MORPHOLOGICAL EFFECTS ON PRONUNCIATION

Petroula Mousikou[1], Patrycja Strycharczuk[2], Alice Turk[3], Kathleen Rastle[1], & James M. Scobbie[2]

[1]Department of Psychology, Royal Holloway, University of London
[2]Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University
[3]Department of Linguistics and English Language, The University of Edinburgh

ABSTRACT

Converging, albeit inconsistent, empirical evidence suggests that the morphological structure of a word influences its pronunciation. We investigated this issue using Ultrasound Tongue Imaging in the context of an experimental cognitive psychology paradigm. Scottish speakers were trained on apparently homophonous monomorphemic and bimorphemic novel words (e.g. *zord, zorred*), and tested on speech production tasks. Monomorphemic items were realised acoustically with shorter durations than bimorphemic items; however, this difference was not statistically significant. Progressive coarticulatory effects were also observed in the monomorphemic condition for some speakers. A dynamic analysis of the articulatory data revealed that the observed differences in the pronunciations of the two types of items could be due to factors other than morphological structure. Our results, albeit inconclusive, make a significant contribution to the literature in this research domain insofar as the presence or absence of morphological effects on pronunciation has important implications for extant theories of speech production.

Keywords: Morphology; Word learning paradigm; Acoustics; Ultrasound; Kinematics.

1. INTRODUCTION

English speakers can effortlessly attach the past tense morpheme *–ed* to the newly learned verb *unfriend* to convey the meaning of having removed someone from their social network. This ability of speakers to generalise familiar linguistic units to novel words is fundamental for language learning and effective communication. Thus, morphology is known to play an important role in speech production, yet little is known about the specific speech production processes that are influenced by morphological aspects of a language.

Some of the most prominent theories of speech production assume, for example, that the morphological complexity of a word affects the time taken to prepare a verbal response [17]. However, whether it also affects its pronunciation is currently under debate [9, 16, 18, 20, 22, 24, 25]. In particular, the past tense morpheme /d/ or /t/ (e.g. *rapped*) is found to be longer in duration than non-morphemic /d/ or /t/ (e.g. *rapt*) in English [18], and inter-gestural timing for bimorphemic words is shown to be more variable than for apparently homophonous monomorphemic words in Korean [9]. On the assumption that each productively used morpheme constitutes a separate lexical entry [6], so that productively affixed forms consist of multiple lexical entries, these results can be accounted for by Articulatory Phonology [7], which posits that timing relations are more stable between gestures belonging to a single lexical item than across different lexical items. In contrast, the above results cannot be explained by theories in which the output of phonology has no information concerning word-internal morphological structure [10], or theories which pose that phonetic encoding does not have access to morphological information [17].

However, a serious limitation of previous studies is that they typically fail to control for important lexical and/or orthographic variables that are known to affect pronunciation systematically. For example, *rapped* is less frequent and has more letters than *rapt*. Both word frequency and orthographic length are thought to influence acoustic duration, so that low frequency and longer words tend to be produced with longer durations [4, 19, 26]. As such, morphological and lexical/orthographic effects on pronunciation have often been confounded in the literature. The present study overcomes these methodological shortcomings by using an innovative word-learning paradigm from cognitive psychology [13] in combination with Ultrasound Tongue Imaging (UTI) [27] to determine whether there are morphological effects on pronunciation. The use of novel words in our study permits exquisite control of the experimental stimuli, in a way that could never be achieved using existing words [21]. Similarly, the use of ultrasound is critical insofar as effects that are difficult to detect acoustically may nevertheless be present in articulation [15].

In the present study, five speakers of Scottish English produced segmentally identical novel words with different morphological structure (e.g. *zord* and *zorred*). Thus, lexical and orthographic variables such as word frequency and orthographic length were controlled for in our experimental paradigm. We measured the acoustic signal resulting from their productions, and tracked their tongue movements with UTI. We hypothesize that if morphological structure influences pronunciation, the monomorphemic items (e.g. *zord*) will be realised acoustically with shorter durations than the corresponding bimorphemic items (e.g. *zorred*) [18]. We also expect that there will be more prominent coarticulatory effects in the monomorphemic items than in the bimorphemic items [9]. We test these hypotheses using synchronised acoustic and articulatory (ultrasound) data.

