the effect of different vowels and onset types (including null onsets) on articulatory and acoustic RT and speech production.

**Keywords:** Speech reaction time, articulation, ultrasound.

## 1. INTRODUCTION

Voice keys are devices which are used to measure speech reaction time (RT) [6, 8]. Voice keys are known to have a phonetic bias [10, 12, 16]. While part of this bias is most likely to be a signal detection problem – hence the variable bias across different voice key devices – there is also a known production bias due to the first segments of the target utterance [12].

The bias caused by the production of different consonantal onsets with three different vowels following the onset consonant was studied by Rastle et al. [13]. The crucial part of their experiment was to alter the standard delayed naming instruction to allow the participants *mentally* prepare and even rehearse speaking the target utterance as much as the participants wanted. After this the participants were asked to produce the by now known utterance in a speeded trial after a randomly delayed go-signal. The participants were asked to keep their vocal tract at rest until receiving the go-signal. The results show systematic effects of consonant quality on the acoustic onset time (Figure 1). However, Rastle et al. did not record articulation at all, nor measure the bias of onsetless – i.e. vowel initial – utterances.

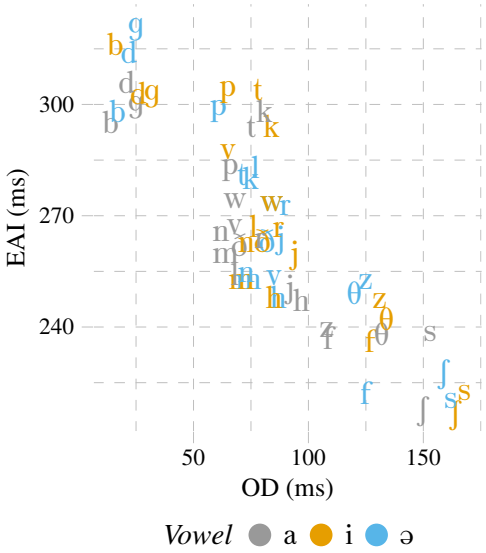


**Figure 1:** Influence of consonantal onset duration on acoustic naming latency in delayed naming [13]. OD = onset consonant duration, EAI = execution to acoustic interval i.e. acoustic RT.

Articulatory measurements of speech RTs have been reported as well [9, 14]. Kawamoto et al. [9] studied lip articulation in RT tasks using the Rastle instructions and standard delayed naming instructions. They provide systematic measurements of Articulatory onset to Acoustic onset interval (AAI) across several task conditions, but only for lip articulations. In contrast, Schaeffler et al. [14] measured the difference between lingual and labial RTs in a picture naming task. They found that on average there was no difference, and also that there was no clear link between an initial labial consonant and lip movement and lingual and tongue movement. However, they did not control for the effect of the vowel of the onset syllable.

Another aspect of speech preparation are audible prespeech sounds such as breath intake, tongue clicks and lip smacks [15]. These sounds – and in-

deed silent articulations which just fail to produce a sound – are not purely associated with speech production at a lexical or segmental level but result from general processes of preparation, such as breathing and swallowing. However, they might be more evident in normal speech than in single-word RT contexts.

Taking into account all of the results summarised above, it is clear that there is a need for a comprehensive approach tackling multiple articulators. In this study we undertake a pilot delayed naming experiment to answer the following research question: Does articulatory RT (ArtRT) measured with tongue ultrasound repeat the general trend seen in Figure 1 or is the trend produced by differences in the articulatory preparation that is needed before speech sounds can be produced? Based on the results in previous studies [9, 13, 14] it seems likely that that the differences in AcRT are produced by differences in articulatory preparation (i.e. AAI) rather than articulatory onset times.

## 2. MATERIALS AND METHODS

### 2.1. Participant

One native Scottish English male speaker with corrected to normal vision and no known hearing problems participated in the experiment (age 25 years).

### 2.2. Stimuli

We carried out a partial replication of the Rastle et al. delayed naming experiment [13] 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 /CVC/ and /VC/ type.

