

Ultrasound investigation of tongue movements in syllables with different onset structure

Tanja Kocjančič

Queen Margaret University

E-mail: tkocjancic@qmu.ac.uk

Abstract

This study is an attempt to describe syllables with different onset structure not only in terms of durational changes but using ultrasound also in terms of the distance the tongue travels over a syllable by using ultrasound and to compare the ratio between the two parameters, expressed as speed. Results indicate that both measures increase with an increasing number of onset segments but not to the same degree for all targets. Therefore speed was not constant over all of them. Additionally, type of onset constituent greatly influenced the three parameters and there were large between-speaker similarities in case of durational changes.

1 Introduction

Ultrasound is a safe and non-invasive imaging technique that enables visualisation of the tongue inside the mouth during speech without placing any obstructions on the tongue. Ultrasound images are recorded by placing the probe under the speaker's chin. The probe emits ultrasound waves which are reflected at a boundary between two media of different density (i.e. tongue/air or tongue/bone boundary). Based on the time lapse between incident and reflected wave, a point of reflection is calculated and an image is created at that point. Since the reflection at the air boundary is almost 100 % it cannot image a raised tongue tip with an air pocket below it. Nevertheless, ultrasound allows observation of tongue during speech and therefore provides a good tool for investigating most aspects of tongue movement.

Properties of differently structured syllables have been investigated in several studies and one of the

general conclusions is that syllable duration increases with an increasing number of onset segments [5, 2]. The increase in duration, however, is not simply a sum of the individual durations of constituent segments. Due to coarticulation individual segments influence each other's articulatory realisation and consequently duration. Most of the studies report shortening of segments in clusters. However, only average group results and not measurements of individual segments are usually reported. Detailed inspection reveals that segments in cluster onsets can be either shorter, longer or the same as in single onset position depending on the number of the onset segments, the position within onset, a segment's intrinsic duration and the identity of adjacent segments. Since the onsets in the present study are composed of the consonants /p/, /s/, and /l/, some past findings regarding these segments are described.

It has been observed that in English a singleton syllable-initial /s/ is the longest of the three consonants, followed by /p/, and then by /l/, which is the shortest [8]. When these segments are part of an onset cluster they each change in their own specific way. Initial /s/ is shorter when followed by a stop [8, 7] and longer [5], equal [2] or shorter [8] when followed by /l/ [5, 6, 9]. Shortening of initial /s/ when followed by a consonant was also found to occur in Italian [4].

English /p/ in an initial onset position lengthens when followed by a voiced consonant [8, 9], and shortens as a second segment in a cluster [8, 7]. In Italian however, the duration of stop consonants is not affected by the onset size or by the position within the onset [4].

In contrast to the other two consonants /l/ can not be the first segment of a consonant cluster in En-

glish. In the second position of a cluster // shortens when preceded by a fricative [8, 6, 9] and either shortens [8] or stays the same when it follows an unvoiced stop [9].

The three segments do not differ only in their durations but also in their articulation. The relationship between these two parameters is not completely clear. O'Shaughnessy [8] stated that shorter durations of segments in a cluster result from shorter distances the articulators have to travel in the realisation of a cluster. Umeda [9], on the other hand, relates the duration of a consonant to the articulator and/or type of gesture shared between the consonant and its adjacent consonant. Duration of the consonant is different when the gesture is overlapping than when it is conflicting. Sharing a gesture additionally prevents consonant's shortening, and sharing an articulator (e.g. tongue in /st/) lowers the variance of timing. O'Shaughnessy [8] reported that in French consonants were shorter in a cluster of two segments sharing place of articulation and longer when their place of articulation differed. The same was not observed in a study of English [1].

The main difference between the segments /p/, /s/ and //, from the point of tongue investigation, is that /p/ is a non-lingual consonant while /s/ and // are both lingual. As such, /p/ does not have any tongue movement necessary for its realisation but the other two require the correct tongue position. Fricatives have to restrict the tongue dorsum position to achieve the necessary tongue constriction while articulation of laterals is less constrained [3].

The aim of this study was to investigate how tongue movements are affected by changing the type and number of syllable onset segments. Tongue movements are described by the tongue's distance of travel over an utterance, the duration of an utterance and the ratio between the two measurements, expressed as speed of the tongue's distance of travel. The following hypotheses were tested in this study: (i) the addition of either a lingual or non-lingual consonant to the syllable onset increases the duration of the syllable, (ii) the addition of a lingual consonant increases the distance the tongue travels over a syllable while the addition of non-lingual consonant does not. Therefore, for example, /a lay/ will have shorter duration than /a play/ but they will have similar distance, resulting in higher distance to duration ratio (speed) for /a lay/ than for /a play/.