2. METHOD

2.1. Participants

Five monolingual native speakers of Scottish English, aged 18-27 years (3F, 2M), with no history of visual, hearing, reading or speech impairments, were paid £30 to participate in the study.

2.2. Materials and design

The experimental stimuli consisted of 16 novel words: 8 novel nouns (*dard, gord, lerd, mord, sard, tord, vard, zord*) and 8 novel verbs (*to dar, to gor, to ler, to mor, to sar, to tor, to var, to zor*), which were associated respectively with 8 unfamiliar objects taken from [21], and 8 unfamiliar definitions of actions. Participants were trained aurally on these associations and never saw the written form of the novel words. For example, they heard a recording of *zord* or *to zor* produced by a female speaker of Scottish English while the corresponding unfamiliar object or definition of an action was visually presented on the computer screen. The associations of nouns and verbs with their corresponding objects and definitions of actions were randomized across participants. Participants were trained on the two types of items in separate blocks. The order of training block was counterbalanced across participants.

Training occurred over the course of two days to ensure long-term memory integration of the newly learnt words [13]. During training, participants were asked to complete orally the sentence *It's a __ again* with the name of the newly learnt object that appeared on the screen (e.g. *zord*), and the sentence *Yesterday, Tessa __ again* with the newly learnt verb that corresponded to the definition that appeared on the screen. Hence, in the latter case, participants were elicited to produce the past tense form of the learnt verb (e.g. if a participant had learnt that *to zor* is *to fake a smile* she/he was expected to produce *Yesterday, Tessa <u>zorred</u> again*).

On the third consecutive day, participants were tested on the newly learnt words in a similar way: they were asked to complete orally the phrases/sentences *A __. It's a __. It's a __ again*, with the name of the visually presented object, and the phrases/sentences *To __. Yesterday, Tessa __. Yesterday, Tessa __ again*, with the appropriate form of the verb that corresponded to the visually presented definition. The two types of items were tested in separate blocks, in the same order in which each participant had learnt them during training. Participants produced six repetitions of each item, making a total of 96 recordings per person.

2.3. Apparatus and procedure

During training (Days 1 and 2), stimulus presentation was controlled by DMDX software [14]. In each training session, each experimental item was presented 15 times. It was also produced by each participant 12 times, and half of these times participants received aural feedback (i.e. they heard the correct answer) after producing their response. All participants achieved 100% accuracy by the end of Day 1. The same training procedure was used on Day 2.

During testing (Day 3), stimulus presentation was controlled by Articulate Assistant Advanced (AAA) software, version 2.15 [3]. 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 stabilized with a headset [1]. The audio data were captured using a lavalier Audio-Technica AT803 condenser microphone connected to a synchronisation unit [2]. 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 [23].

2.4. Analysis

2.4.1. Acoustic data

This analysis included data from five speakers (1, 2, 3, 5, and 6) in the sentence-final (*It's a <u>zord</u>* vs. *Yesterday, Tessa <u>zorred</u>*) and pre-vocalic (*It's a <u>zord</u> again* vs. *Yesterday, Tessa <u>zorred</u> again*) contexts. In particular, we measured the acoustic duration of the monomorphemic and bimorphemic items based on the segmentation criteria established in the ANDOSL database [11]. Each item's onset was labeled on the basis of these criteria (onsets of stop consonants were labeled at the beginning of the closure). The end point of each item's duration was labeled at the release of the final stop.