The words were chosen from a pronunciation lexicon of Standard Scottish English generated automatically with Unisyn [7]. The dictionary was searched for word triads, which had the same onset consonant, one each of the vowels /a,i,ɔ/, and the final consonant was a stop – i.e. one of /k,p,t/. To make it possible to get an adequate number of repetitions for each word, it was necessary to limit the number of onset consonants. The set used in this study was selected by maximising the lexical frequency of the target words.

The target words used in this 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*.

### 2.3. Ultrasound and audio recordings

The experiment was run with synchronised ultrasound and sound recording controlled with Articulate Assistant Advanced (AAA) software [2] 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 [1]. Ultrasound recordings were obtained at a frame rate of ∼83 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.

### 2.4. Procedure

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 he was ready to speak the word, he indicated so by pressing a button on a keyboard. (3) The key press activated the sound and ultrasound recording. The experimental software automatically initiated the ultrasound recording about 0.5 s after the sound recording began providing an adequate window to examine the stability of the participant's articulation before the go-signal was given. (4) After a random delay, which was uniformly distributed between 1200 ms and 1800 ms, the computer produced a go-signal – a 50 ms long 1000 Hz pure tone.

It was emphasised to the participant that it was important to keep his mouth (lips, tongue, etc.) at rest during phases (1) and (2), and when he heard the go-signal to "read the word out loud as fast and as accurately as possible keeping in mind that this is a speeded trial." The experiment consisted of four sessions on two separate days. The 48 stimuli were repeated seven times throughout the whole experiment in seven internally randomised blocks.

### 2.5. Acoustic and ultrasound segmentation

The acoustic recordings were segmented with Praat [4]. The relevant phonetic segmentation boundaries were acoustic onset (AcRT) and the acoustic offset of then initial consonant to provide onset consonant duration (OD). In addition to these the beginning of the go-signal was also annotated with Praat to pro-

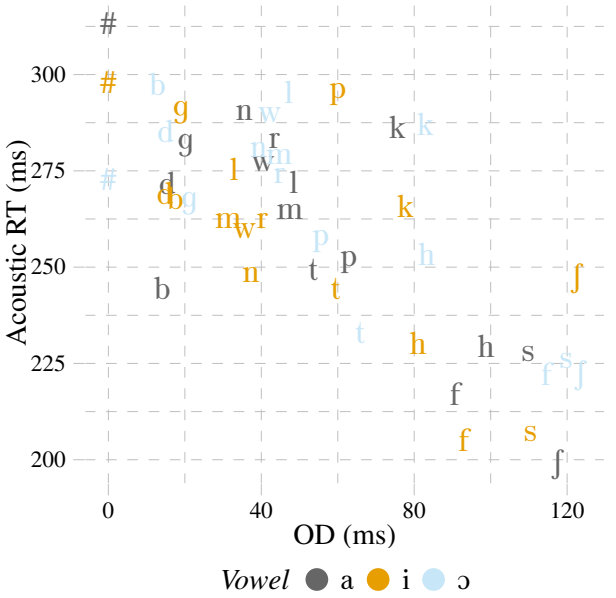


**Figure 2:** Acoustic RT as a function of onset duration (OD). Tokens without onset consonant (i.e. /VC/ words) are marked with #.

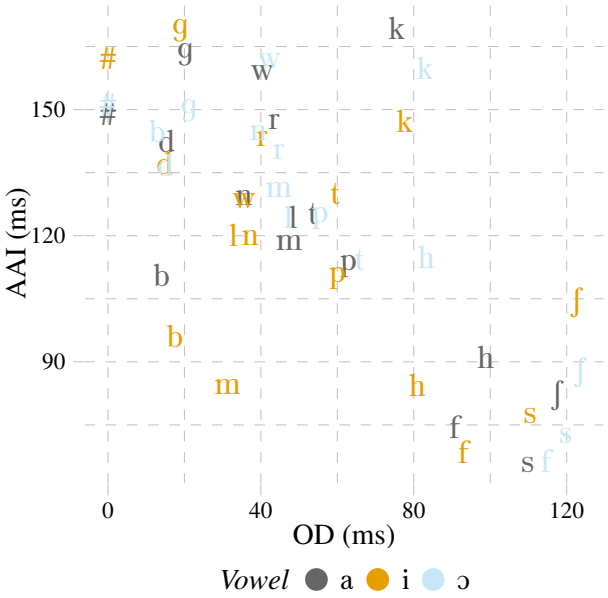


**Figure 3:** Articulatory onset to Acoustic onset Interval (AAI) as a function of onset duration (OD). Tokens without onset consonant (i.e. /VC/ words) are marked with #.

vide time of the onset of stimulus for both acoustic and articulatory RT measures.

