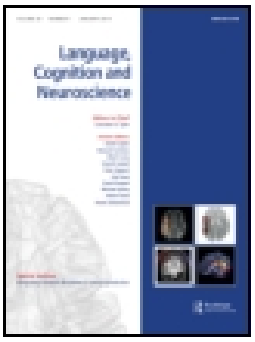

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: http://www.tandfonline.com/loi/plcp21

Acoustic differences in morphologically-distinct homophones

Scott Seyfarth, Marc Garellek, Gwendolyn Gillingham, Farrell Ackerman & Robert Malouf

LANGUAGE, COGNITION AND NEUROSCIENCE, 2017
https://doi.org/10.1080/23273798.2017.1359634

Acoustic differences in morphologically-distinct homophones

Scott Seyfarth[a], Marc Garellek[a], Gwendolyn Gillingham[a], Farrell Ackerman[a] and Robert Malouf[b]

[a]Department of Linguistics, University of California, San Diego, CA, USA; [b]Department of Linguistics and Asian/Middle Eastern Languages, San Diego State University, San Diego, CA, USA

ABSTRACT
Previous work demonstrates that a word's status as morphologically-simple or complex may be reflected in its phonetic realisation. One possible source for these effects is phonetic paradigm uniformity, in which an intended word's phonetic realisation is influenced by its morphological relatives. For example, the realisation of the inflected word *frees* should be influenced by the phonological plan for *free*, and thus be non-homophonous with the morphologically-simple word *freeze*. We test this prediction by analysing productions of forty such inflected/simple word pairs, embedded in pseudo-conversational speech structured to avoid metalinguistic task effects, and balanced for frequency, orthography, as well as segmental and prosodic context. We find that stem and suffix durations are significantly longer by about 4–7% in fricative-final inflected words (*frees*, *laps*) compared to their simple counterparts (*freeze*, *lapse*), while we find a null effect for stop-final words. The result suggests that wordforms influence production of their relatives.

ARTICLE HISTORY
Received 12 October 2016
Accepted 3 July 2017

KEYWORDS
Paradigm uniformity; morphology; articulation; spreading activation; speech production

1. Introduction

When a language-user produces a spoken word, its exact articulation is influenced by a wide range of linguistic and psycholinguistic variables, such as the word's position in a phrase (Oller, 1973), overall frequency in the language (Gahl, 2008), and its predictability in context (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Lieberman, 1963). How do the morphological properties of a word influence its phonetic realisation? Discrete, sequential processing architectures (e.g. Levelt, Roelofs, & Meyer, 1999) and other non-interactive models of language production (e.g. Kiparsky, 1982) propose that when phonetic attributes such as duration and pitch are encoded from a phonological representation, the word's morphological status is inaccessible. However, a growing body of work demonstrates that morphology does interact with phonetic characteristics such as formant trajectory alignment (Scobbie, Turk, & Hewlett, 1999), /l/-darkening (Hayes, 2000; Lee-Kim, Davidson, & Hwang, 2013; Sproat & Fujimura, 1993; Strycharczuk & Scobbie, 2015, 2017), and segment duration (Plag, Homann, & Kunter, 2017; Pluymaekers, Ernestus, Baayen, & Booij, 2010; Smith, Baker, & Hawkins, 2012). For example, the /t/ is aspirated in the derived word *mistime*, but not in the morphologically-simple word *mistake*, even though it occurs in the same phonological environment in both words (Baker, Smith, & Hawkins, 2007; Smith et al., 2012; Zuraw & Peperkamp, 2015).

What causes these effects? One possible mechanism is *phonetic paradigm uniformity*: the influence of an intended word's morphological relatives on the articulatory realisation of that word (Ernestus & Baayen, 2006; Frazier, 2006; Hayes, 2000; Roettger, Winter, Grawunder, Kirby, & Grice, 2014; Steriade, 2000). There is some existing evidence that morphological families affect speech production latencies (Baayen, Levelt, Schreuder, & Ernestus, 2007; Hay & Baayen, 2005). However, these effects are nonetheless compatible with a model that segregates morphology and phonetics, as they arguably involve competition during lexical retrieval processes, rather than during speech encoding or articulation (see Goldrick, Baker, Murphy, & Baese-Berk, 2011; Goldrick, Keshet, Gustafson, Heller, & Needle, 2016). In this paper, we argue that an intended word's morphological relatives also interact with that word's phonetic realisation, and test this hypothesis by looking at the durational influence of freestanding English stems on the wordforms in their inflectional paradigms. More generally, this work addresses broader questions about interaction among different components of the linguistic signal, and the role of analogy between wordforms in phonological representation.

1.1. Paradigm uniformity

A morphological paradigm is the set of words that have a lemma in common. For example, the inflectional

CONTACT Scott Seyfarth sseyfarth@ucsd.edu
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2017.1359634.

paradigm of the English verb FREE is *free*, *frees*, *freeing* and *freed*. Paradigm uniformity is a pressure for invariance among the phonological forms of an inflectional or derivational paradigm (Hayes, 2000; Steriade, 2000). This phenomenon occurs in the pronunciation of the American English words *capitalistic* and *militaristic*. The unstressed syllable /tə/ in the word *capitalistic* is normally produced with an alveolar tap [ˌkæpɪɾəˈlɪstɪk]. This follows the phonological pattern in which intervocalic /t/ is tapped when it is unstressed. However, the same syllable /tə/ in the word *militaristic*—which is unstressed, just as in *capitalistic*—can be pronounced with an aspirated [t] ([ˌmɪlɪtʰəˈɹɪstɪk]), even though this violates that phonological pattern (Withgott, 1982). This can be accounted for by uniformity pressures within the two words' derivational paradigms (Steriade, 2000). The syllable that corresponds to /tə/ is unstressed in *capital* /ˈkæpɪtḷ/, but is stressed in *military* /ˈmɪlɪˌtɛɹi/. Even though /tə/ is unstressed in the derived *militaristic*, the pressure for paradigmatic uniformity with *military* prevents it from being realised as a tap in *militaristic*. On the other hand, there is no such influence on *capitalistic*, because the /t/ is also realised as a tap in *capital* (see also Davis, 2005, for a different uniformity-based analysis).

While paradigm uniformity has been formalised in several symbolic phonological theories (e.g. Benua, 1997; McCarthy, 2005, see Steriade, 2000 for a summary), it has also been argued to influence more fine-grained production patterns (Frazier, 2006; Hayes, 2000; Steriade, 2000). As one instance, paradigm uniformity may account for incomplete voicing neutralisation patterns in Germanic languages (Ernestus & Baayen, 2006, 2007; Kaplan, 2016; Roettger et al., 2014; Winter & Roettger, 2011). For example, the German words *Rad* "wheel" and *Rat* "council" are typically considered to be homophones, ending in a final voiceless segment: both are pronounced [ʁaːt]. *Rad* is morphologically related to *Räder* "wheels", in which the corresponding segment is voiced [d] ([ʁɛːdɐ]). However, *Rat* has no such voiced relative. A body of research demonstrates that there are fine-grained phonetic differences between *Rad* and *Rat* such that *Rad*, but not *Rat*, is produced with some of the phonetic cues associated with a final voiced segment (see Winter & Roettger, 2011 and Roettger et al., 2014 for recent theoretical and experimental reviews). One account for these results is that incomplete neutralisation is the result of paradigm uniformity effects: when a speaker produces the form *Rad*, their production is influenced by the morphologically-related voiced form *Räder*, which affects how voicing cues are realised in *Rad*.