2.4.2. Articulatory data: dynamic approach

The ultrasound data from five speakers (1, 2, 3, 5, and 6) were analysed using TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) [8]. This method analyses the distance-intensity data in ultrasound images, and reduces the information in the image to a set of principal components (PCs), which account for the greatest amount of variance in the set. For each speaker, we applied the Principal Component Analysis (PCA) to the set of ultrasonic frames corresponding to the acoustic duration of each experimental item (e.g. all frames in *zord* and *zorred*). We then extracted a set of PCs corresponding to 80% of the variance. The PCs were subsequently used in a Linear Discriminant Analysis (LDA). We trained the LDA to distinguish between monomorphemic and bimorphemic items based on a random sample corresponding to 10% of the data, with equal number of frames from both conditions. The classifier was then used to predict the response category (monomorphemic vs. bimorphemic items) in the remaining 90% of the data. The procedure was repeated 1000 times for each speaker to obtain a distribution of classification success, defined as the percentage of correctly identified frames.

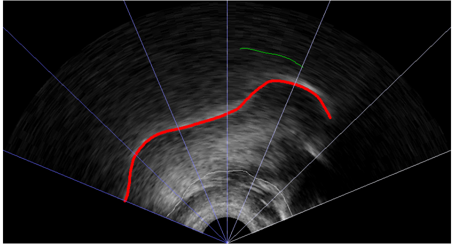
The success of the classification algorithm was evaluated against baseline data from the carrier sentence. We applied the same classification analysis to ultrasonic frames corresponding to the word *again,* which occurred in the carrier sentence in both conditions. If there is a morphological effect contributing to a different pronunciation in the items *zord* and *zorred,* we would expect the classification to perform significantly better in the experimental items than in the word *again*.

2.4.3. Articulatory data: static approach

In this analysis, we compared the tongue shapes in the two conditions of interest (monomorphemic vs. bimorphemic) for each speaker separately. The ultrasound image for Speaker 1 was not sufficiently clear for this type of analysis; hence, her data were excluded. For each item, the tongue contour was manually traced at the onset of the final-/d/ (see Figure 1). We focused on the onset of the final stop closure hypothesizing that progressive coarticulation from the /r/ in the stem might operate differently in the two conditions. In particular, we reasoned that it would be greater in monomorphemic forms compared to bimorphemic ones. Hence, we expected the tautomorphemic /d/ to be more /r/-like in shape at the onset of the stop closure than the suffix /d/ in the bimorphemic form. This specific choice of acoustic closure to represent the /d/ was also influenced by the fact that it could be identified consistently in the acoustic form.

For each speaker (2, 3, 5, and 6), the tongue contours of all repetitions of all items were superimposed in a single workspace, enabling the creation of average tongue shapes and the comparison of the two conditions using the AAA difference function [3]. Average tongue shapes in the two conditions were then statistically compared using Smoothing Splines (SS) ANOVA [12].

Figure 1: Traced tongue contour at the onset of the final-/d/ closure in *zord* for Speaker 5.



3. RESULTS

3.1. Acoustic data

Dysfluent utterances or erroneous responses (2.2% of the data) were discarded. We analysed the duration of the experimental items using linear mixed effects modelling [5]. The preferred model included duration (log-transformed) as the dependent variable, and as fixed effects the interaction between morphological structure and context, as well as the effect of repetition order. Intercepts for participants and items were included as random effects, and so were by-participant random slopes for the interaction between morphological structure and context, and the effect of repetition order (logItem_duration ~ Morph_Structure * Context + rep_order + (1 + Morph_Structure * Context + rep_order | Participant) + (1 | Item)). Outliers with a standardized residual greater than 2.5 standard deviations from zero were removed from the fitted model (1.7% of the data). The results indicated that the items in the prevocalic context were uttered with significantly shorter durations ($t = -4.504$, $p < .01$) than the items in the sentence-final context. Monomorphemic items were uttered with shorter durations compared to bimorphemic items, both in the sentence-final (545 vs. 554 ms) and prevocalic (431 vs. 442 ms) contexts, however, the effect of morphological structure was not significant ($t < 1$).

