Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

This postprint was originally published as:

Mousikou, P., Strycharczuk, P., Turk, A., & Scobbie, J. M. (2021). Coarticulation across morpheme boundaries: An ultrasound study of past-tense inflection in Scottish English. *Journal of Phonetics*, *88*, Article 101101. https://doi.org/10.1016/j.wocn.2021.101101

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, nontransferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.

Provided by:
Max Planck Institute for Human Development - Library and Research Information
library@mpib-berlin.mpg.de

Coarticulation across morpheme boundaries:
An ultrasound study of past-tense inflection in Scottish English.

Petroula Mousikou[a,b,c]*, Patrycja Strycharczuk[d]*, Alice Turk[e], & James M. Scobbie[a]

[a]Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University, Queen Margaret University Drive, Musselburgh, EH21 6UU, United Kingdom

[b]Department of Psychology, Royal Holloway, University of London, United Kingdom

[c]Max Planck Institute for Human Development (MPIB), Berlin, Germany

[d]Linguistics and English Language, The University of Manchester,
Oxford Rd, Manchester M13 9PL, United Kingdom

[e]Department of Linguistics and English Language, The University of Edinburgh,
Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, United Kingdom

Author Note

*These authors contributed equally to this work.

*Corresponding authors: Petroula Mousikou, Max Planck Institute for Human Development, Lentzeallee 94, 14195, Berlin, Germany. Email: bettymousikou@gmail.com, and Patrycja Strycharczuk, Linguistics and English Language, The University of Manchester, Oxford Rd, Manchester M13 9PL, United Kingdom. Email: pstrycharczuk@gmail.com

Word Count: 8,781

Abstract

It has been hypothesized that morphologically-complex words are mentally stored in a decomposed form, often requiring online composition during processing. Morphologically-simple words can only be stored as a whole. The way a word is stored and retrieved is thought to influence its realization during speech production, so that when retrieval requires less time, the articulatory plan is executed faster. Faster articulatory execution could result in more coarticulation. Accordingly, we hypothesized that morphologically-simple words might be produced with more coarticulation than apparently homophonous morphologically-complex words, because the retrieval of monomorphemic forms is direct, in contrast to morphologically-complex ones, which might need to be composed online into full word forms. Using Ultrasound Tongue Imaging, we tested this hypothesis with nine speakers of Scottish English. Over two days of training, participants learned phonemically identical monomorphemic and morphologically-complex nonce words, while on the third consecutive testing day, they produced them in two prosodic contexts. Two types of articulatory analyses revealed no systematic differences in coarticulation between monomorphemic and morphologically-complex items, yet a few speakers did idiosyncratically produce some morphological effects on articulation. Our work contributes to our understanding of how morphologically complex words are stored and processed during speech production.

Keywords: speech production, morphology, coarticulation, Ultrasound Tongue Imaging

1.0 Introduction

Speakers can effortlessly compose new words on the basis of predictable changes in existing words (e.g., *snack* + *able* = *snackable*, “suitable to be eaten as a snack”). Morphological productivity in speech production was empirically evidenced in Berko’s (1958) seminal “wug” test, where children were presented with a nonce word that was associated with an unfamiliar object (e.g., “This is a *wug*.”) and were asked to produce its plural form via an elicitation task (e.g., “Now there are two of them. There are two [*wugs*].”). Since then, Berko’s results have been replicated in a number of studies, leading to the widely held assumption that speakers are uniformly capable of analyzing the internal structure of words, decomposing existing morphologically-complex ones into their constituent morphemes, which can then be used to form novel complex words. Accordingly, some theoretical models of morphological processing posit that regular morphologically-complex words, even frequently used ones, are always stored in their decomposed form and computed online (e.g., Clahsen, 1999; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Pinker & Ullman, 2002; Prasada & Pinker, 1993).

It has been proposed that the way a word is stored and retrieved can influence its realization during speech production (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Pierrehumbert, 2002; Pluymaekers, Ernestus, & Baayen, 2005). For example, frequent and predictable words, which are thought to be retrieved more easily, tend to be produced with less articulatory effort (see Bürki, 2018, for a review, and Tomaschek, Arnold, Bröker, & Baayen, 2018). Bell et al. (2009) proposed a mechanism that coordinates the pace of higher-level processes (e.g., word retrieval) and the execution of the articulatory plan. Accordingly, when word retrieval requires less time, the execution of the corresponding articulatory plan will be faster (see Tomaschek et al., 2020; Tomaschek, Tucker, Fasiolo, & Baayen, 2018; and Aylett & Turk, 2004, for alternative views of these effects). Faster articulatory execution is likely to result, among other things, such as lenition/weakening, faster speed of movement, and shorter "hold" phases, in more coarticulation (Gay, 1981; Matthies, Perrier, Perkell, & Zandipour, 2001), that is, more overlapping of speech segments.

On the assumption that there is a connection between the time required for word retrieval and the speed of articulatory execution (Bell et al., 2009), which could then influence coarticulation, we would predict that a morphologically-simple word might be produced with more coarticulation than an apparently homophonous morphologically-complex word with similar characteristics. This is because the retrieval of monomorphemic words, which can only be stored and retrieved as a whole, should require less time than the retrieval of comparable morphologically-complex words, which might need to be composed online into full word forms. Also, given the nature of the composition process (e.g., stem + affix), we would expect differences in coarticulation between the two types of items to be observed at segments that correspond to morphological boundaries in morphologically-complex words. We investigated this idea in adult speakers of English using Ultrasound Tongue Imaging. By studying adult coarticulation patterns in speech production, we could gain an insight into the nature of lexical storage in the mental lexicon.

1.1 Previous empirical evidence

Previous studies suggest that there may be a systematic relationship between morphological complexity and the degree of coarticulation.

Cho (2001) compared utterances of monomorphemic and bimorphemic segmentally-identical words in Korean (e.g., /sapi/, private expense, vs. /sap-i/, *shovel* + Nom.) using Electromagnetic Articulography and Electropalatography. Intergestural timing for the interval between the middle of the labial constriction and the end of the next vowel was found to be more variable in bimorphemic items. Similarly, there was more variability in palatalization in

Consonant-Vowel sequences that crossed a morpheme boundary (e.g., /ti/ in /mat-i/, *the eldest*) compared to when there was no morphological boundary (e.g., /ti/ in /mati/, *knot*). However, Cho also reported more gestural overlap in heteromorphemic /ti/ sequences than in tautomorphemic ones. A tentative explanation for the latter result is that increased flexibility in gestural timing for heteromorphemic forms may result in more overlap. In sum, these findings show that articulations for segmental sequences spanning morpheme boundaries might differ from those executed within a single morpheme, even when the segments in the two environments are phonemically identical.

Song, Demuth, Shattuck-Hufnagel, and Ménard (2013) used Ultrasound Tongue Imaging to compare coda clusters (i.e., /ks/) in the monomorphemic word *box* versus the bimorphemic word *rocks*. The study was carried out with adults and 2-year old children. Results showed a perseveratory coarticulation effect of /k/ on /s/ in *box* and an anticipatory coarticulation effect of /s/ on /k/ in *rocks*. This led the authors to conclude that the primary articulatory target for /ks/ in monomorphemic *box* must have been the /k/, whereas the primary articulatory target for /ks/ in bimorphemic *rocks* must have been the plural morpheme /s/. Even though it is difficult to generalize these results, given that the experimental materials were limited to a single item pair (i.e., *box* vs. *rocks*), this finding has been taken as evidence for a morphological effect on coarticulation.

Taken together, the available empirical evidence offers support for the idea that morphological complexity might modulate coarticulation. However, the scarcity of articulatory studies in this domain and the associated difficulty to generalize the findings (given the different aspects of coarticulation being examined in each study, the different languages being investigated, and the limited number of experimental materials that were used) warrants a more controlled investigation into this topic.

2.0 Present study

Lexical frequency effects on the retrieval of morphologically-complex words have been reported by previous studies (e.g., Baayen, 1992; Hay, 2003). Hay (2003), in particular, argued that derived words that are more frequent than their base form (e.g., *swiftly* vs. *swift*) are more likely to be retrieved as a whole, whereas derived words that are less frequent than their base form (e.g., *softly* vs. *soft*) are more likely to be decomposed. To test our hypothesis, we needed to ensure that the mechanism supporting morphological (de)composition would be likely to occur during the processing of morphologically-complex words in our study. For this reason, we opted for using nonce words, which have no prior lexical frequency and need to be learned and created on the fly (when inflected), as it is the case with the wug-test. Furthermore, English frequent words tend to be produced with more contracted vowels (e.g., Munson & Solomon, 2004) and are more prone to t/d deletion (Jurafsky, Bell, Gregory, & Raymond, 2000), while frequent syllables are often produced with more coarticulation than infrequent syllables (Herrmann, Whiteside, & Cunningham, 2009). Hence, the use of nonce words, and more specifically, of homophonous (in terms of phoneme sequence) nonce-word pairs, further allows us to control for the influence of confounding lexical and sublexical variables on coarticulation.

Accordingly, 10 speakers of Scottish English produced phonemically identical pairs of nonce words that were either monomorphemic (e.g., *zord* /zɔrd/), corresponding to unfamiliar objects, or bimorphemic (e.g., *zorred* /zɔrd/), corresponding to the past-tense form of unfamiliar verbs in infinitive form (i.e., *to zor* /zɔr/).[1] Importantly, the regular past-tense inflection *-ed* applies to 86% of the 1000 most common English verbs (Pinker, 1999), and is readily realized in

---

[1] As we note in the Results section, the quality of the /r/ in the coda differed across speakers. Note also that Scottish English is a rhotic accent.

coda clusters word-finally (e.g., cared /kɛrd/).