1.2. Mechanisms for paradigm uniformity effects

Phonetic paradigm uniformity effects like incomplete voicing can be operationalised in terms of spreading activations among wordforms (Ernestus & Baayen, 2006, 2007; Goldrick & Blumstein, 2006; Roettger et al., 2014; Winter & Roettger, 2011). Following this proposal, when a speaker retrieves a target wordform for production, semantically- and phonologically-related words are co-activated (Dell, 1986; Goldrick, 2006, 2014; McMillan, Corley, & Lickley, 2009; Peterson & Savoy, 1998; Rapp & Goldrick, 2000). For example, the process of activating the target word *cat* also involves partial activation of the semantic relative *dog*. In non-discrete production models, the activation of both *cat* and *dog* cascades into phonological planning, such that the forms /kæt/ and /dɑg/ are both activated to some extent. Further, activation of the phonological form /kæt/ feeds back to activate phonological relatives such as *hat*, due to the segmental overlap of /kæt/ and /hæt/ (Rapp & Goldrick, 2000). By definition, the other words in a target's morphological paradigm are close semantic relatives, and are likely to be phonological relatives as well (see e.g. Bybee, 1985, on lexical connections within a paradigm). In some proposals, morphological relatives may always be co-activated (Ernestus & Baayen, 2007), regardless of semantic or phonological similarity. In either case, retrieval of a target form *Rad* [ʁaːt] leads to co-activation of the phonological form of *Räder* ([ʁɛːdɐ]), due to the spreading activations within and between the semantic and phonological retrieval processes.[1]

Evidence from speech errors suggests that cascading activation from non-target phonological forms can have gradient influences on articulatory processes (Goldrick et al., 2011, 2016; Goldrick & Blumstein, 2006; McMillan et al., 2009). There are at least two pathways through which this might happen. First, partially-activated non-target forms may contribute to an articulatory plan, such that the resulting plan is a mix of target and non-target forms (e.g. Gafos, 2003; Goldrick & Blumstein, 2006). Alternatively, multiple gestural plans may be constructed and simultaneously implemented (or partially implemented), leading to gestural blending in cases where two gestures cannot be executed simultaneously, and overlap when they can be (e.g. Pouplier & Goldstein, 2010).

For the case of target *Rad* and co-activated *Räder*, the prediction is that the relatively strong influence of the close relative *Räder* should affect the production of [ʁaːt], even in normal speech (see also Gafos, 2003). In particular, the influence of co-activated /d/ should lead to a blend of corresponding /t/ and /d/ realizations, such that *Rat* is produced as partially voiced (Ernestus &

Baayen, 2006, 2007; Gafos, 2003; Goldrick & Blumstein, 2006; Roettger et al., 2014; Winter & Roettger, 2011). This mechanism generates predictions about phonetic paradigm uniformity effects more broadly (and see references above). When a word has a high degree of semantic, grammatical, and/or phonological similarity to its paradigm members, its final articulatory realisation should be influenced by those members (Goldrick, Folk, & Rapp, 2010).

1.3. Paradigm effects on English inflected forms

One proposed and measurable type of phonetic uniformity effect involves word and segment duration (Frazier, 2006; Steriade, 2000). There is evidence that durational targets are specified in a wordform's phonological plan (Katz, 2010, 2012; Seyfarth, 2014; Tauberer & Evanini, 2009), and duration has previously been used as a test case for interaction among plans (Goldrick et al., 2011). Here, we investigate the effects of monosyllabic English words such as free on an inflected paradigm member with a heavier coda, such as *frees*. The paradigm uniformity account predicts that the timing of the segments in *frees* should be influenced by the durational targets of *free* (see Frazier, 2006 for a similar proposal based on moraic structure).[2] As a baseline for what the timing of *frees* should be if there were no interference from paradigm members, we compare each inflected word to a segmentally-identical but morphologically-simple homophone, such as *freeze*.

The inflected word *frees* should show the following uniformity effects from the influence of *free*. The first kind of effect arises from differences in syllable weight. The word *free* [fɹi] has no coda, and therefore the nucleus is longer than if it were in a closed syllable (Katz, 2010, 2012; Munhall, Fowler, Hawkins, & Saltzman, 1992; Shaiman, 2001). If the lighter form *free* influences the plan for *frees*, the nucleus should be relatively longer in *frees* compared to *freeze*, where it is not influenced by a longer wordform (Frazier, 2006).

The second kind of uniformity effect arises from differences in prosodic alignment. This can be illustrated in a gestural score. Figure 1(a) shows a partial gestural score for the rime [i] in the freestanding word *free*, which shows the tongue-body constriction gesture associated with the vowel. In addition to constrictions, prosodic effects are also modelled as gestures (Byrd, Krivokapić, & Lee, 2006; Byrd & Saltzman, 2003). Prosodic gestures (*π*, shown in grey) overlap with constriction gestures, and change the overall rate that the production system moves through the gestural score. For example, because *free* is a prosodic word, it ends in a prosodic-word gesture. While the prosodic gesture is activated, it slows the rate at which constrictions are produced (Byrd et al., 2006; Byrd & Saltzman, 2003), thus causing word-final lengthening (Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Because *free* ends with a tongue-body constriction, the prosodic-word gesture overlaps mainly with this constriction, resulting in elongation of the [i].

By comparison, Figure 1(b) shows a partial score for the [iz] rime of *freeze*. In this score, the prosodic-word gesture overlaps mainly with the [z] constriction. Thus, the most elongated segment in *freeze* is [z], while in *free*, [i] is relatively more elongated. This follows empirical work showing that final lengthening effects are greatest on segments immediately adjacent to a prosodic