3.2. Articulatory data: dynamic approach

For each participant, we compared the distribution of classification success in the experimental items vs. the word *again* using one-sample t-tests. The results showed that the classification success was lower for the experimental items compared to the word *again*, i.e. the word *again* pronounced in the two conditions of interest (monomorphemic vs. bimorphemic) was more distinguishable than the experimental items (e.g. *zord* vs. *zorred*). This was the case for 4 out of 5 participants. One participant (5) showed better discrimination of the experimental items compared to the word *again* in the two conditions. The results from this analysis are presented in Table 1.

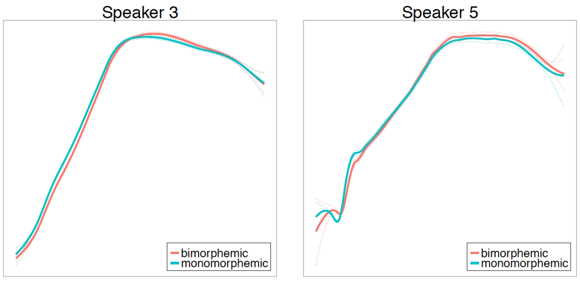
Table 1: Results of t-tests comparing the success of the classification algorithm to predict condition on the basis of monomorphemic vs. bimorphemic items and the word *again* in the two conditions.

	Items	*again*	*t*	*p*
Speaker 1	94.31%	94.44%	5.74	<.001
Speaker 2	82.03%	86.91%	171.35	<.001
Speaker 3	84.77%	92.10%	364.63	<.001
Speaker 5	79.38%	74.59%	-176.63	<.001
Speaker 6	86.92%	90.84%	1940.27	<.001

3.3. Articulatory data: static approach

The results from the SS-ANOVA showed a significant difference between the two conditions for two of the speakers (3 and 6). However, for the other two speakers (2 and 5), no significant differences were observed. Example data are illustrated in Figure 2. This figure shows the mean tongue contours in the two conditions for Speakers 3 and 5, rotated to the occlusal plane [23]. For Speaker 3 (left panel), the tongue dorsum is more retracted in the monomorphemic condition compared to the bimorphemic one. For Speaker 5 (right panel), the tongue shapes in the two conditions do not differ significantly from each other.

Figure 2: Results of SS-ANOVA for Speakers 3 and 5. Non-overlapping confidence intervals indicate a significant difference between the two conditions.



4. DISCUSSION AND CONCLUSIONS

The role of morphology in speech production is important, yet the available empirical evidence on *whether* and *how* the morphological structure of a word may influence its pronunciation is inconsistent in the literature. In the present study, we used UTI in the context of a word-learning cognitive psychology paradigm to investigate whether there are morphological effects on pronunciation. Based on results from previous studies [9, 18], we hypothesized that monomorphemic items would be shorter in duration, and more coarticulated than bimorphemic items.

The acoustic analyses showed a non-significant durational difference in the expected direction, that is, monomorphemic items were acoustically shorter than bimorphemic items. Moreover, the static articulatory analysis was suggestive of progressive coarticulatory effects in the monomorphemic condition, yet this was the case for half of the speakers only, indicating significant inter-speaker variability. Last, the dynamic articulatory analysis indicated that for all speakers except one, the monomorphemic and bimorphemic items were less distinguishable than the word *again* in the two conditions. Consequently, although there seems to be a difference between monomorphemic and bimorphemic forms, this difference may be due to factors other than morphological structure. As such, the results from both acoustic and articulatory analyses are currently inconclusive.

It is worth pointing out that the finding that the neutral word *again* was more distinguishable in the two conditions compared to the monomorphemic and bimorphemic items (e.g. *zord* vs. *zorred*) is unexpected. However, given that the two types of items were presented in separate blocks, it could well be the case that prosodic differences in the two blocks due to fatigue, for example, exerted a strong influence on the pronunciations of most speakers. Thus, morphological effects may have been confounded with an effect of block order.