The critical comparison in the present study is the articulation of the consonant cluster /rd/ in the two types of items. Several studies that investigated the articulation of consonant clusters in adults have shown robust coarticulatory effects (e.g., Byrd, 1996; Davidson, 2005; Hardcastle, 1985), while coarticulation in consonant clusters is a well-studied phenomenon (for a review, see Recasens, 2018). The choice of the /rd/ cluster was motivated by the fact that liquids are known to exert a strong coarticulatory influence on other sounds (West, 1999). Furthermore, the postvocalic /r/ in Scottish English has a detectably different articulation compared to /d/. It is typically realized as either a bunched or retroflex approximant (Lawson, Scobbie, & Stuart-Smith, 2011). Bunched or retroflex articulations share a characteristic raising of the pre-dorsum, and a complex tongue shape that is markedly different from a plain coronal /d/ (Lawson, Stuart-Smith, & Scobbie, 2018). We reasoned that the articulatory differences between /r/ and /d/ would allow us to detect and quantify coarticulation more easily, thus making putative differences between monomorphemic and bimorphemic items easier to recover.

Speakers' tongue movements were tracked using Ultrasound Tongue Imaging. In the present study, we focused exclusively on articulation, rather than on both articulation and acoustics (as is usually the case in most ultrasound studies), because previous studies had already failed to find significant acoustic differences between the past-tense morpheme /t/ and /d/ and non-morphemic /t/ and /d/ (e.g., Losiewicz, 1992, Seyfarth, Garellek, Gillingham, Ackerman, & Malouf, 2018; Zimmermann, 2016). In this respect, it is worth noting that articulatory techniques like Ultrasound Tongue Imaging have previously discovered "covert" phonological contrasts (i.e., ones not found in the acoustic signal) via the detection of contrasting articulatory gestures (e.g., Gick, Michelson, & Radanov, 2006).

2.1 Hypotheses regarding coarticulation effects

The present study seeks to test whether morphologically-simple words might be produced with more coarticulation than apparently homophonous morphologically-complex words. Articulatory studies generally show that intergestural timing (i.e., the temporal organization of adjacent articulatory gestures) is more stable (or less variable) word-internally than across word boundaries (e.g., McClean, 1973; Hardcastle, 1985; Holst & Nolan, 1995; Byrd, Kaun, Narayanan, & Saltzman, 2000); within a syllable than across syllable boundaries (Byrd, 1994); and within a morpheme than across morphological boundaries (Cho, 2001). Observations about stability of gestural timing (or variability thereof) allow us to make detailed predictions about coarticulation. In particular, stable intergestural timing is hypothesized to reflect greater 'bonding strength' or degree of cohesion between certain gestures (Browman & Goldstein, 2000). In line with our reasoning then, tautomorphemic gestures should be more coarticulated than heteromorphemic gestures. Note, however, that while this hypothesis follows from the above-mentioned theories, it is not supported by Cho's (2001) observation of relatively *less* coarticulation across morphological boundaries.

Drawing on recent work in this domain (Rubertus & Noiray, 2020), which shows great similarity between anticipatory and perseverative coarticulation due to the temporal overlap of articulations, we do not make any specific predictions about the direction of potential coarticulation effects in our study.

3.0 Method

3.1 Participants

Ten monolingual native speakers of Scottish English (3 males and 7 females) participated in the study for monetary compensation. Participants were between 18 and 33 years old ($M = 22$, $SD = 5$), had normal or corrected-to-normal vision, and reported no hearing, reading, speech, or language difficulties. Two speakers yielded a very low-quality tongue image, for another speaker the probe moved during the experimental session, while for another, the probe was positioned in an unusual way, resulting in a tilted tongue shape. The data from these four speakers were excluded from parts of the articulatory analyses (more details are provided in the corresponding sections of the analyses). The study was approved by the ethics committee of Queen Margaret University and participants provided written consent prior to participating in the study.

3.2 Materials

The experimental stimuli consisted of 16 nonce words. Half of these words corresponded to nouns with a CVCC structure (*dard, gord, lerd, mord, sard, tord, vard, zord*) and were associated with eight unfamiliar objects consisting of black-and-white line drawings taken from Rastle, McCormick, Bayliss, and Davis (2011). The other half corresponded to verbs in the infinitive form (*to dar, to gor, to ler, to mor, to sar, to tor, to var, to zor*) and were associated with eight unfamiliar definitions of actions. All pictures had the same width and height (550 pixels). The unfamiliar objects and definitions of actions that corresponded to the nonce nouns and verbs, respectively, are shown in the Appendix. Since participants had to learn a significant number of nonce words (N = 16), we focused on a single suffix, that is, the past-tense morpheme -*ed*, to prevent cognitive overload.

3.3 Design

Participants underwent a training phase and a test phase. The training phase consisted of three stages.

3.3.1 First training stage. During this stage, participants were told that they had to learn some new words, which they would be asked to produce later. On each trial, participants saw on the computer screen a picture of an unfamiliar object (e.g., new animal or fruit or vehicle) or the definition of an action (e.g., “to fake a smile”) and heard their corresponding unfamiliar lexemes (e.g., *zord* and *to zor*, respectively) via headphones. The audio presentation of the nonce nouns and verbs used recordings of a female native speaker of Scottish English who was naïve to the purpose of the study. Participants never saw a written form of the nonce words. They studied each picture for as long as they needed for learning, and controlled the time at which the next picture was presented with a button press.

Training on the nonce nouns and verbs took place in separate blocks. Half of the participants were trained first on the nonce nouns and the other half on the nonce verbs. Within each block, the items were presented to each participant in a different random order. Also, the associations of nouns and verbs with their corresponding objects and definitions of actions were randomized across participants. This meant that the item *zord*, for example, may have corresponded to an unfamiliar animal for some participants, but to an unfamiliar fruit or vehicle for others (see Appendix). Similarly, the item *to zor* corresponded to different definitions of actions across participants.

3.3.2 Second training stage. During this stage, we assessed whether participants remembered the nonce words they had just learned. Pictures of objects that corresponded to the nonce nouns, or definitions of actions that corresponded to the nonce verbs were presented visually again in a random order, while participants were asked to provide their names. If participants produced the name of the newly-learned nouns and verbs correctly, the word "correct" was displayed on the screen; if they produced it incorrectly, the word "incorrect" was displayed on the screen. In both cases, the correct name of the item was also presented aurally via the headphones.

3.3.3 Third training stage. During this stage, we assessed again whether participants remembered the nonce words they had learned during the first and second training stages. This time, there was an incomplete sentence below each new object or definition of action. Participants were told that their task was to complete the sentence with the missing name, while reading it out loud accurately and without hesitating. For the nouns, the following example was given: If the picture depicts a bird, the sentence below will read “It’s a ___ again.” Your task is to say out loud “It’s a *bird* again.” Hence, in this case, participants were expected to produce a sentence like “It’s a *zord* again.” For the verbs, the procedure slightly differed. Participants were told that the sentence would require them to put the newly-learned verb in the past-tense form. Accordingly, the following example was given: If the definition refers to the verb “talk”, the sentence below will read “Yesterday, Tessa ___ again”. Your task is to say “Yesterday, Tessa *talked* again”. Hence, in this case, participants were prompted to produce the past-tense form of the learned verb (e.g., if a participant had learned that *to zor* is “to fake a smile” and the latter definition appeared on the screen, she/he was expected to produce “Yesterday, Tessa *zorred* again.”). Hence, the produced pairs of nonce words in the monomorphemic (noun) and complex (verb) conditions were phonemically identical (i.e., /zɔrd/).

The first training stage was administered five times. Then the second and third training stages were administered. Subsequently, the second and third stages were administered again. This procedure occurred three times. Therefore, participants were exposed to the nonce words 15 times and were asked to produce them another 12 times. They also received feedback on their response half of the times they produced them. Training occurred over the course of two consecutive days, with all three training stages taking place on both days to ensure long-term memory integration of the newly-learned words (Dumay & Gaskell, 2007).

3.3.4 Test phase. On the third consecutive day, participants were tested. However, they were told first that they would see the pictures of the objects and the definitions of the verbs they had previously learned, and that below each picture or definition there would be three incomplete sentences. Their task was to complete these sentences with the missing name and read them out loud one after the other, accurately, and without hesitating. For the nouns, the following example was given: If the picture depicts a bird, the sentence below will read “A ___. It’s a ___. It’s a ___ again.” Your task is to say out loud “A *bird*. It’s a *bird*. It’s a *bird* again.” For the verbs, the following example was given: If the definition refers to the verb “talk”, the sentence below will read “To ___. Yesterday, Tessa ___. Yesterday, Tessa ___ again.” Your task is to say “To *talk*. Yesterday, Tessa *talked*. Yesterday, Tessa *talked* again.” Before testing began, participants practiced uttering the above phrases and sentences with eight pictures (line drawings) of real objects that were easy to name, and eight real verbs that were presented in the infinitive form.