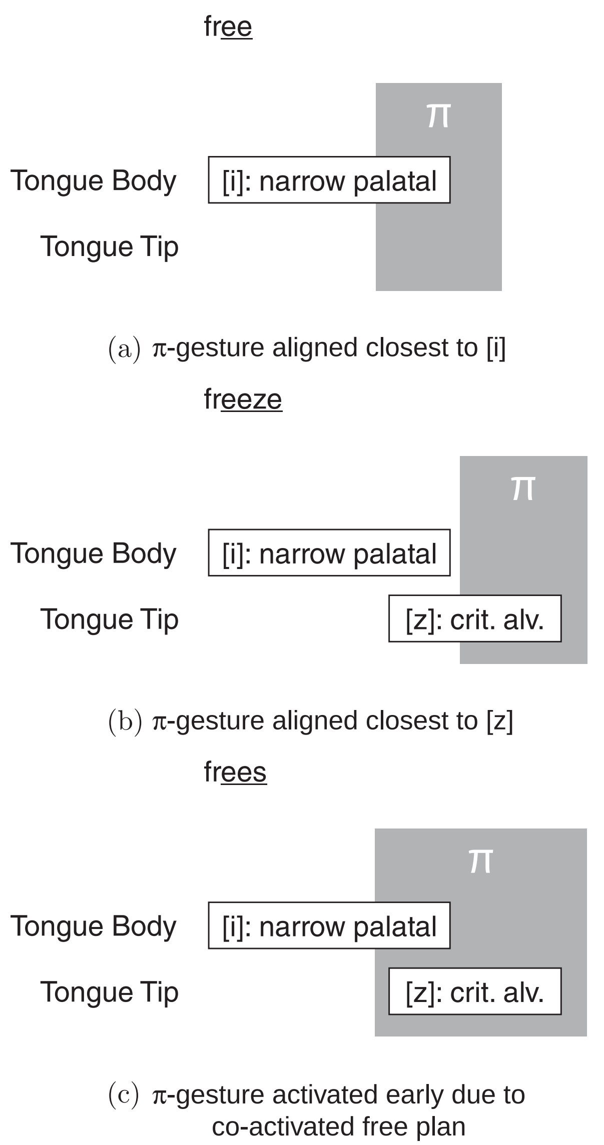

Figure 1. Partial articulatory scores showing the activation of gestures over time (from left to right) during the rime of *free* (a), *freeze* (b), and *frees* (c). Prosodic (*π*) gestures (grey) at the ends of words cause other gestures to be lengthened in duration.

boundary (Byrd et al., 2006; Byrd & Saltzman, 2003; Shattuck-Hufnagel & Turk, 1998).

Figure 1(c) shows a partial score for the morphologically-complex *frees*. If the gestural score for *free* is co-activated and influences the production of its morphological relative *frees*, the prosodic gesture in *free* is predicted to influence the timing of the corresponding gesture in *frees*. Thus, the prosodic-word gesture in *frees* will be activated earlier than would otherwise be expected. The result is that the domain of word-final lengthening will extend earlier into the production of *frees* than of *freeze*, and the string [fɹi] is predicted to be overall longer in inflected *frees* compared to morphologically-simple *freeze*.

In addition, if the prosodic gesture is activated earlier in *frees* than in *freeze*, it will overlap more of the tongue-tip [z] gesture than in *freeze*, slowing the production of the word-final consonantal constriction. Since English sibilants are highly elastic with respect to domain-edge prosodic effects (Cho, Lee, & Kim, 2014; Hofhuis, Gussenhoven, & Rietveld, 1995; Klatt, 1976; Oller, 1973), the expectation is that the durations of English word-final [s, z] suffixes should be lengthened by a longer prosodic gesture in English inflected words. However, not all segments are equally sensitive to prosodic lengthening effects (Fougeron, 2001; Keating, 2006; Oller, 1973). In particular, word-final [t, d] are less elastic than vowels and sibilants (Berkovits, 1993; Hofhuis et al., 1995; Klatt, 1976). Thus, lengthening in English words with [t, d] inflections (*ducked, tied*) is less likely to be observable on the suffixes themselves.

1.4. Previous evidence

Prior work has compared the durations of simple and inflected English homophones (*tax/tacks*), but with unclear results. Two laboratory studies report that suffix durations are longer in inflected words than their simple homophones (Losiewicz, 1992; Walsh & Parker, 1983), and two report the same pattern for vowel or stem durations (Frazier, 2006; Sugahara & Turk, 2009, as well as mixed results discussed in Sugahara & Turk, 2004). However, a major concern with the interpretation of these findings is that the word productions were elicited in short lists of homophones and in short phrases intentionally designed to highlight contrasts between the target words. It has been shown that phonetic variation between orthographically-distinct homophones increases when the target homophones are dictated in an isolated-word list or in contrastive sentences, as compared to when the target words are disguised in longer contexts (Fourakis & Iverson, 1984; Kharlamov, 2014; Port & Crawford, 1989, see also Roettger et al., 2014; Winter & Roettger, 2011). Thus, while the participants in these studies may have been encouraged by the experimental design to produce phonetic distinctions, those distinctions may have been motivated by orthography or metalinguistic knowledge as much as by the words' morphological properties (Fourakis & Iverson, 1984; Jassem & Richter, 1989; Kharlamov, 2014; Mousikou, Strycharczuk, Turk, Rastle, & Scobbie, 2015, see also Ernestus & Baayen, 2006; Warner, Good, Jongman, & Sereno, 2006; Warner, Jongman, Sereno, & Kemps, 2004).

More broadly, the generalizability of previous reports has also been criticised (see Bermúdez-Otero, 2010; Hanique & Ernestus, 2012; Mousikou et al., 2015; Plag, 2014; Plag et al., 2017), including the findings of one corpus study that reports longer suffix durations for inflected words (Song, Demuth, Evans, & Shattuck-Hufnagel, 2013). These studies have often tested very few items (3 homophone pairs in Walsh & Parker, 1983; 6 pairs in Losiewicz, 1992; 9 non-homophonous words in Song et al., 2013), found the effect only at a slow speech rate (Sugahara & Turk, 2009), only utterance-finally (Song et al., 2013), or were not robust to current statistical practices (Plag, p.c., on Losiewicz, 1992). Additionally, the inflected and simple words in the prior laboratory work were not balanced for frequency, which is well known to influence acoustic duration.[3] Several authors also raise a concern about orthographic differences (Fourakis & Iverson, 1984; Winter & Roettger, 2011 on incomplete neutralisation; Mousikou et al., 2015; Sugahara & Turk, 2004, 2009, on duration), which might affect production independently of morphological status (Brewer, 2008; Bürki, Spinelli, & Gaskell, 2012; Ernestus & Baayen, 2006; Warner et al., 2006, 2004). Specifically with regard to duration, Warner et al. (2004, 2006) and Brewer (2008) find that words spelled with more letters are produced with longer durations.

In addition to these concerns, a recent study of a larger number of non-homophonous inflected and simple words in conversational speech reports the *opposite* pattern for English [s] suffix durations: final [s] is *shorter* when it signals an inflectional suffix (Plag et al., 2017). Since this study found the opposite pattern as prior laboratory experiments, one interpretation is that the experimental work may have been confounded by task effects or other methodological issues. At the same time, a corpus-based analysis raises a different set of analytical and interpretability challenges due to the heterogeneous word types in the data, as well as the unbalanced prosodic contexts that English inflected and uninflected words tend to appear in. We return to these questions in the discussion Section 4.3.

1.5. The current study

In the current study, we analyse the stem and suffix durations in forty pairs of monosyllabic English homophones, in which one member of the pair is inflected (*frees*) and the other is morphologically-simple (*freeze*). Under the paradigm uniformity account, the prediction is that stem durations should be relatively longer in inflected words like *frees*, compared to simple *freeze*, due to the prosodic influence of a lighter word (*free*) on the inflected but not simple words. This account also predicts that the suffix duration in *frees* should be lengthened relative to the same segment in *freeze* as a result of the longer prosodic gesture in *frees*. Because the theory predicts no differences between different morphological suffixes, we include a variety of [s, z, t, d] English suffixes in our stimulus set. Besides these planned tests, we also use the data to explore the influence of probabilistic variables on inflected words (Cohen, 2014; Cohen Priva, 2012; Hay, 2003; Rose, Hume, & Hay, 2015; Schuppler, van Dommelen, Koreman, & Ernestus, 2012). In particular, we evaluate whether the predicted influence of morphological relatives is stronger if these relatives are more frequent, in either absolute or relative terms.