The ultrasound recordings were segmented with AAA [2]. The only articulatory boundary was articulatory onset (ArtRT). The ultrasound videos were also inspected for movement before the go-signal to exclude tokens where the articulatory onset occurred before or during the go-signal (see next Section).

## 3. RESULTS

The recording of three tokens had been stopped before the participant had had time to pronounce the target word. These tokens as well as 12 tokens where the articulatory onset happened in less than 60 ms from or even before the go-signal onset, were excluded. As were four clear outliers: two tokens with ac_RT exceeding 450 ms, one /bɔt/ with ac_OD > 50 ms (on listening, a mispronunciation) and one /sat/ with ac_OD > 200 ms. Summary statistics are reported in Table 1.

**Table 1:** Means and standard deviations for the measured variables. Over all sample size $n = 315$.

| Variable | Mean (ms) | Stdev (ms) |
|---|---|---|
| AAI | 123.0 | 40.0 |
| AcRT | 261.6 | 45.5 |
| ArtRT | 138.6 | 33.0 |
| OD | 53.9 | 38.3 |

As a first step towards answering the research

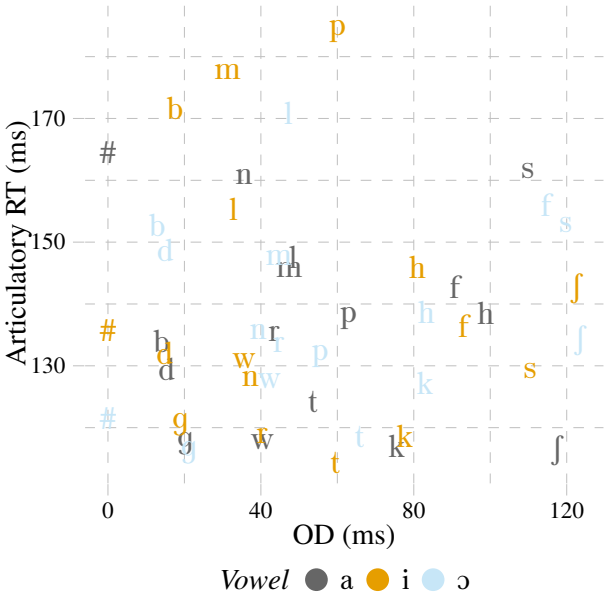


**Figure 4:** Articulatory RT as a function of onset duration (OD). Tokens without onset consonant (i.e. /VC/ words) are marked with #.

questions we made scatter plots of the averages for each onset consonant / vowel combination (Figures 2-4). As can be seen in Figure 2, the present data does indeed pattern in the same way as the original data [13]. It can also be seen that the /VC/ utterances fit the pattern.

Looking at Figures 3 and 4, we see that the for-

mer reproduces the general pattern, while the latter does not. This informal analysis seems to confirm the hypothesis made in the Introduction. To see if this result can be considered statistically significant, the data was analysed in R [11] by iteratively fitting linear models (step-up process [3]) to explain the variation in AcRT.

The iteration process identified two statistically significant models that successfully predict AcRT. Given in R formula notation they are:

(1) $log(AcRT) \sim ArtRT + OD + trial$

and

(2) $log(AcRT) \sim ArtRT + C1 + trial$,

where trial is the running trial number, C1 is the onset consonant and AcRT has been logarithmised to correct for the skewness of its distribution. Including OD and C1 in Models 1 and 2, respectively, are both highly statistically significant (P-values $\approx 0$). The $R^2$ values of Models 1 and 2 respectively are: $R_1^2 = 0.4276$ and $R_2^2 = 0.6294$, meaning that Model 2 has greater explanatory power. Finally, it should be noted that some of the variation in ArtRT can be explained with C1 (procedure as for the other models, P-value = 7.91e-06, $R^2 = 0.1541$).