2 Methodology

2.1 Speakers

Ultrasound data from 10 native English female speakers, aged between 19 and 30 years, was analysed in this study. Speech material consisted of six mono-syllabic real English words: "pay", "say", "lay", "play", "slay", "splay".

2.2 Speech material and recording

A midsagittal view of the tongue was recorded with Concept M6 (Dynamic Imaging) ultrasound with a frame rate of 30fps. A special helmet was used to fix the probe under the speaker's chin. Participants repeated each of the words five times in a frame sentence "a [word] today". Both ultrasound and audio signals were recorded at the same time using Articulate Assistant Advanced which allows temporal synchronisation of the two signals.

2.3 Data analysis

Articulate Assistant Advanced was used for annotating and tracing tongue contours on the recorded ultrasound images. All the reported results are for the "a [word]" part of the recording. The distance of tongue travel in a target utterance was calculated as a sum of average nearest neighbour distances (aNND) between every pair of consecutive tongue contours of the utterance. aNND is an average of all the nearest neighbour distances measured between the points on the two contours of a pair. Duration was measured from audio signal and speed was calculated as the ratio of distance and duration. Stat. sig. was tested using a non-parametric Friedman's ANOVA with a post-hoc Wilcoxon signed-rank test and Bonferroni correction.

3 Results and discussion

Group results for the ten speakers are presented in Figure 1. As can be seen in Figure 1a, the duration of the target increases with increasing number of onset segments and the target with single onset /s/ is the longest of the single onset ones and the one with // the shortest. These two observations support findings of previous studies. Additionally, duration

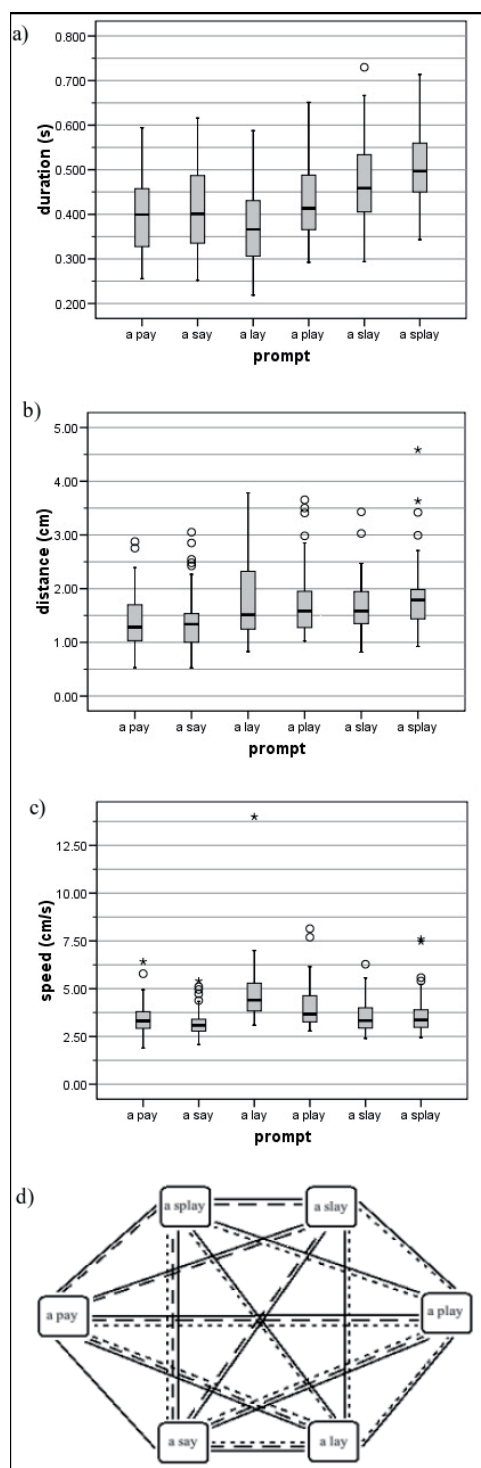


Figure 1: Duration (a), distance of tongue's travel (b), speed of tongue's distance of travel over the target (c) and stat. sig. ($p < 0.0033$) between pairs of targets (d; dashed line = tongue's distance of travel, solid line = duration, dotted line = speed.)

is significantly different between all targets except "a pay"- "a say" and "a say"- "a play" pairs (solid lines in Figure 1d; lines between the pairs of targets represent stat. sig. difference) showing a strong effect of the type of onset segments on the total duration. However, despite the difference there is a similarity in the distribution of measurements of the six targets as they all show similar variation. Moreover, similarities were observed between individual speakers as well. They showed very similar patterns of increasing duration, with nine out of ten having the shortest duration for "a lay" and longest for "a splay".