In order to elicit more natural speech and avoid metalinguistic task effects, yet still maintain a phonetically-controlled context, we adapt a method used by Port & Crawford (1989), Baker et al. (2007), and Smith et al. (2012) in which the homophone pairs are embedded in conversational dialogues that are matched for prosodic and segmental context. These dialogues are read by pairs of naïve participants who are already friends, and who are familiarised with the dialogues prior to participation in the experimental task (see Warner, 2012).

This method has several crucial advantages over previous work. First, by using controlled dialogues rather than completely spontaneous speech, we are able to collect productions of a large number of homophone pairs which are matched for frequency and orthographic length, factors that have been potential confounds in previous work (Hanique & Ernestus, 2012; Mousikou et al., 2015; Plag, 2014; Plag et al., 2017). The use of homophone pairs allows us to compare matched stems, and to be sure that durational differences are truly independent of segmental content.

Second, by concealing the task and target words within a meaningful conversation, speakers are unlikely to explicitly attend to orthographic or morphological differences in the targets. In particular, because Plag et al. (2017)—who looked at spontaneous speech—found a durational effect in the opposite direction as previous experimental work, this method allows us to evaluate whether that difference can be ascribed to metalinguistic task effects in the lab reading experiments. While our hypothesis does not predict the result in Plag et al. (2017), if we do replicate their effect in an experiment that uses conversational speech styles, it would suggest that task effects did in fact confound prior experimental work, and thus help reconcile that study's findings with earlier work.

Third, our hypothesis requires that the target words be parsed as having both prosodic and morphological structure. By embedding the words in a meaningful conversational dialogue, speakers are much more likely to generate an appropriate prosody and morphological parse, as compared to if they produce items from a word list or in a fixed carrier phrase. The dialogue context further encourages participants not to produce items with a uniform list intonation, which is argued to interfere with lexical effects on word durations (Gahl, 2008, 2015, Gahl & Strand, 2016).

2. Methods

2.1. Participants

Forty participants were recruited from the UC San Diego community. They each brought a friend to the experiment, and together read through a list of short conversational dialogues that included the target words. All participants and their friends reported that they had started learning to speak English before age 6. Both members of each pair gave informed consent, and both optionally received course credit in exchange for participation.

2.2. Stimuli

The target words were 40 pairs of English homophones in which one member of the pair was uninflected, and the other had an inflectional suffix. 26 pairs had fricative [s] or [z] suffixes (e.g. plural *lapse/laps*, third-person singular *freeze/frees*) and 14 had stop [t] or [d] suffixes (past *duct/ducked*, participle *tide/tied*).

2.2.1. Dialogues

The two homophones in each pair were embedded in phonetically-matched dialogues (Baker et al., 2007; Smith et al., 2012). Each dialogue was a short conversation between two people, and was preceded by a one-sentence description of the scenario in which the conversation took place. For example, the descriptions and dialogues for the target words *freeze/frees* were the following:

Two housemates are wrapping up a surprise birthday party that they put on for a friend.

B: It looks like most people are leaving now. I guess I'm going to start cleaning up a little bit.
A: There's so much cake leftover. I don't want it to go bad.
B: If we freeze it, it should be fine.

Two rural neighbours are talking about a friend, Rich, who is an avid hiker and animal-lover.

B: Rich decided to take care of the injured hawk that he found yesterday.
A: They don't do well in captivity. Wouldn't it be better to let it go?
B: If he frees it, it won't survive.

The complete set of 40 dialogue pairs is given in the Supplemental data. All of the target words received nuclear accent in their phrase. Within each pair of dialogues, each of the two target homophones were preceded by the same number of syllables and stresses in the phrase. If the target homophones were not in the first phrase of the speaker's turn, there were also the same number of syllables, stresses, and phrases between the beginning of the turn and each of the two target homophones. To manage the possibility that a suffix could be resyllabified, the targets were followed by the same segment, or by a phrase boundary.[4] To control for the spread of phrase-final lengthening, each pair of target homophones was followed by the same number of syllables in the phrase and turn. In addition, the targets bore the same type of focus, occurred on the same conversational turn (e.g. the third turn in the dialogue), and where it was possible, the target words (or their phrases) had the same discourse relation with the preceding utterance.

2.2.2. Frequency and orthography

Across pairs, the morphologically-simple and inflected words were not significantly different on log SUBTLEX wordform frequencies (mean of differences $= 0.21$, paired $t(39) = 1.03$, p>0:3), log SUBTLEX word frequency specific to the words' part-of-speech ($\mu_d = 0.20$, $t(39) = 0.87$, p>0.39), or on orthographic length ($\mu_d = -0.33$ letters, $t(39) = 1.65$, p>0.1).

In addition to being matched across the stimulus pairs overall, both frequency measures were matched across the 26 fricative-final pairs alone (frequency: $\mu_d = 0.40$, $t(25) = 1.39$, p>0:17; part-of-speech-specific frequency: $\mu_d = 0.51$, $t(25) = 1.66$, p>0.11) and across the 14 stop-final pairs alone (frequency: $\mu_d = -0.13$, $t(13) = 0.48$, p>0.64; part-of-speech-specific frequency: $\mu_d = -0.39$, $t(13) = 1.49$, p>0.16). Orthographic length was matched overall, as well as across the fricative-final pairs alone ($\mu_d = 0.19$, $t(25) = 0.93$, p>0.36). However, orthographic length was not matched across the stop-final pairs ($\mu_d = -1.29$, $t(13) = -4.84$, p<0.001); we discuss this issue further in Section 4.1.

2.2.3. Predictability norming experiment

Beyond the effects of frequency on word and segment durations, it is well-known that words that are predictable in the discourse context are shortened (e.g. Bell et al., 2009; van Son & Pols, 2003). We estimated the contextual predictability of each word by recruiting 40 different participants for a cloze norming task via Amazon Mechanical Turk, using the JavaScript library `jsPsych` (de Leeuw, 2015). On each trial, each cloze participant saw the first part of one dialogue (including the one-sentence description), which was truncated immediately before the target word. They were asked to complete the partial dialogue with the first word, phrase, sentence or sentence(s) that came to mind. Cloze participants saw half of the 80 dialogues (i.e. only one member of each dialogue pair). We collected 20 individual completion judgments per dialogue.