## 4. DISCUSSION

It is quite convincing to see that the data (Figure 2 from a single speaker aligns so well with previous results from a larger study [13, see also Figure 1]. The average OD range is smaller in the present data: Rastle et al. had a range of roughly 0-150 ms, present data has 0-120 ms, as well as there being a less pronounced difference between the Acoustic RTs.

It is furthermore, very interesting to see that AAI as a function of OD inherits practically all of the inverse relationship evident in AcRT as a function OD. There are only two obvious groups which break the inversely proportional relationship of AAI to OD: A group of bilabial consonants in the lower left corner and all /kVC/ words on the opposite side of the main trend. Bilabial consonants are likely to have an early start with lip articulation. Considering this would probably increase the AAI for /bak/, /bit/ and /mit/ and thus move them into better alignment with the general AAI/OD trend.

On the other hand the /kVC/ words are probably right where they should be for this speaker. The AAI of the /kVC/ words clearly aligns with those of /gVC/ words. At the same time the long aspiration produced by the speaker lengthens OD considerably and moves the /kVC/ words away from the general trend. This does, however, leave open the question of why /pVC/ and /tVC/ words do align with the trend. Analysing data from more participants is likely to shed light on this unvoiced plosive pattern. More importantly, it should be noted that, the event of interest (vowel onset) is overshadowed by the aspiration and the event itself really belongs to the articulatory domain and thus should be measured based on articulation rather than acoustics.

So what about the /VC/ words? We have interpreted them as having a consonantal onset length of 0 ms. This claim is backed up by the patterns in Figures 2 and 3. Further, more rigorous, evidence is provided by extremely good fit of the statistical models (Models 1 and 2). It might seem at first counter intuitive that acoustic RTs would be long for /VC/ words in comparison to /CVC/ words. However, comparing our results with results from rhythmical speech alignment studies [5], there is a certain logic to the situation. When a listener is asked to align words to a regular rhythm they tend do so at a point which lies within the initial consonant – not at utterance onset. Perhaps we are seeing the same mechanism at work here.

## ACKNOWLEDGEMENTS

We wish to thank Steve Cowen for assistance with the ultrasound recordings and Professor Alan Wrench for advice on post-processing of the data.

## 5. REFERENCES

[1] Articulate Instruments 2008. *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd.

[2] Articulate Instruments 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

[3] Baayen, R. H. 2008. *Analyzing Linguistic Data, A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

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

[5] Browman, C. P., Goldstein, L. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45, 140 – 155.

[6] Cattell, J. M. 1886. The time taken up by cerebral operations. *Mind* 11, 220 – 242.

[7] Fitt, S. Unisyn lexicon release. version 1.3, retrieved 26 November 2014 from http://www.cstr.ed.ac.uk/projects/unisyn/.

[8] Kapusinski, D. A., Rosenquist, H. S. 1973. A brief history of the voice key. *Proceedings of the Annual Convention of the American Psychological Association* Volume 8, 945 – 946.

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

[10] Kessler, B., Treiman, R., Mullenix, J. 2002. Phonetic biases in voice key response time measurements. *Journal of Memory and Language* 47, 145 – 171.

[11] R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[12] Rastle, K., Davis, M. H. 2002. On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance* 28(2), 307 – 314.

[13] Rastle, K., Harrington, J. M., Croot, K. P., 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.

[14] Schaeffler, S., Scobbie, J., Schaeffler, F. 2014. Measuring reaction times: Vocalisation vs. articulation. *Proceedings of 10th ISSP* 383 – 386.

[15] Scobbie, J. M., Schaeffler, S., Mennen, I. 2011. Audible aspects of speech preparation. *Proceedings of 17th ICPhS* Hong Kong. 1782 – 1785.

[16] Tyler, M. D., Tyler, L., Burnham, D. K. 2005. The delayed trigger voice key: An improved analogue voice key for psycholinguistic research. *Behavior Research Methods* 37(1), 139 – 147.