The second measured parameter, distance of travel, has revealed the same effect of the number of onset segments as duration; as the onset increases, the distance travelled increases as well (Figure 1b). The measurement was the shortest in "a pay" and "a say", which are not significantly different from each other but are significantly shorter than all the other targets (Figure 1d). This result was expected for "a pay" since /p/ is not a lingual segment but not for "a say". A possible explanation is that although /s/ has a lingual component in its realisation, the movement from the /ə/ with a tongue positioned more or less in the centre of the mouth, is very small. Additionally, this result could be affected by the limitation of ultrasound which does not image a raised tongue tip. The greatest movement could be done with the tip of the tongue, and not recorded, while the body stays in a position similar to /ə/. Another observation regarding this parameter is that "a lay" had similar distance travelled to targets with cluster onsets (Figure 1d). This is the result of the /l/ having the greatest lingual component of the three consonants and the others having less influence on the total distance travelled over targets when in clusters with /l/. In fact, /p/ should not have any influence and /s/ should show some. Following this, it was expected that the tongue would travel less distance over "a play" than over "a slay" but the results showed no difference. Further inspection of relevant ultrasound frames has revealed that in case of the /pl/ cluster the tongue movement for /l/ is not restricted by /p/ and can be thus fully realised. On the other hand, in case of /sl/, tongue movement for /s/ restricts movement for /l/ and gives the tongue less space to cover in the articulation of the later. Similarly, in the case of /spl/ the tongue articulates /s/, stays in the same position dur-

ing /p/ and then realizes /l/, which is consequently again restricted by the last preceding lingual consonant (/s/). The result is a similar distance travelled over a /sl/ than over a /spl/ target (Figure 1d). A very notable difference between the six targets is the distribution of the measurements, with “a lay” being the most variable and “a splay” the least (based on interquartile range). This can be due to singleton /l/ being the least constrained in its articulation of the three consonants [4] and possibly the most constrained of all segments when in cluster. Another difference was observed between speakers when comparing the increasing pattern of distance travelled over the target. Individual speakers showed very different patterns with only one common factor; most of them had shorter travelled distances in targets with single onsets than in those with clusters. Such a result may be influenced by the differences in the size of the tongue and oral cavity between speakers.

As already seen, both duration and the distance travelled showed an effect of the increasing number of onset segments but their ratio, expressed as speed, did not (Figure 1c). Although the change of both parameters was in the same direction (increase) the extent of each of them was not the same for all targets. The exceptions were “a lay” and “a play”, which were significantly faster from almost all the others (similar speed over “a pay” and “a play”) with “a lay” being even faster than “a play” (Figure 1d). These results, however, simply reflect the measured changes in distance travelled and duration. Targets with shortest durations and longer distance of travel had higher distance to duration ratio and thus higher speed of tongue movement over the entire target. This parameter has also shown only one common characteristic between the ten speakers: speed was the highest in “a lay” for seven of them and the second highest for the remaining three.

4 Conclusion

The data has confirmed the expected result of duration increasing with increasing number of onset segments, either lingual or non-lingual, but not the expectation that the tongue’s distance of travel would increase only with the addition of a lingual consonant. Distance showed a greater effect of individual segments, e.g. /l/ having the greatest tongue

movement of the three onset consonants as a singleton and the greatest effect on the tongue’s distance of travel over a cluster onset target. Overall, both measured parameters increased with the increasing number of onset segments but not always to the same degree causing their ratio, speed, to be constant for some targets but not for all, which reflects the characteristic of individual consonants and their combinations in clusters. Comparing individual speaker data revealed that duration was the least speaker dependant and had the most similar variability across targets of the three parameters.

5 Acknowledgements

Many thanks to my supervisor, Dr Nigel Hewlett. Financial support provided by the Marie Curie EdSST programme.

References

- [1] T. Crystal and A. House. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4):1553–1573, 1988.
- [2] T. Crystal and A. House. Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, 88(1):101–112, 1990.
- [3] E. Farnetani. *Coarticulation and connected speech processes*. Blackwell Publishers, 1997.
- [4] E. Farnetani and S. Kori. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 5(1):17–34, 1986.
- [5] S. Greenberg, H. Carvey, L. Hitchcock, and S. Chang. Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics*, 31(3):465–485, 2003.
- [6] M. Haggard. Abbreviation of consonants in English pre- and post-vocalic clusters. *Journal of Phonetics*, 1(1):9–24, 1973.
- [7] D. Klatt. The duration of [s] in English words. *Journal of Speech, Language and Hearing Research*, 17(1):51–63, 1974.
- [8] D. O’Shaughnessy. Consonant durations in clusters. *IEEE Transactions on Acoustics Speech and Signal Processing*, 22(4):282–295, 1974.
- [9] N. Umeda. Consonant duration in American English. *The Journal of the Acoustical Society of America*, 61(3):846–858, 1977.