The predictability of each target word was considered to be the proportion of individuals who wrote down that word immediately following the partial context ($\mu = 0.09$, $\sigma = 0.18$, range $= 0.00$–1.00). On this measure, there was no significant difference between inflected and morphologically-simple words, either across pairs overall or across fricative pairs or stop pairs, by either paired t-test or paired Wilcoxon test (since the distribution of predictability was highly non-normal; all p>0.15). We also used this experiment to estimate the probability of inflectional agreement in the dialogues containing inflected words. For each dialogue containing an inflected word, the probability of inflectional agreement was considered to be the proportion of individuals who wrote down a word with the same inflection immediately following the partial context. For example, in the *frees* dialogue above, the probability of third-person singular agreement was the proportion of participants who completed the truncated phrase in the third turn of the dialogue *"If he …"* with any third-person singular verb. We explore these data further in Section 3.3.3 (see also Cohen, 2014; Pluymaekers, Ernestus, & Baayen, 2005; Rose et al., 2015).

2.3. Procedure

2.3.1. Lists

Each participant pair in the primary experiment read through one of four lists containing half of the dialogues. Each list included one member of each of the 40 homophone pairs, comprising 20 inflected targets and 20 simple targets. The fricative-final and stop-final pairs were evenly divided between the inflected and simple targets, so that the inflected targets in a particular list included half of the 26 fricative-final pairs and half of the 14 stop-final pairs. The first list was constructed by pseudo-randomly selecting one member from each dialogue pair, and sorting them in a random order. The second list was the mirror-image of the first list (i.e. the first list began with the dialogues containing *prize*, while the second list began with *pries*). To control for possible trial order effects, the third and fourth lists were reversed versions of the first two lists. The order of the dialogues is provided in the Supplemental data.

Lists were randomly assigned to participants so that each list was seen by 10 participant pairs. Participants were given their experimental list at least one day in advance. They were instructed to familiarise themselves with the dialogues and to share the list with their friend before arriving for the experiment. They were asked to try to read the dialogues as conversationally and as naturally as possible. During the recording session, participants were given additional time before each item to silently review each dialogue before reading it out loud. To avoid clear speech styles, participants were told not to worry if they stumbled or misspoke, and just to start over where they left off as they would normally do. This resulted in some excluded data, described below in Section 3.1.

2.3.2. Recording

Participants were given the same role (speaker A or speaker B—see example in Section 2.2.1) for all of the dialogues in their list. The target words were always produced by speaker B. Each participant pair sat together in a sound-attenuated booth in a quiet room. Speaker B was sometimes the original participant and sometimes the friend that had been recruited, assigned arbitrarily based on the order in which they entered the booth. Both participants wore head-mounted microphones, and the person given the speaker B role was recorded at a 44.1 kHz sampling rate with 16-bit depth. Although both microphones were set up in the same way, the person assigned to the speaker A role was not recorded.

2.4. Segmentation

Each target word was extracted from the dialogues recorded by the participant pairs, and segmented into two regions. The stem region was the word onset to onset of the final [s, z, t, d] suffix segment. For example, for the words *freeze/frees* [fɹiz], the stem was [fɹi]. For the words *mist/missed* [mɪst], the stem was [mɪs]. The suffix region was the final segment [s, z, t, d].

Segmentation was performed using the waveform and broadband spectrogram view in Praat (Boersma & Weenink, 2016). The acoustic criteria that were used to mark the onset boundary for the stem region are given in Table 1, with the following additional procedures. For five pairs, an onset plosive followed another plosive segment (e.g. in the phrase *ba**d** **b**ruise*). If the preceding segment was unreleased, the midpoint of the two-segment closure was used as the stem onset boundary (e.g. the midpoint of the [db] closure in *bad bruise*). For [l]-initial pairs, if the intensity contour was flat, the onset of a low F2 or high F3 plateau was used as a boundary instead.

The [s, z] fricative suffixes were segmented from the onset to the offset of sibilant noise in the range above 3500 Hz. If there was broadband aspiration noise following the sibilant noise, it was not included in the suffix duration. If a plosive preceded the sibilant (e.g. *lax/lacks* [læks]), the plosive release burst (if any) was not included in the suffix duration.

The [t, d] stop suffixes were segmented from the onset of the closure to the offset of a release burst (if present), or to the end of the closure, if a release burst was not visible. Closure durations were also segmented; all results for [t, d] suffixes were the same when closure durations were analysed alone. If there was no burst, no closure (complete or incomplete), and a relatively small drop in intensity, the segment was considered to be an approximant. If there was also no drop in intensity, no audible percept of a coronal stop, and no visible F2 transition (when adjacent to non-front vowels), it was considered to be deleted. Plosives that were part of a coda cluster (e.g. *duct/ducked* [dʌkt]) were segmented beginning after the first segment's release burst. If no release burst was visible, the midpoint of the two-segment closure was used as the suffix onset boundary.

3. Results

The experiment was run until reaching 40 included participant pairs, with a total of 1600 tokens of the target

Table 1. Criteria used to mark the onset of the stem region in the target words.

Word-initial segment	Example	Onset boundary
[p, t, k, tʃ, b, d, g, m, n]		beginning of closure
[f]	we **f**reeze	onset of broadband frication noise
[s]	we **s**eize	onset of sibilant noise >3500 Hz
[h]	the **h**ose	intensity drop following a vowel
[l]	the **l**aps	onset of low intensity trough
[ɹ]	alread**y** **w**rapped	onset of intensity rise
[oʊ]	a**n** **o**de	end of preceding nasal closure

words (1 of 2 words in each of 40 homophone pairs ∗ 40 participants). Data from one additional participant was excluded without being annotated because of a lisp.

3.1. Exclusions

65 tokens (4.1%) were excluded from all analyses because the target word was disfluent, which was defined as a hesitation immediately before the word, a mispronunciation or speech error on the target word (whether or not the speaker corrected it), or laughter during the word. 40 additional tokens (2.5%) were excluded because the speaker misread the target phrase (e.g. they said *had packed it* instead of *had it packed*). 54 tokens (3.4%) were excluded from the duration analyses because the suffix segment was judged to be deleted (see criteria given in Section 2.4; an analysis of deletion rates appears below in Section 3.3.1), as well as 5 other tokens (0.3%) which had no visible landmarks on the spectrogram that could be used for segmentation.

For the stem duration analysis only, 22 tokens (1.4%) were excluded because they were 2.5 standard deviations or more from the mean stem duration of their respective items. For the suffix duration analysis only, 33 (2.1%) [t, d] tokens were excluded because they were approximated, 29 (1.8%) because they were spirantised, and 6 (0.4%) because they were phrase-final but unreleased, which made it impossible to identify the suffix offset. Additionally, 25 tokens (1.6%) were excluded from the suffix duration analysis because they were 2.5 standard deviations or more from the mean suffix duration of their respective items.

3.2. Models

Stem and suffix durations were analysed in separate linear mixed-effects models. Fixed effects were word type (simple or inflected) and suffix manner (fricative [s, z] or stop [t, d]), plus the interaction. These analyses were planned, designed to replicate a significant interaction found with 20 different participants and variant dialogues in a pilot experiment.[5] Models also included by-item intercepts and slopes for word type, and by-participant intercepts and slopes for all three fixed effects. Each homophone pair was treated as a single item for the purpose of random groupings. Mixed-effects models were fit using the R package `lme4` (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2015).