Testing on the nonce words took place using Ultrasound Tongue Imaging. Word-final clusters with a /t/ or /d/ in final position are sometimes pronounced without the stop segment (e.g., Labov, 1968; Wolfram, 1969). To ensure that the final segment of the monomorphemic and complex items, which is critical for the hypotheses tested in the present study, would be produced, both a sentence-final and a pre-vocalic context (provided by the sentences ending in “again”) were used. Nouns and verbs were tested in separate blocks, but blocks were presented to each participant in the same order as during training. Speakers 1, 3, 5, 7, and 9 were tested on the nouns first and speakers 2, 4, 6, 8, and 10 were tested on the verbs first. We chose to present the monomorphemic (i.e., nouns) and complex (i.e., verbs) items in separate blocks, rather than in one mixed block, to avoid triggering participants’ awareness of the homophonous nature of the items in the two conditions. Had participants realized that *zorred* and *zord* sound identical, they could have potentially tried to consistently hyperarticulate the items in one condition to make them sound distinct from the similar-sounding items in the other condition (Lindblom, 1990). This could lead us to a confound, because a difference between the two conditions in this case would have been due to factors other than morphological structure. Within each block, the items were presented to each participant in a different random order. Each item was presented six times across the session, resulting in a total of 96 recordings per speaker.

3.4 Apparatus and Procedure

During training (Days 1 and 2), stimulus presentation was controlled by DMDX software (Forster & Forster, 2003). Participants were trained individually in a quiet room, seated approximately 60 cm in front of a CRT monitor. The pictures of the objects and the definitions of the verbs appeared on a white background in the center of the screen. Each training session lasted approximately 45 minutes and all participants achieved 100% accuracy already by the end of the first day of training. During testing (Day 3), stimulus presentation was controlled by Articulate Assistant Advanced (AAA) software, version 2.15 (Articulate Instruments Ltd., 2014). The same software was used to collect time-synchronized articulatory and audio data. Tongue movement data were captured using a high-speed Sonix RP ultrasound system (to nearest integer, Frame Rate = 121 fps, Scanlines = 63, Pixels per Scanline = 412, Field of Vision = 135°, Pixel offset = 51, Depth = 80 mm). The ultrasonic probe was placed under the participant's chin and was stabilized with a headset (Articulate Instruments Ltd., 2008). The audio data were captured using a lavalier Audio-Technica AT803 condenser microphone connected to a synchronisation unit (Articulate Instruments Ltd., 2010). More technical detail about the system is described in Wrench and Scobbie (2016). At the beginning of each experimental session, participants were recorded swallowing water, in order to image the hard palate, and biting on a piece of plastic, in order to image the occlusal plane (Scobbie, Lawson, Cowen, Cleland, & Wrench, 2011).

4.0 Analyses and Results

4.1 Segmentation

We analyzed whole-word productions from nine speakers in two morphological conditions (monomorphemic and complex) and two contexts (sentence-final and pre-vocalic). Hence, in utterances like the following in the monomorphemic condition: "A *zord*. It's a *zord*. It's a *zord* again." the last two productions of *zord* were analyzed. Similarly, in utterances like the following in the complex condition: "To *zor*. Yesterday, Tessa *zorred*. Yesterday, Tessa *zorred* again." the last two productions of *zorred* were analyzed. P.M. labeled the relevant acoustic boundaries for each item via visual inspection of the waveform and spectrogram using Praat (Boersma, 2001). Each item's onset was labeled on the basis of the segmentation criteria established by Rastle, Croot, Harrington, and Coltheart (2005). The end point of each item's duration was labeled at the release of the final stop. Erroneous responses (2.9% of the data), including hesitations, dysfluencies, and missing utterances in a certain context, or responses that yielded unclear spectrograms were removed. Example segmentations of the items *zord* and *zorred* in sentence-final and pre-vocalic contexts are provided on the Open Science Framework (OSF; see section 4.3).

4.2 Articulatory data

The goals of the articulatory analysis are twofold. Firstly, we want to know whether there is a mean difference between articulation of monomorphemic and articulation of morphologically-complex words. If morphological structure has a systematic effect on articulation, then we can expect to see a significant difference between the two conditions. Secondly, we are also interested in the nature of that putative difference. The primary dimension we are investigating in this context is coarticulation, operationalized in terms of similarity between two neighboring segments. Any two neighboring segments are coarticulated to a certain extent. Our question is whether the degree of coarticulation changes, depending on the morphological condition, more specifically whether or not we see increased similarity between /r/ and /d/ in monomorphemic words compared to morphologically-complex words.

A major challenge to our analysis is that we are not in a position to make strong predictions about the articulatory region where a difference might become apparent, so we need to compare data from the whole tongue. We also cannot predict a specific time point (i.e., a specific ultrasound frame) where the potential articulatory difference between the two conditions is relatively greatest.

In order to address the need for spatial and temporal sensitivity in the analysis, we conducted a dynamic analysis of signal extracted from entire ultrasound frames. This analysis is reported in Section 4.2.1. In this analysis, we relied on abstract measures of coarticulation, such as global similarity between neighboring segments, but we were able to take more spatial and temporal information into account, and we could aggregate over speakers (see Shaw, Carignan, Agostini, Mailhammer, Harvey, & Derrick, 2020, for a recent application of the same method motivated by a similar need for spatial and temporal depth). We followed this up with a comparison of tongue contours within individuals at selected time points. This is a more established and more phonetically interpretable approach to analyzing ultrasound data, but one that involves a degree of data reduction. The trade-off between the methods is sensitivity vs. interpretability. The first method allows us to establish whether there is coarticulation, while the second can tell us how such coarticulation might translate into differences in tongue shape.

4.2.1 Whole-image analysis. The first part of the articulatory analysis was a dynamic analysis of pixel-intensity data extracted from the ultrasound images. This analysis has three steps. First, we extract the pixel data from ultrasound images and reduce the dimensionality of the data, using Principal Component Analysis. Then, we standardize the data, transforming them into two measures of coarticulation, which are essentially abstract dimensions of segmental similarity. This is done using Linear Discriminant Analysis. Finally, we carry out a dynamic statistical analysis of these two coarticulation measures, to establish whether there is a mean difference in coarticulation, depending on morphological condition.

Ultrasound Tongue Imaging data from nine speakers were included in this analysis. The data from speaker 4 were excluded, because we became aware that the ultrasound probe had moved during the recording session, which would have made the results impossible to interpret. For the remaining speakers, we exported ultrasound frames corresponding to the test words as JPEG frames. We used those images as input to a Principal Component Analysis (PCA) of pixel values, as previously developed by Carignan, Mielke, and Dodsworth (2016); Hoole and Pouplier (2017); Hueber et al. (2007); Mielke, Carignan, and Thomas (2017) *inter alia*. For a detailed description of the method, the reader is referred to Mielke et al. (2017). The PCA was run separately for each speaker. The images were filtered and a region of interest mask corresponding to the approximate range of tongue movement for any particular speaker was applied. The pixel values were then collected into a 3D matrix containing pixel coordinates and intensity. That matrix served as the input to PCA, which reduced the variation in the data to a set of orthogonal principal components. We extracted the number of PCs corresponding to 80% of variance in the data. The number of PCs varied depending on the speaker (23–67), with a median value of 44 PCs. We implemented the PCA using MATLAB code by Carignan (2014). The PCA was applied within speaker, due to individual differences in anatomy, as well as the placement and rotation of the ultrasound probe. Consequently, the articulatory interpretation of the PCs is also speaker-dependent.

The PCA reduces the information in the ultrasound images to a set of numeric vectors. However, these cannot be used as input to statistical analysis, because they do not have a standard interpretation across all speakers, or even within the individual speakers. In order to transform the PCs in a way that is comparable across speakers, we used Linear Discriminant Analysis (LDA; see the following for a similar approach to ultrasound tongue data: Shaw et al., 2020; Smith, Mielke, Magloughlin, & Wilbanks, 2019; Strycharczuk & Scobbie, 2017; Strycharczuk & Sebregts, 2018), which was implemented in R (R Core Team, 2018). The LDA allows us to transform the PCs into vectors that express a specific phonetic category. It is appropriate in our case, because we conceptualize coarticulation in terms of overall articulatory similarity. Perseverative coarticulation is operationalized as similarity of /d/ to /r/, whereas anticipatory coarticulation is operationalized as similarity of /r/ to /d/.

In order to extract the similarity measure, we first trained the algorithm to distinguish frames corresponding to /r/ from the remaining articulatory signal. As a proxy for /r/, we used all the ultrasound frames corresponding to the second half of the acoustic /Vr/ sequence (we expect the /r/ maximum to occur in this temporal region). The outcome is a linear combination of predictors that can be interpreted as a scale of similarity to /r/. We refer to this variable ‘LDr’. Based on the rotation, we calculated the ‘LDr’ values for all ultrasonic frames. We then conducted a similar analysis, this time distinguishing final /d/ from the remaining frames. As a proxy for /d/, we used the middle 50% of the acoustic closure phase, since we might expect that maximum tongue-tip raising will occur around the middle of the acoustic closure phase. The ‘non-/d/’ category corresponded to all the remaining ultrasound frames. Subsequently, we calculated the Linear Discriminant values expressing similarity to final /d/ for all ultrasonic frames. We will call this variable ‘LDd’. This procedure was conducted separately for each speaker.

The two outcome LD vectors were collected into a joint dataset. The next step of the analysis focused on how these variables change dynamically during speech. The change is exemplified in Figure 1.