To sum up, the present results do not allow us to conclude that morphology influences pronunciation, which contrasts the conclusions reached in other studies [9, 18, 24]. However, the present data come from a very small sample of speakers, hence further study of the putative acoustic and articulatory differences in the two conditions, using the methods described in this paper, is well worthwhile. A convincing null result in relation to our research question would have as important theoretical implications as observing morphological effects on pronunciation [7, 10, 17].

5. REFERENCES

[1] Articulate Instruments Ltd. 2008. Ultrasound stabilization headset, users manual version 1.4.

[2] Articulate Instruments Ltd. 2010. SyncBrightUp users manual version 1.10.

[3] Articulate Instruments Ltd. 2014. Articulate Assistant Advanced ultrasound module, users manual version 2.15.

[4] Aylett, M., & Turk, A. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J. Acoust. Soc. Am*. 119, 3048–3059.

[5] Baayen, R.H. 2008. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

[6] Booij, G.E. 1985. Word Formation. *Lingua*, 65, 260–270.

[7] Browman, C., & Goldstein, L. 1992. Articulatory Phonology: an Overview. *Phonetica*, 49, 155–180.

[8] Carignan, C. 2014. TRACTUS (Temporally Resolved Articulatory Configuration Tracking of Ultrasound) software suite, http://phon.chass.ncsu.edu/tractus.

[9] Cho, T. 2002. Effects of morpheme boundaries on intergestural timing: Evidence from Korean, *Phonetica*, 58, 129–162.

[10] Chomsky, N., & Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

[11] Croot, K., Fletcher, J., & Harrington, J. 1992. Phonetic segmentation of the Australian National Database of Spoken Language. *Proc. 4th International Conference on Speech Science and Technology*, Brisbane, 86–90.

[12] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am*. 120, 407–415.

[13] Dumay, N., & Gaskell, M.G. 2007. Sleep-associated changes in the mental representation of spoken words. *Psychological Science,* 18, 35–39.

[14] Forster, K.I., & Forster, J.C. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116–124.

[15] Gick, B., Michelson, K., & Radanov, B. 2006. Articulation without acoustics: A case study of Oneida utterance-final forms. *J. Acoust. Soc. Am*. 120, 33–77.

[16] Hanique, I., & Ernestus, M. 2012. The role of morphology in acoustic reduction. *Lingue e linguaggio*, 11, 147–164.

[17] Levelt, W.J.M., Roelofs, A., & Meyer, A.S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.

[18] Losiewicz, B.L. 1992. The effect of frequency on linguistic morphology. PhD dissertation, University of Texas.

[19] Munson, B. 2007. Lexical access, lexical representation, and vowel production. In J.S. Cole & J.I. Hualde (Eds.), *Laboratory phonology* 9 (pp. 201–228). Berlin/New York: Mouton.

[20] Pluymaekers, M., Ernestus, M., Baayen, R.H., & Booij, G.E. 2010. Effects of morphological structure on fine phonetic detail: The case of Dutch –igheid. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology* 10 (pp. 511–531). Berlin: De Gruyter Mouton.

[21] Rastle, K., McCormick, S., Bayliss, L., & Davis, C.J. 2011. Orthography influences the perception and production of speech. *JEP:LMC*, 37, 1588–1594.

[22] Schuppler, B., van Dommelen, W., Koreman, J., & Ernestus, M. 2012. How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and subphonemic level. *Journal of Phonetics*, 40, 595–607.

[23] 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.

[24] 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.

[25] Sugahara, M., & Turk, A. 2009. Durational correlates of sub-lexical constituent structure in English. *Phonology*, 26, 477–524.

[26] Warner, N., Jongman, A., Sereno, J., & Kemps, R. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics*, 32, 251–276.

[27] Wrench, A.A., & Scobbie, J.M. 2011. Very high frame rate ultrasound tongue imaging. *Proc. 9th ISSP* Montreal, 155–162.