Effect sizes were estimated with the R package `lsmeans` (Lenth, 2016) by predicting the appropriate marginal means from each model, and then calculating the difference and *p*-value for the contrast with the default Kenward-Roger approximation for degrees of freedom (Halekoh & Højsgaard, 2014). The multivariate *t* distribution was used for multiplicity correction within each family of tests (stem durations and suffix durations).

Table 2. Summary of by-group parameter estimates and residual error. Word type was coded as −1 for simple and 1 for inflected words; manner was coded as −1 for fricatives and 1 for stops. Overall model estimates are given in Figure 2.

		σ (stem model)	σ (suffix model)
By-item	Intercepts	0.066	0.031
	Slopes for word type	0.014	0.004
By-participant	Intercepts	0.025	0.008
	Slopes for word type	0.004	0.001
	Slopes for manner	0.005	0.008
	Slopes for word type by manner	0.006	0.001
Residual error		0.040	0.024

3.2.1. Stem durations

The left panel of Figure 2 shows a summary of stem durations, by word type and manner of the suffix. Figure 3 shows the durations estimated by the model, and Table 2 provides a summary of the by-group parameter estimates. Crucially, for fricative-final words, stem durations were significantly longer in inflected words (*frees*) compared to simple words (*freeze*) ($\hat{\beta} = 18$ ms, $t(39.73) = 2.91$, $p<0.02$). However, for stop-final words, stem durations were not significantly different between word types ($\hat{\beta} = -16$ ms, $t(47.43) = 1.82$, $p>0.14$), with a non-significant effect in the reverse direction.

3.2.2. Suffix durations

The right panel of Figure 2 shows a summary of suffix durations, by word type and manner of the suffix. Figure 3 shows the durations estimated by the model, and Table 2 provides a summary of the by-group parameter estimates. Crucially, for fricative suffixes, suffix durations were significantly longer in inflected words compared to simple words ($\hat{\beta} = 6$ ms, $t(32.51) = 2.73$, $p<0.03$). However, for stop suffixes, suffix durations were not significantly different between word types ($\hat{\beta} = -2$ ms, $t(33.12) = 0.54$, $p>0.8$). Results were qualitatively the same if release bursts were excluded from the suffix region, and stop suffix durations were considered to be the closure only.

3.3. Additional analyses

3.3.1. Deletion rates

Final [t, d]-deletion is a well-attested process in American English. Further, a variety of studies have found that final [t, d] are deleted more often when they represent an inflectional suffix (as in *paced*) than when they do not (*paste*) (Bybee, 2000; Guy, 1980, 1991; Guy, Hay, &

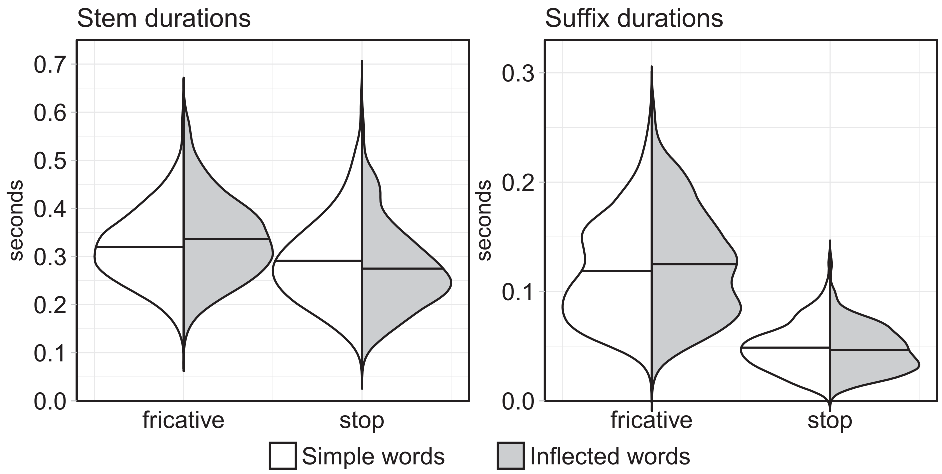

Figure 2. Stem and suffix durations for morphologically-simple and inflected words, by manner of the suffix. The violins are density plots of the empirical durations for simple words (left side of each violin) compared to inflected words (right side of each violin). Horizontal lines show the empirical means for each subgroup.

Walker, 2008; Labov, Cohen, Robins, & Lewis, 1968; Neu, 1980), among others). In our data, deletion rates were roughly the same regardless of inflectional status. Excluding disfluent or misread tokens (Section 3.1), $27/249 = 10.8\%$ of inflected [t, d] suffixes were deleted, and $26/251 = 10.4\%$ of morphologically-simple [t, d] suffixes were deleted (plus 1 token that was both misread and deleted).

To evaluate whether the non-effect was driven by particular items or subjects, we fit a logistic mixed-effects model (using the [t, d] data only) to predict deletion, with a fixed effect of word type (simple or inflected), plus by-item and by-participant intercepts and slopes. The effect of word type was marginally non-significant in the expected direction ($\hat{\beta} = -2.22$, z=1.70, p<0.09). While it is difficult to interpret a null result, the balanced design of the dialogues suggests that the robust differences in deletion rates that have previously been reported may have been partially driven by frequency effects (Guy et al., 2008) or by the different segmental contexts in which simple and inflected words tend to appear (cf. Bybee, 2002).

3.3.2. Frequency and dual-route models

Besides wordform frequency, several other probabilistic measures may influence the realisation of the inflected words in our study. For example, in a dual-route model of morphological processing, morphologically-complex words are accessed through both whole-word representations, and through decomposed constituent forms (e.g. Baayen, 1992; Caramazza, Laudanna, & Romani, 1988; Frauenfelder & Schreuder, 1992; Hay, 2003; Schreuder & Baayen, 1995, and see footnote 1). If the complex wordform has a high frequency relative to its components, it is predicted to behave more like a morphologically-simple form, potentially including stem reduction (Cohen, 2014; Hay, 2003; Losiewicz, 1992; Zuraw & Peperkamp, 2015, though see Hanique & Ernestus, 2012). Additionally, two studies have also found that suffixes are lengthened (Cohen, 2014) or less likely to delete (Schuppler et al., 2012) with higher relative frequency.

This processing model potentially has implications for the paradigm-uniformity account. For example, Winter & Roettger (2011) and Roettger et al. (2014) predict that if a paradigm member is highly frequent, it should exert a stronger influence on its morphological relatives during speech processing. We explored this prediction by examining whether inflected words with high-frequency freestanding stems (such as *guys*) show stronger or weaker effects.

Following Hay (2001) (and others), we also tested whether inflected words with a high frequency relative to their freestanding stems (such as *bored*) show different effects. If relative frequency conditions uniformity effects, this might account for Cohen's (2014) finding that high relative frequency of a complex form is associated with reduced stems (cf. Hay, 2003). High relative frequency means that the inflected form is relatively more frequent than the freestanding stem, and so the influence of the freestanding stem word's independent plan may be weaker (cf. Zuraw & Peperkamp, 2015). In Section 1.3, we argued that the stem word's influence should produce relatively longer durations. Therefore, when this influence is weaker, the stems in inflected words should be relatively shorter.