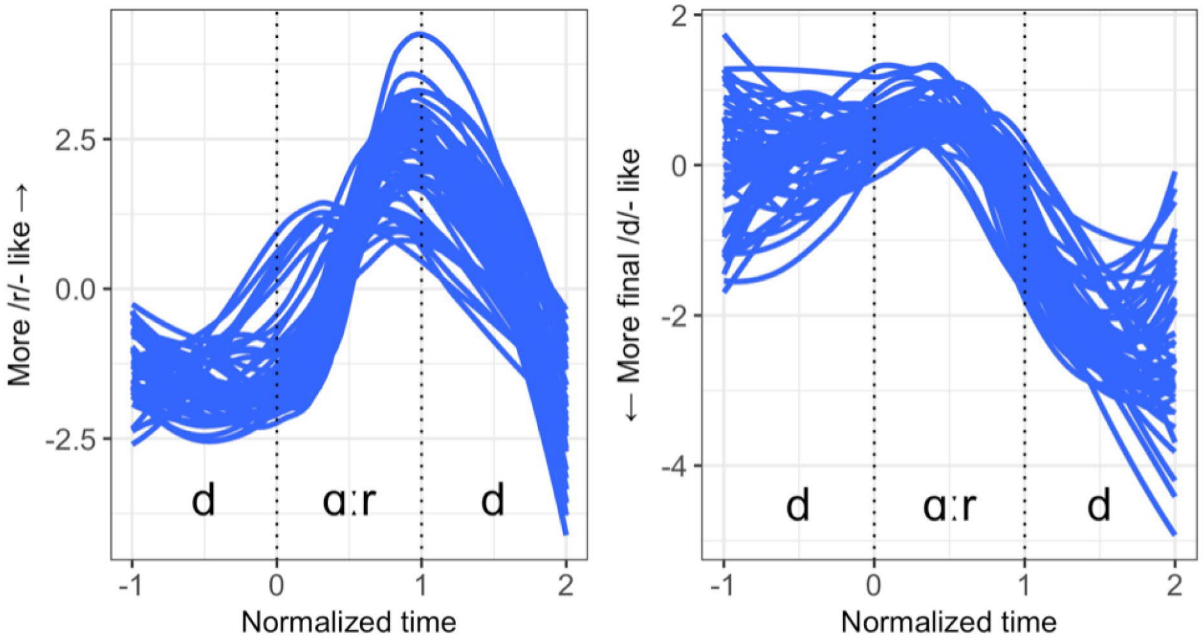


Figure 1. Individual LD trajectories for the item *dard* in sentence-final context pooled across speakers.

The left panel of the figure shows individual trajectories of the linear discriminant trained on /r/ (LDr) for the item *dard* pooled across all speakers. The peak values tend to fall ca. 75% into the /Vr/ interval (i.e., around the center of the category of the time region on which the LDA was trained). There is some variability in the location of the peak, which we ascribe to individual variation in the timing of /r/. The trajectories represent the level of similarity to the articulation at this time point (which we interpret as more /r/-like). The right panel of the figure shows the LDd values. In this case, lower values mean 'more /d/-like'. Note that the initial /d/ shows some lowering of LDd compared to the following vowel, but it still remains distinct from the final /d/, on which the classification was trained. This is expected, since the final /d/ is in a different segmental and prosodic context, both of which affect its articulation.

Although the LD values cannot be assumed to precisely capture the timing of the /r/ and /d/ gestures, they serve as indices of articulatory similarity between different parts of the word.

Therefore we used them as measures of perseverative and anticipatory coarticulation. Similarity to /r/ in the following /d/ is an index of perseverative coarticulation, whereas similarity to final /d/ before its onset is an index of anticipatory coarticulation. Our question is whether the degree of coarticulation can change systematically, depending on the presence or absence of a morphological boundary separating /r/ and /d/. We hypothesize that coarticulation is likely to be greater when there is no morphological boundary. If that is the case, and perseverative coarticulation is increased for monomorphemic items, we expect that these items will show relatively higher values for LDr in the final /d/, which was defined acoustically (see 4.1 section for details). On the other hand, if anticipatory coarticulation is increased for monomorphemic items, lower LDd values preceding the final /d/ might be expected.

We tested these predictions using Generalized Additive Mixed modeling, (GAMMs; Sóskuthy, 2017; Wieling, 2018; Wood, 2006), as this method is well-suited to analyzing non-linear time-series data. The GAMMs were run using the *mgcv* and *itsadug* packages (van Rij, Wieling, Baayen, & van Rijn, 2015; Wood, 2006). We modeled the two variables, LDr and LDd, based on the following predictor variables.

- main effect of morphological condition (monomorphemic vs. complex)
- main effect of prosodic context (sentence-final vs. pre-vocalic)
- smooth term for normalized time
- smooth term for normalized time by morphological condition
- smooth term for duration of annotation
- smooth term for trial time
- smooth term for normalized time by prosodic context
- tensor product interaction between normalized time and duration of annotation
- random by-speaker smooths for normalized time by morphological condition

- random by-item smooths for normalized time

Time was normalized for each of the three annotated intervals (initial consonant, vowel + /r/, and final /d/, as illustrated in Figure 1) and used as the main temporal factor in the analysis. The annotations themselves are segmental boundaries, based on the acoustic segmentations (see section 4.1). In addition, the raw segmental durations were included as a smooth, and in an interaction with normalized time, because segmental duration is correlated with speech rate, and it may affect articulation systematically (e.g., articulatory displacement may be more limited at faster speech rates). Prosodic condition was included as a main effect and in interaction with normalized time for similar reasons, i.e., to account for the possibility that the articulatory trajectory differs systematically depending on the prosodic context. Trial time is a variable expressing time elapsed between the start of the experiment and the recording of a particular token, measured in minutes. It was included to correct for systematic change in the image as a result of potential probe movement throughout the recording session. The random effects account for by-speaker and by-item variability. This includes by-speaker random smooths of morphology. The random smooths modeling individual dynamic morphology effects were included to avoid Type I errors (Sóskuthy, 2021). For item, the same level was coded for pairs of items, e.g., *dard* and *darred* would be coded as ‘d_rd’, since we used a paired design.

The model described above is a rich model, which allows for time-variable differences, depending on morphological structure, while controlling for variability related to trial, prosodic condition, raw duration (correlated with speech rate), speaker, and segmental environment (captured as item variability). Statistical significance was established using Maximum Likelihood (ML) comparisons between the model detailed above and a nested model with no effect of morphology.

We applied models of LDr and LDd to different subsets of the data. Since changes in LDr measure perseverative coarticulation, we analyzed them throughout the duration of the final /d/ (defined acoustically). Changes in LDd, on the other hand, were modeled throughout the duration of Vr, since this is where we expect to see anticipatory coarticulation.

4.2.2 Whole-image results. In modeling LDr, we fitted a model as described above in Section 4.2.1, and a nested model with no main effect of morphological condition, and no smooth for normalized time by morphological condition. Removing these two predictors did not significantly affect the ML scores (difference = 1.36, *df* = 3, *p* = 0.44), which suggests there is no significant difference in the main values for LDr between monomorphemic and complex words, and there is no significant difference in the dynamic profile of LDr between the two morphological conditions either. In contrast, the random smooth for a dynamic morphological effect within speaker did improve the model. Removing this predictor led to a significant change in ML scores (difference = 1828.82, *df* = 3, *p* < .001). This suggests a morphological effect on articulation for some speakers, yet this result should be interpreted with caution, because there was considerable by-speaker variability. Hence, morphology does not seem to affect perseverative coarticulation systematically across speakers.

A similar result was obtained from the modeling of LDd. Comparing models with and without an effect of morphology (main effect and a smooth for normalized time by morphological condition), there was a slight improvement in ML scores after removing the effect of morphology (difference = -1.3, *df* = 3), although a difference in ML scores as small as this is not meaningful. However, further removal of the random smooth for normalized time varying by speaker and by morphology resulted in a large and significant change in ML scores (difference = 4053, *df* = 3, *p* < .001). Once again, this points to individual variation in the way morphology affects the change of LDd scores in normalized time, so that a morphological effect on articulation was apparent in some speakers, yet such effect is not generalizable across the sample.

4.2.3 Tongue-contour analysis. The whole image analysis reported above in Section 4.2.2 focused on an abstract measure of coarticulation, operationalized in terms of similarity between the articulatory signal at different timepoints. We follow up whole-image analysis with analysis of tongue contours. The purpose of this analysis is to add an interpretable articulatory dimension to the analysis of coarticulation, and to explore the individual variation signaled by the results obtained thus far.

For all items, we analyzed the corresponding ultrasound image in each ultrasonic frame by fitting curves (splines) to the tongue contour visible in the image. This was done automatically in Articulate Assistant Advanced software (Articulate Instruments Ltd., 2014), and followed with manual corrections carried out by P.M. The quality of the ultrasound image did not allow for reliable contour tracking in three speakers: 1, 2, and 10. Furthermore, speaker 4 was excluded due to probe displacement during the recording (see section 3.1). This left a subset of six speakers to be included in this type of analysis.

The cartesian coordinates of the individual splines were rotated to the occlusal plane and exported for further analysis in R. We analyzed the change in tongue shape dynamically, using GAMMs, and modeled tongue height (Y) as a function of tongue position (X). The cartesian coordinates were transformed to polar coordinates before fitting the model, using the *rticulate* package in R (Coretta, 2017). GAMMs were fitted individually for each speaker.

For this analysis of tongue contours, we focus on selected time points: 75% /Vr/, onset of closure for the final /d/, and 25% through the final /d/ closure. Our motivation here is that these time points represent the transition from /r/ to /d/, and this transition is also where the morphological boundary falls (for complex items). Therefore, we expect that any potential coarticulation effects for monomorphemic items will be most prominent in this part of the word.