These analyses should be interpreted with caution, in particular because the stimuli were not selected to include a broad range of either frequency measure. Following the procedure in Section 3.2, we fit a separate linear mixed-effects model to predict stem durations.

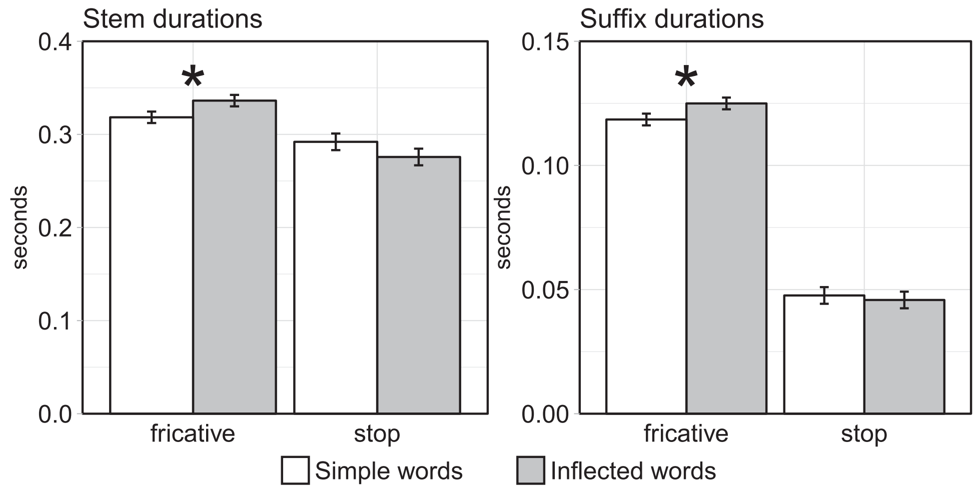

Figure 3. Model estimates for stem and suffix durations for morphologically-simple and inflected words, by manner of the suffix. Error bars show +/− the standard error of the difference between groups.

The model included stem word frequency as an additional fixed effect, as well as all interactions with word type and manner, plus by-participant random slopes. We also fit three models which replaced the stem word frequency parameters with relative frequency, in order to predict stem durations, suffix durations, and stop deletion rates. Stem frequency was the log wordform frequency of the freestanding stem in SUBTLEX. Relative frequency was the log ratio of the inflected word frequency to the freestanding stem frequency.

There was no effect of stem word frequency on stem durations ($p>0.3$), or of relative frequency on stem or suffix durations ($p>0.4$). However, there was a marginally non-significant effect in which inflectional stop suffixes were less likely to delete as relative frequency increased ($\hat{\beta} = 0.99$ per SD, $z=1.95$, $p<0.051$). This supports Schuppler et al. (2012), who had a similar finding for Dutch /t/ suffixes in a corpus of spontaneous speech.

3.3.3. Predictability effects

In addition to frequency, we explored a possible effect of the probability of an inflection in context (Cohen, 2014; Rose et al., 2015). Inflectional probability was estimated using the cloze norming experiment, as described in Section 2.2.3. We fit additional models, following the procedures in Sections 3.2 and 3.3.2, to predict suffix durations and stop deletion rates. As before, because the stimuli were not selected to include a broad range of the predictability measure, these results should be interpreted with caution. There was a marginal effect of inflectional probability, such that inflectional fricative suffixes (but not stop suffixes) were non-significantly longer with lower inflectional probability ($\hat{\beta} = 12$ ms per SD, $t(39.40) = 2.05$, $p<0.10$). There was no significant effect of inflectional probability on stop deletion rates ($p>0.17$).

3.3.4. Speech rate

Differences in participant speech rates should be captured in the by-participant intercepts, which reflect overall participant differences in mean durations. Because of the relatively large number of participants, and the balanced design of the experimental lists (Section 2.3.1), it is unlikely that the results could have been influenced by different global speech rates. In an exploratory analysis, fixed-effects parameters for experimental list were added to the models (Section 3.2), but no list was produced with significantly longer or shorter durations than any other list overall. To evaluate whether participants' speech rates changed over the course of the experiment, an additional parameter for trial number was included. There was an effect of trial number on stem (but not suffix) durations, such that participants slowed down over the course of the experiment ($\hat{\beta} = 12.5$ ms over 40 trials, $p<0.001$), which did not affect the crucial results. There were no significant interactions between trial number and word type or manner of the suffix, and by-participant slopes for trial number did not significantly improve the model.

More generally, it is not clear that a more direct measure of speech rate would be a useful control. A standard way to measure speech rate is to count the number of syllables or segments in a local region near the target. However, many of the target words by design had narrow focus in their phrase, were part of a short phrase, or were phrase- or utterance-final, which means that a local average syllable or segment rate would not necessarily be indicative of how the target word was produced. Prior analyses that use a direct speech rate measure have in fact excluded phrase-final tokens and tokens in short phrases, due to these issues (Bell et al., 2009; Gahl & Strand, 2016; Gahl, Yao, & Johnson, 2012; Seyfarth, 2014).

4. Discussion

English words with English [s, z] inflectional suffixes (e.g. *frees*) had significantly longer stems and suffixes than morphologically-simple homophones (*freeze*). This result supports the phonetic paradigm uniformity account, which predicted that inflected words such as *frees* should be influenced by the phonological plan of their freestanding stems, such as *free*. In particular, we predicted (following Frazier, 2006) that the stem in an inflected word should be lengthened, because the same stem has a lighter coda and a longer duration when it occurs as a word on its own. Further, the free-standing stem word *free* is subject to prosodic domain-final lengthening. This should influence the timing of prosodic gestures within the inflected relative *frees*, such that the stem and suffix are lengthened when they occur within the inflected word *frees*.

4.1. Results for stop suffixes

In addition to the positive results for [s, z] suffixes, we found a null effect for stem and suffix durations when the final suffixes were [t] or [d]. As discussed in Section 1.3, final [t, d] segments are less elastic and less sensitive to domain-final lengthening effects than [s, z], and thus it is not necessarily surprising that we did not find evidence for observable lengthening of [t, d]. Only one prior study has investigated [t, d] durations (Losiewicz, 1992), but also found a null result using a mixed-effects analysis (Plag, p.c.). The prosodic paradigm uniformity hypothesis does predict that the stems should be longer in monosyllabic inflected words, regardless of the final consonant, but it is difficult to interpret a null result. It should be stressed that a null result is not incompatible with the theoretical proposal, but we would consider a positive result in the opposite direction (i.e. significantly shorter stems or suffixes in the inflected words) to be evidence against the hypothesis.

Table 3. Distribution of part-of-speech for simple and inflected words, by manner of the suffix.

Word type	Noun	Verb	Other
Fricatives			
Simple	17	7	2
Inflected	16	10	0
Stops			
Simple	10	2	2
Inflected	0	12	2

There are several factors that may have contributed to the null result. First, in Section 2.2.2, it was observed that orthographic length was unbalanced across stop-final pairs. However, complex words had more letters than simple words, and the stem duration effect for stop-final pairs is in the opposite direction: complex words are non-significantly shorter than simple words. Therefore, it is unlikely that an orthographic confound caused the null result.