For each speaker, we followed the same model comparison procedure, which involved comparing a full model with a model without morphology. The dependent variable was tongue height (Y). The predictor variables for the full model were the following:

- interaction between morphological condition (monomorphemic or complex) and time point (75% /Vr/, onset of closure for the final /d/, and 25% through the final /d/ closure)
- prosodic context (sentence-final vs. pre-vocalic)
- smooth term for tongue position
- smooth term for trial time
- smooth term for duration of annotation
- tongue position by the interaction between morphological condition and time point
- tongue position by prosodic context
- tensor product interaction for tongue position by trial time
- tensor product interaction for tongue position by duration of annotation
- random by-item smooths for tongue position

Time normalization was done, as for full images, based on annotated intervals (onset, /Vr/, closure of final /d/, as per section 4.2.1). Prosodic context was included to account for potential variability associated with prosodic position. Including duration of annotation interacting with tongue position allows us to model differences in tongue shape associated with variability in speech rate. Trial time was entered as a smooth term and in an interaction with time position to capture some of the variability in tongue shape associated with potential time movement of the ultrasound probe throughout the experiment. The random by-item smooth accounts for differences in tongue shape depending on the segmental environment, such as the quality of the vowel preceding the /r/. As in the whole image analysis above, potential homophones, like *dard* and *darred* were coded as an item pair.

For each speaker, we compared the full model with a model where the interaction between time point and morphological condition was substituted by time point effects only (both in the main term part and in the smooth part). The purpose of the comparison is to establish whether including morphological effects improves the model fit. The comparison was done using the *compareML* function in *itsadug* package. For all speakers, the results were significant, which suggests that morphology did affect the tongue contours at the selected time points. We then analyzed the nature of this interaction, having corrected for residual autocorrelation, using the method of Baayen, van Rij, de Cat, and Wood (2018).

4.2.4 Tongue-contour results. Figures 2 - 7 show by-speaker tongue comparisons between monomorphemic and complex items at the selected time points for the sentence-final context (note that the tongue front is on the right). Prosodic context was modeled as an additive effect. The figures show GAMM-fitted tongue contours by morphological condition, along with 95% confidence intervals.

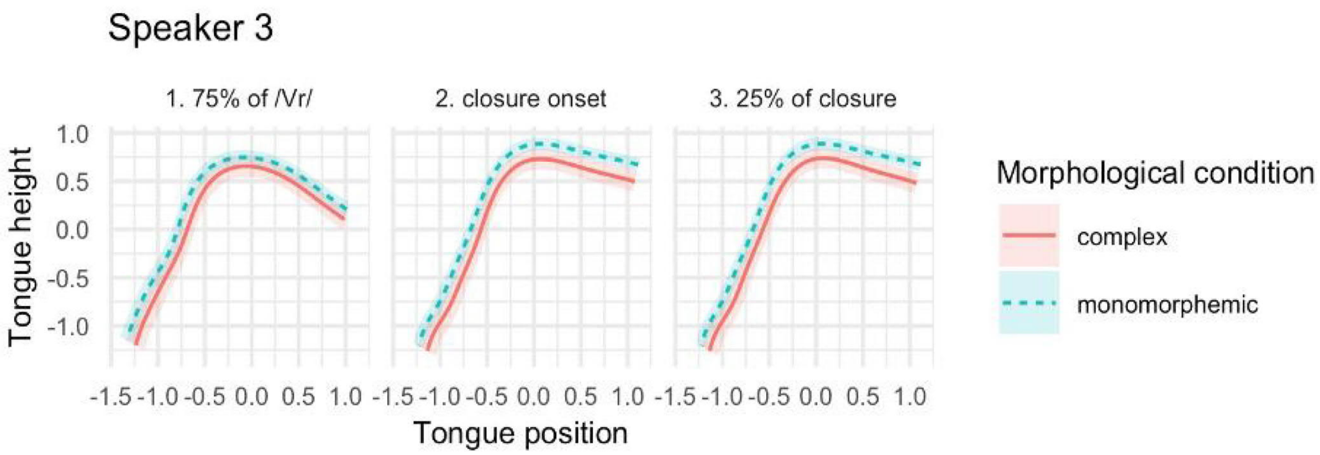


Figure 2. Morphological effects on tongue shape (sentence-final context) at selected time points for speaker 3.

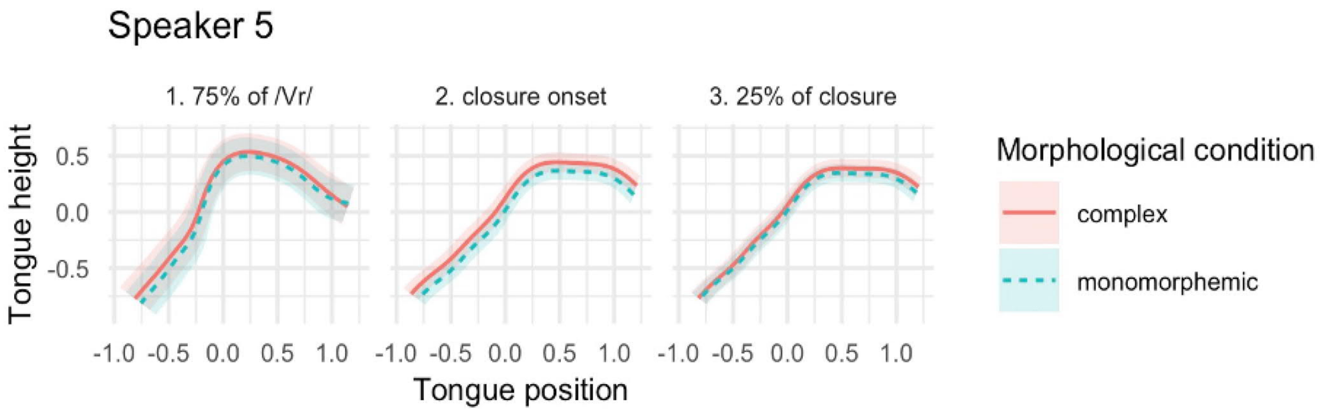


Figure 3. Morphological effects on tongue shape (sentence-final context) at selected time points for speaker 5.

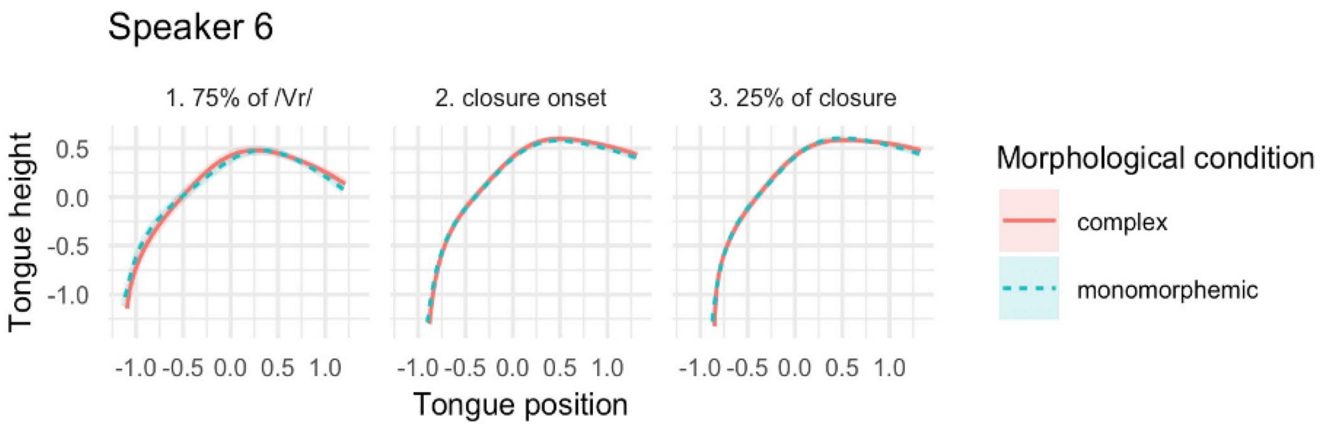


Figure 4. Morphological effects on tongue shape (sentence-final context) at selected time points for speaker 6.

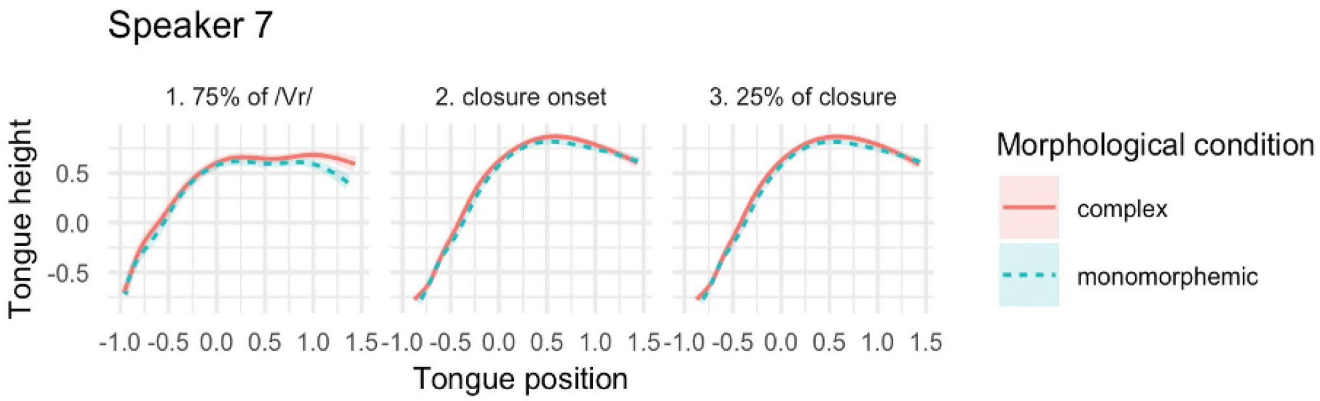


Figure 5. Morphological effects on tongue shape (sentence-final context) at selected time points for speaker 7.

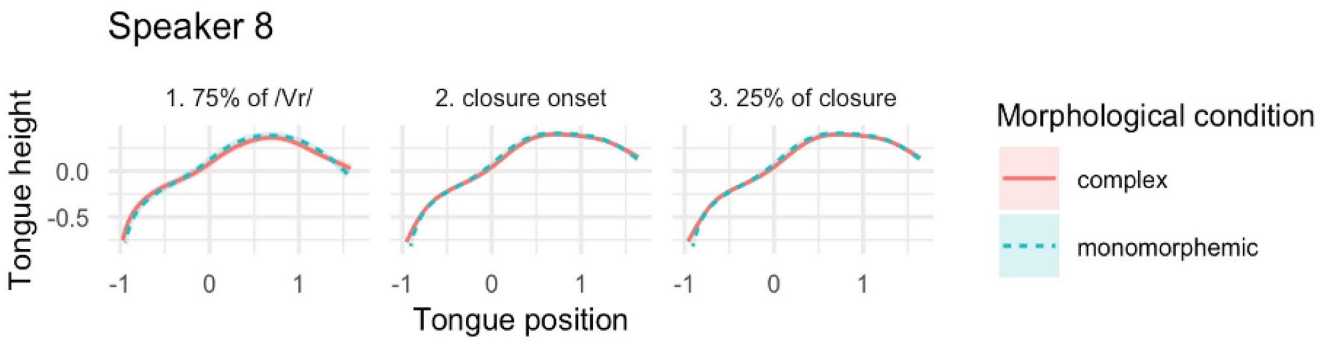


Figure 6. Morphological effects on tongue shape (sentence-final context) at selected time points for speaker 8.

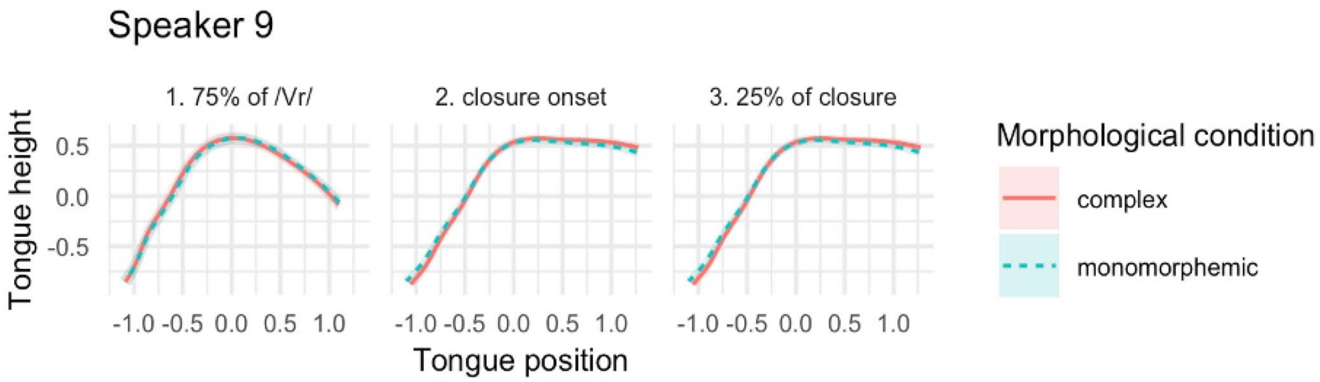


Figure 7. Morphological effects on tongue shape (sentence-final context) at selected time points for speaker 9.

We note that five out of six speakers produced a tip-down (bunched) coda /r/ (Lawson et al., 2011). For these speakers, we would expect that increased anticipatory coarticulation manifests itself as early or more advanced tongue-tip raising in anticipation of an antagonistically apical final /d/ for monomorphemic items. Increased perseverative coarticulation in monomorphemic items would be realized as increased pre-dorsal raising (bunching) persisting into the closure for final /d/. The predictions concerning coarticulation are less clear for Speaker 7, who produced a tip-up (retroflex) coda /r/. However, we might expect less retroflexion in final /r/ for monomorphemic than for complex items if there is increased anticipatory coarticulation. If there is perseverative coarticulation, we would expect that traces of retroflexion might be detectable in final /d/.

The predictions concerning coarticulation, as specified above, are arguably only borne out for speaker 9 (Figure 7), who shows increased tongue tip raising during /r/ in monomorphemes, consistent with anticipatory coarticulation, and tongue lowering during final /d/, also in monomorphemes. This is consistent with increased coarticulation, both anticipatory and perseverative in monomorphemic items.

Other speakers show some morphologically conditioned differences, but not the ones we predicted. Speaker 3 (Figure 2) shows overall increased tongue height for complex items, however this effect is present throughout the /r/ and /d/, so we cannot interpret it in terms of coarticulation, as coarticulatory effects should be more localized. Furthermore, it is worth noting that speaker 5 (Figure 3) shows a difference in the opposite direction. Speaker 6 shows a difference in tongue root, such that it is more advanced for complex than for monomorphemic words at 75% of /Vr/. Speakers 7 and 8 show a similar difference, but in the opposite direction: the tongue root is more advanced at 75% of /Vr/ in monomorphemes.

In summary, the tongue-contour data suggest that the presence of a morphological boundary may affect tongue shape, yielding statistically significant within-speaker effects. We need to note, however, that our initial hypothesis that the presence of a morphological boundary limits coarticulation between /r/ and /d/ was not borne out. Considering the variable nature of the observed effects across speakers in combination with the results of the whole image analysis, which indicated by-speaker differences in the effect of morphology, it seems that there are some apparent effects of morphology on articulation, yet such effects cannot be generalized across speakers.

It is worth pointing out that we further considered whether the absence of a mean significant difference in the whole image analysis might be due to the fact that this analysis included data where the tongue contour was imaged less clearly. We were only able to reliably trace tongue contours for six out of the nine speakers. In theory, including speakers with less clear image could have increased variability in the data, thus obscuring an underlying difference. To test this idea, we reran the whole-image analysis based on data from the six speakers that were included in the tongue-contour analysis. We obtained a similar pattern of results: no significant effect of morphology on either type of coarticulation was observed, yet there was significant individual variation. Therefore, the absence of a significant morphological effect on coarticulation in the whole-image analysis cannot be due to image quality.

4.3 Statement on reproducibility

All data and the R code corresponding to the analyses are available via the OSF (https://osf.io/94f5m/). The OSF repository includes model summaries.

5.0 General Discussion

Morphologically-complex words may be stored in a decomposed form, which would mean that they are computed into a single word each time they are used (e.g., Pinker & Ullman, 2002). Morphologically-simple words can logically only be stored as a whole. The way a word is stored and retrieved is thought to influence its realization during speech production, so that faster retrieval would translate into faster execution of the articulatory plan (Bell et al., 2009). Faster articulatory execution might result in more coarticulation (Gay, 1981; Matthies et al., 2001). On the basis of these related assumptions, we hypothesized that morphologically-simple nonce words (e.g., *zord*) would be produced with more coarticulation than apparently homophonous morphologically-complex nonce words (e.g., *zorred*), because the retrieval of the former does not require online computation, and so it should be faster than the retrieval of the latter. We tested this idea with nine speakers of Scottish English using Ultrasound Tongue Imaging. Our participants learned 16 nonce words (i.e., eight paired items in each condition) through aural training over the course of two days. On the third consecutive day, that is, the testing day, they produced these nonce words, which were placed within phrases and sentences in two prosodic contexts (sentence-final vs. pre-vocalic).

Two types of analyses were performed on the articulatory data. The first type, which was based on the information present in whole midsagittal ultrasound images of the tongue, assessed whether the degree of coarticulation changed systematically, depending on the absence or presence of a morphological boundary between /r/ and /d/ in items such as *zord* and *zorred*. As we mention above, we hypothesized that coarticulation would be greater in monomorphemic items than in morphologically-complex items. More specifically, we reasoned that coarticulation might be perseverative, in which case the final /d/ should be more /r/-like in monomorphemic items; or anticipatory, in which case the gestures preceding the final /d/ should be more /d/-like in monomorphemic items. Nevertheless, the results from our whole-image analysis provided no evidence for systematic perseverative or anticipatory coarticulation in the monomorphemic condition. Of course, coarticulation across a morphological boundary is only one possible domain where effects of morphology may be manifested, so our findings do not preclude the presence of morphological effects that are of a different nature.

The second type of analysis focused on selected time points representing the transition from /r/ to /d/, which is where the morphological boundary for complex items falls. We hypothesized that potential coarticulation effects for monomorphemic items would be most prominent in this part of the word. The results from the tongue-contour analysis revealed some significant differences between the monomorphemic and complex condition for most speakers. However, such differences were observed at different parts of the tongue across speakers, in contrast to our predictions. Even though these results indicate some morphological effects on articulation, they should be interpreted with caution, because the mere presence of a morphological boundary does not seem to affect tongue shape systematically. Taken together, the results from both types of analyses suggest that morphology might influence articulation, yet the observed effects are not generalizable across speakers.

On the assumption that there is a link between lexical selection and articulation (Bell et al., 2009), the lack of a systematic difference in coarticulation between monomorphemic and morphologically-complex items raises the possibility that the latter are not always stored in a decomposed form, even when they are unfamiliar or newly acquired. This is also the case for low-frequency words or most words in children's mental lexicon, which are stored as a whole. Our finding is consistent with the view that morphologically-complex words might be stored as whole items, *as well as* in their decomposed form, as dual-route models of morphological processing postulate (e.g., Caselli, Caselli, & Cohen-Goldberg, 2016; deVaan, Ernestus, & Schreuder, 2011; Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 2015). Given the characteristics of the experimental paradigm used in the present study, one possibility is that morphological composition initially occurred for the complex items. Yet, the repeated production of the inflected word forms over two consecutive days during training could have led our speakers to store these newly-learned forms holistically. As a result, no differences in the articulatory realization of monomorphemic and complex items would be apparent on the third (testing) day. The amount of by-speaker variability, however, that we observed does not offer support for this idea.

Another possibility, raised by a reviewer, is that some speakers might have learned the nonce words better than other speakers. All speakers achieved 100% accuracy by the end of the first day of training, yet it might have been easier for some speakers to retrieve the newly learned words during production. Greater learnability is associated with more skillful articulation (Tomaschek, Tucker, Ramscar, & Baayen, 2021), while speakers with greater lexical proficiency were found to produce stronger articulatory differences between morphological categories (Tomaschek, Tucker, Ramscar, & Baayen, 2019). Accordingly, the speakers in our study who could readily retrieve the nonce words during production might have been the ones that yielded morphological effects on articulation. In contrast, the speakers who could not retrieve the nonce words as fast might have articulated the items in a less effortful manner (Tomaschek et al., 2021), thus making potential articulatory differences between the two morphological conditions more difficult to detect.

The presence of morphological effects on articulation in some speakers cannot be explained by some traditional modular models of speech production (e.g., Levelt, Roelofs, & Meyer, 1999) and of phonology–morphology interaction (e.g., Bermúdez-Otero, 2018; Chomsky & Halle, 1968; Kiparsky, 1982), according to which, there is a distinction between lexical and post-lexical phonological processes (but see Sugahara & Turk, 2009). Lexical phonological processes involve the retrieval of the phonological representations of morphemes and their assembly, which form then the input to post-lexical processes. Post-lexical processes involve the phonological retrieval/encoding of segmental representations and the phonetic encoding of articulatory gestures (see Goldrick & Rapp, 2007). Since morphological information is not available at “later” post-lexical stages in these models, it cannot influence the articulatory realization of words.

Our findings are not fully consistent with those of previous studies (e.g., Cho, 2001; Song et al., 2013), which were taken as evidence for a systematic relationship between morphological complexity and the degree of coarticulation. It is worth noting, however, that the results from these studies cannot be interpreted in a straightforward manner. As we already mentioned earlier, in contrast to his other results, Cho (2001) reported more gestural overlap in heteromorphemic /ti/ sequences than in tautomorphemic ones, thus indicating *less* coarticulation across morphological boundaries. As far as the study of Song et al. (2013) is concerned, only a single item pair with different lexical characteristics (e.g., lexical and sublexical frequency) was used (i.e., *box* vs. *rocks*). These two words further differ in terms of their onsets (/b/ vs. /r/), which require different articulators. More importantly, the coarticulatory effect of /r/ on other segments tends to be very strong. Taking into account the inconsistent results of Cho (2001) and the methodological shortcomings of Song et al. (2013), the reported *systematic* morphological effects on coarticulation should be considered with caution.

It is worth pointing out here that using nonce words in a study such as the present one has an important advantage over the use of real words. As Bürki (2018) pointed out, the way words are articulated may be influenced by many variables at once. Therefore, if the influence of a certain variable on articulation needs to be determined, it is important that all other variables are properly controlled for; or they may act as confounding variables, thus masking potential effects of the variable under investigation and leading to empirical confounds. Matching the words on these (other) variables in the different conditions, or statistically controlling for them in the analyses, may be a solution to the problem. However, controlling for all potential confounding variables is most of the times not possible (Buz & Jaeger, 2016) and may lead to inconsistencies (see Mousikou, Strycharczuk, Turk, Rastle, & Scobbie, 2015; Plag, 2014; Seyfarth et al., 2018).

The last issue concerns the empirical results reported in the literature, which often seek to report a significant difference between the experimental conditions under investigation. Non-significant or non-generalizable results raise the question of whether an effect is indeed absent, or merely very small, and therefore difficult to detect. Non-significant results, however, might be just as important as significant results in terms of trying to understand the workings of the speech production system. Our results suggest that there might be morphological effects on articulation, yet these effects seem to emerge under specific conditions and manifest in an inconsistent manner. Neither of these characteristics is well-suited to providing convincing evidence to support or refute general theories of the architecture of the speech production system. Within our sample, differences in tongue shape between the two morphological conditions were indeed observed, yet the pattern of results was not readily interpretable.

6.0 Conclusion

Our own work does not provide clear-cut evidence in favor of morphological effects on coarticulation. Bearing in mind previous findings in this area, and conscious of the overall empirical complexity, we remain cautious about the theoretical interpretation of our findings. We do, however, see a clear need for future theoretical work in this domain to give serious consideration to generating falsifiable predictions about pronunciation variation phenomena. Well-controlled studies of such phenomena are needed as a means to inform our understanding of the speech production system.

Acknowledgments

This research was supported by British Academy Postdoctoral Fellowships awarded to P.M. and P.S., and an Experimental Psychology Society Small Grant awarded to P.M. The authors have no competing interests to declare.

References

Articulate Instruments Ltd. (2008). Ultrasound stabilization headset, users manual version 1.4.

Articulate Instruments Ltd. (2010). SyncBrightUp, users manual version 1.10.

Articulate Instruments Ltd. (2014). Articulate Assistant Advanced ultrasound module user manual, versions 2.15, 2.16. Edinburgh, UK: Articulate Instruments Ltd.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. doi:10.1177/00238309040470010201

Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1991* (pp. 109–149). Kluwer Academic Publishers.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In *Mixed-effects regression models in linguistics* (pp. 49-69). Springer, Cham.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*, 92–111. doi:10.1016/j.jml.2008.06.003

Berko, J. (1958). The child’s learning of English morphology. WORD, *14*, 150–177. doi:10.1080/00437956.1958.11659661

Bermúdez-Otero, R. (2018). Stratal phonology. In S. J. Hannahs & A. Bosch (Eds.), *Routledge handbook of phonological theory* (pp. 100–134). London: Routledge.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5:9/10*, 341–345. Available online at:

http://www.fon.hum.uva.nl/praat/manual/FAQHowtocitePraat.html

Browman, C. P. & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, no5, 25–34.

Bürki, A. (2018). Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin & Review*, *25*, 1973–2004. doi:10.3758/s13423-017-1423-4

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, *31*, 404–424. doi:10.1080/23273798.2015.1105984

Byrd, D. (1994). Articulatory Timing in English Consonant Sequences. *UCLA Working Papers in Phonetics*, *86*, Ph.D. dissertation, UCLA.

Byrd, D. (1996). Influences on articulatory timing in consonantal sequences. *Journal of Phonetics*, *24*(2), 209–244. doi:10.1006/jpho.1996.0012

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. B. Broe and J. B. Pierrehumbert, (Eds.). *Papers in Laboratory Phonology V*, Cambridge: Cambridge University Press, 70–87.

Carignan, C. (2014). TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) software suite. ms. doi:10.13140/RG.2.2.27085.03047. http://christophercarignan.github.io/TRACTUS.

Carignan, C., Mielke, J., & Dodsworth, R. (2016). Temporal dynamics of /æ/ tensing in Northern American English. In M-H. Côté, R. Knooihuizen & J. Nerbonne. (Eds.), *The future of dialects: Selected papers from Methods in Dialectology XV (Language Variation 1),* 313–319. Berlin: Language Science Press.

Caselli, N. K., Caselli, M. K., & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology*, *69*(3), 432–454. doi:10.1080/17470218.2015.1054847

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean, *Phonetica*, *58*, 129–162. doi:10.1159/000056196

Chomsky, N., & Halle, M. (1968). The Sound Pattern of English. Harper & Row. New York.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *The Behavioral and Brain Sciences*, *22*(6), 991–1013. doi:10.1017/S0140525X99002228

Coretta, S. (2017). rticulate. https://github.com/stefanocoretta/rticulate

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*, *19*(6-7), 619–633. doi:10.1080/02699200500114077

deVaan, L., Ernestus, M., & Schreuder, S. (2011). The lifespan of lexical traces for novel morphologically complex words. *The Mental Lexicon*, *6*(3), 374–392. doi:10.1075/ml.6.3.02dev

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35–39. doi:10.1111/j.1467-9280.2007.01845.x

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods Instruments and Computers, *35*, 116–124. doi: 10.3758/BF03195503

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Geert Evert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1991*, 165–183. Dordrecht, Netherlands: Springer.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, *38*, 148–158. doi:10.1159/000260020

Gick, B., Michelson, K., & Radanov, B. (2006). Articulation without acoustics: A case study of Oneida utterance-final forms. *Journal of the Acoustical Society of America*, *120*(5), 3377. doi:10.1121/1.4781608

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260. doi:10.1016/j.cognition.2005.12.010

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, *4*, 247–263. doi:10.1016/0167-6393(85)90051-2

Hay, J. (2003). *Causes and consequences of word structure*. Abingdon: Routledge.

Herrmann, F., P. Whiteside, S., & Cunningham, S. (2009). An acoustic investigation into coarticulation and speech motor control: High vs. low frequency syllables. In *Proceedings of Meetings of the Acoustical Society of America*, *4*, 1–11. doi:10.1121/1.3085742

Holst, T., & Nolan, F. (1995). The influence of syntactic structure on [s] and [ʃ] assimilation. In *Phonology and phonetic evidence: papers in laboratory phonology* IV (B. Connell & Arvaniti, editors), pp. 315–333. Cambridge: Cambridge University Press.

Hoole, P. & Pouplier, M. (2017). Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech & Language*, *45*, 253–277. doi:10.1016/j.csl.2017.03.002

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., & Stone, M. (2007). Eigentongue feature extraction for an ultrasound-based silent speech interface. In *IEEE International conference on acoustics, speech and signal processing.* Honolulu, 1245–1248. HI: Cascadilla Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2000). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), Frequency and the emergence of linguistic structure (pp. 229–254). Amsterdam, The Netherlands: Benjamins.

Kiparsky, P. (1982). Lexical morphology and phonology. In *Linguistics in the morning calm* (pp.3–91). Seoul: Hanshin Publishing Company.

Labov, W. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City: Vol. I. Phonological and grammatical analysis. New York, NY: Columbia University.

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, *15*(2), pp. 256–268. doi:10.1111/j.1467-9841.2011.00464.x

Lawson, E., Stuart-Smith, J., & Scobbie, J. M. (2018). The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study. *Journal of the Acoustical Society of America*, *143*, 1646–1657. doi:10.1121/1.5027833

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75. doi:10.1017/s0140525x99001776

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.

Losiewicz, B. L. (1992). The effect of frequency on linguistic morphology. PhD dissertation, University of Texas.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*(3), 189–256.

doi:10.1006/cogp.1995.1015

Matthies, M., Perrier, P., Perkell, J., & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech Language and Hearing Research*, *44*(2), 340–353. doi:10.1044/1092-4388(2001/028)

McClean, M. (1973). Forward coarticulation of velar movements at marked junctural boundaries. *Journal of Speech and Hearing Research*, *16*, 286–296. doi:10.1044/jshr.1602.286

Mielke, J., Carignan, C., & Thomas, E. R. (2017). The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *The Journal of the Acoustical Society of America*, *142*, 332–349. doi:10.1121/1.4991348

Mousikou, P., Strycharczuk, P., Turk, A., Rastle, K., & Scobbie, J. M. (2015). Morphological effects on pronunciation. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighbourhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*, 1048–1058. doi:10.1044/1092-4388(2004/078)

Pierrehumbert, J. B. (2002). Word-specific phonetics. In Gussenhoven, C. & Warner, N., (eds.). *Laboratory phonology, 7*, 101–139. New York: Mouton de Gruyter.

Pinker, S. (1999). *Words and rules: the ingredients of language*. New York, NY: HarperCollins.

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Science*, *6*(11), 456–463. doi:10.1016/s1364-6613(02)01990-3

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics*, *26*, 209–228.

Pluymaekers, M., Ernestus, M., &Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*, 2561–

2569. doi:10.1121/1.2011150.

Prasada, S. & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*(1), 1–56. doi:10.1080/01690969308406948

R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Retrieved from https://www.R-project.org/

Rastle, K., Croot, K. P., Harrington, J. M., & 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,* 1083–1095. doi:10.1037/0096-1523.31.5.1083

Rastle, K., McCormick, S., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1588–1594. doi:10.1037/a0024833

Recasens, D. (2018). Coarticulation. In *Oxford Research Encyclopedia of Linguistics*. doi:10.1093/acrefore/9780199384655.013.416

Roettger, T., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics*, *43*, 11–25. doi: 10.1016/j.wocn.2014.01.002

Rubertus, E., & Noiray, A. (2020). Vocalic activation width decreases across childhood: Evidence from carryover coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 7. doi:10.5334/labphon.228

Schreuder, R., & Baayen, R. H. (2015). Modeling morphological processing. In Laurie Beth Feldman (ed.), *Morphological aspects of language processing*, 131–154. New York, NY, USA: Psychology Press.

Scobbie, J. M., Lawson, E., Cowen, S., Cleland, J., & Wrench, A. A. (2011). A common co-ordinate system for mid-sagittal articulatory measurement. *QMU CASL Working Papers* WP–20.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, *33*, 32–49. doi:10.1080/23273798.2017.1359634

Shaw, J. A., Carignan, C., Agostini, T. G., Mailhammer, R., Harvey, M., & Derrick, D. (2020). Phonological contrast and phonetic variation: The case of velars in Iwaidja. *Language*, *96*(3), 578–617. doi:10.1353/lan.2020.0042

Smith, B. J., Mielke, J., Magloughlin, L., & Wilbanks, E. (2019). Sound change and coarticulatory variability involving English /ɹ/. *Glossa: A Journal of General Linguistics*, *4*, 63. doi:10.5334/gjgl.650

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics, 41,* 281–295. doi:10.1016/j.wocn.2013.03.004

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*. *84,* 101017. doi:10.1016/j.wocn.2020.101017

Sóskuthy, M. (2017). Generalized Additive Mixed Models for dynamic analysis in linguistics: A practical introduction. arXiv:1703.05339.

Sóskuthy, M., Foulkes, P., Hughes, V., & Haddican, B. (2018). Changing words and sounds: The roles of different cognitive units in sound change. In J. P. Chevrot, K. Drager, & P. Foulkes (Eds.), *Topics in cognitive science*, 1–16. Wiley Online Library.

Strycharczuk, P., & Scobbie. J. M. (2017). Whence the fuzziness? Morphological effects in interacting sound changes in Southern British English. *Laboratory Phonology: Journal of* *the Association for Laboratory Phonology* 8. 7.

Strycharczuk, P., & Sebregts, K. (2018). Articulatory dynamics of (de)gemination in Dutch. *Journal of Phonetics*, *68*, 138–149. doi:10.1016/j.wocn.2018.03.005

Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, *26*, 477–524. doi:10.1017/S0952675709990248

Tomaschek, F., Arnold, D., Bröker, F., & Baayen, R. H. (2018). Lexical frequency co-determines the speed-curvature relation in articulation. *Journal of Phonetics*, *68*, 103–116. doi:10.1016/j.wocn.2018.02.003

Tomaschek, F., Arnold, D., Sering, K., Tucker, B.V., van Rij, J., Ramscar, M. (2020). Articulatory variability is reduced by repetition and predictability. *Language and Speech*. doi:10.1177/0023830920948552.

Tomaschek, F., Tucker, B. V., Fasiolo, M., & Baayen, R. H. (2018). Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*, *4*, 1–13. doi:10.1515/lingvan-2017-0018

Tomaschek, F., Tucker, B. V., Ramscar, M., & Baayen, R. H. (2019). How is anticipatory coarticulation of suffixes affected by lexical proficiency? doi:10.31234/osf.io/gv89j

Tomaschek, F., Tucker, B. V., Ramscar, M., & Baayen, R. H. (2021). Paradigmatic enhancement of stem vowels in regular English inflected verb forms. *Morphology*, *31*, 171–199. doi:10.1007/s11525-021-09374-w

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). itsadug: Interpreting time series and autocorrelated data using gamms.

West, P. (1999). Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics, 27*, 405–426. doi:10.1006/jpho.1999.0102

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86-116. doi:10.1016/j.wocn.2018.03.002

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics*, *59*, 122–143. doi:10.1016/j.wocn.2016.09.004

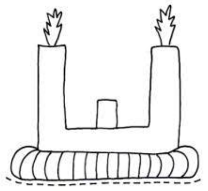
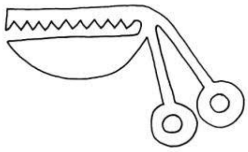
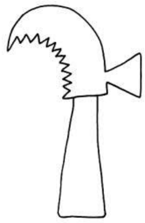
Wolfram, W. A. (1969). A sociolinguistic description of Detroit Negro speech. Washington, DC: Center for Applied Linguistics.

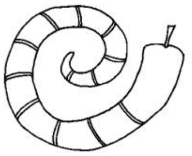
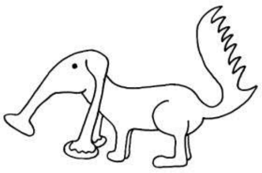
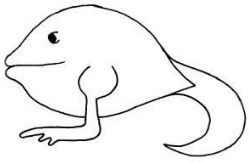
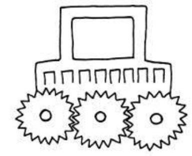
Wood, S. (2006). Generalized additive models: An introduction with R. CRC press.

Wrench, A. A., & Scobbie, J. M. (2016). Queen Margaret University ultrasound, audio and video multichannel recording facility (2008-2016). *CASL Working Papers, Queen Margaret University, WP-24*. http://eresearch.qmu.ac.uk/4367/

Zimmermann, Julia. (2016). Morphological status and acoustic realization: Is there a difference between Bra[d] Pitt and a grille[d] cheese omelet, or between Kate Mo[s] and killer robot[s]? In J. Wahle, M. Köllner, H. R. Baayen, G. Jäger, & T. Baayen-Oudshoorn (Eds.), *Proceedings of 6th Conference on Quantitative Investigations in Theoretical Linguistics (QITL-6)*. Tübingen: Universität Tübingen. http://hdl.handle.net/10900/68949

Appendix. Objects and definitions of actions that corresponded to the nonce noun and verb items used in the study.

To spend on food and drink

To fake a smile

To speak foolishly

To walk without shoes

To pretend to be ill

To hum continuously

To laugh loudly

To eat while lying down