An alternative explanation comes from different parts-of-speech between the simple and inflected words. Several corpus studies report significant differences in word duration as a function of part-of-speech (Gahl, 2008; Gahl et al., 2012; Seyfarth, 2014). In particular, nouns tend to be longer in duration than verbs. Because these differences are generally attributed to systematic differences in phrase position and accent in spontaneous English speech (Gahl, 2008; Gahl et al., 2012), and because the items in the current study were matched for phrase position and accent (see Section 2.2), part-of-speech was not intentionally balanced across pairs. Impressionistically, we found that participants were very reliable at accenting the expected word. During segmentation of the data, we noted only 21 tokens (1.3% of the data) in which an unexpected word was accented; exclusion of these data did not qualitatively affect the results.

Nevertheless, it is also plausible that nouns are more likely to attract a stronger prominence than verbs, even with all else held equal, which might result in longer durations. Table 3 shows the distribution of part-of-speech for simple and inflected words (see also Supplemental data), within fricative-final pairs (*freeze/frees*) and within stop-final pairs (*tide/tied*). While the fricative-final pairs were balanced for part-of-speech ($\chi^2 = 2.56$, p>0.27), the stop-final pairs included mainly verbs among the inflected words, but mainly nouns among the simple words ($\chi^2 = 17.14$, p<0.001). This confound could thus have led to a null result for the stop suffixes (with a non-significant trend such that the inflected words were shorter).

To explore this possibility (cf. Conwell, 2016; Li, Shi, & Hua, 2010), we measured homophone vowel pitch as a proxy for phonetic prominence. We fit maximal two mixed-effects models predicting average pitch and pitch slope (maximum pitch minus minimum pitch, divided by the length of the vowel) as a function of part-of-speech (noun, verb, and other). However, there were no significant differences in either pitch measure between the parts-of-speech. This suggests that prominence did not differ between nouns and verbs in our experiment.

4.2. Other accounts

4.2.1. Internal hierarchical prosodic structure

Although we argue that the lengthening of inflected words derives from the influence of their morphological relatives, there are other accounts that may accommodate this result. One proposal is that an English inflected word like *tacks* has a hierarchical prosodic structure, such that the inflectional suffix is adjoined to an internal prosodic-word constituent corresponding to the stem *tack* (Goad, White, & Steele, 2003; Sugahara & Turk, 2009, contra Hall, 2001; Raffelsiefen, 2005). Because syllable rhymes are lengthened before prosodic boundaries at various levels of prosodic constituency (Wightman et al., 1992), the stem within an inflected word should be lengthened.

While this analysis is possible, it entails that the final [z] of *frees* either comprise its own syllable, or else be extra-syllabic (not part of any syllable) (Goad et al., 2003; Sugahara & Turk, 2009). However, psycholinguistic evidence is lacking to support either possibility; and from a formal perspective, there are alternative accounts for the phenomena that extra-syllabicity has been used to explain (Hall, 2002). The exception is the durational lengthening observed here (Sugahara & Turk, 2009, pp. 482–485), which we claim can be explained by a more general uniformity mechanism. Sugahara & Turk (2009, p. 506) argue against the uniformity account on the grounds that it would require including duration—which is highly variable in usage—in phonological representation. However, there is more recent evidence which supports this assumption (Katz, 2010, 2012; Seyfarth, 2014; Tauberer & Evanini, 2009).

4.2.2. Communicative enhancement

One reason to expect that inflectional suffixes might be lengthened is because a suffix like the [z] in *frees* signals a morphosyntactic property (third-person singular agreement, in our materials), whereas the same [z] suffix in a word like *freeze* carries no additional information beyond that conveyed by any other word-final segment (Cohen Priva, 2012; Hanique & Ernestus, 2012; Pluymaekers et al., 2010; Rose et al., 2015). This suggests an alternate explanation for our finding that inflectional [s, z] suffixes were longer: speakers may use the details of phonetic implementation to enhance the perceptibility of a morphological property. The prediction is that [s, z] suffixes should be lengthened when they are unpredictable in context. While previous work has found suggestive effects (Cohen, 2014; Rose et al., 2015, though see Hanique, Ernestus, & Schuppler, 2013), we did not find a significant effect of inflectional probability on suffix duration in our analysis. It may be the case that inflectional suffixes are lengthened in general to enhance intelligibility, independent of the local probability of an inflection. While we consider this account to be plausible, it does not straightforwardly predict that the *stems* in inflected words should be lengthened as well (cf. Cohen, 2014, who found that stems are *shortened* with lower inflectional probability).

A variant of this hypothesis is that small durational (or other phonetic) differences between suffixes might be used as a cue to identify morphological complexity. Kemps, Wurm, Ernestus, Schreuder, & Baayen (2005) and Blazej & Cohen-Goldberg (2014) investigate whether listeners are sensitive to durational differences in the initial syllable of words that do or do not have an additional suffix (*–er, –ly, –less, –ness*). The behavioural results indicated that participants were able to anticipate an upcoming suffix. However, because the experiments compared morphologically-simple monosyllabic words to morphologically-complex polysyllabic words, it is not clear whether listeners were anticipating an additional morphological constituent, or simply an additional upcoming syllable in the stem (see also Lehiste, 1972). There are well-known durational differences in syllable length within monosyllabic and polysyllabic words (Turk & Shattuck-Hufnagel, 2000; White & Turk, 2010), and listeners are able to take advantage of those differences to identify whether a syllable is likely to be part of a polysyllabic word (distinguishing e.g. *captain* from *cap tucked*; Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). Walsh & Parker (1983) tested listeners' sensitivity to durational differences in monosyllabic English simple and inflected homophones (*lapse, laps*), but found a null result (cf. Section 1.4). There is therefore currently no unambiguous evidence that listeners take advantage of morphologically-conditioned durational differences.

4.2.3. Planning costs

A reviewer asks whether the longer durations for the inflected [s, z] words can be interpreted as the result of planning costs. Under this proposal, a word like *frees* is assembled from a phonological stem constituent plus an additional suffix /–z/, whereas a word like *freeze* does not have such multiple phonological constituents that need to be assembled (e.g. Cohen-Goldberg, 2015; Cohen-Goldberg, Cholin, Miozzo, & Rapp, 2013; Levelt et al., 1999). Due to the additional planning complexity and cost during phonological encoding, *frees* is phonetically elongated compared to *freeze*, which is not as costly to assemble.

While greater planning costs for a target word might intuitively be associated with greater naming *latencies*, articulation duration does not necessarily scale with planning costs (Buz & Jaeger, 2015; Kirov & Wilson, 2013). In order to explain our results, this proposal requires a link between planning cost and phonetic duration (see Arnold & Watson, 2015; Buz & Jaeger, 2015; Jaeger & Buz, 2017, for discussion). One proposed link that is relevant to our data involves production fluency. In order to maintain fluent delivery, speakers slow down articulation while waiting for speech planning to be completed (Bell et al., 2009; Christodoulou, 2012; Zerkle, Rosa, & Arnold, 2017), even within a word as encoding and articulation proceed from beginning-to-end (Watson, Buxó-Lugo, & Simmons, 2015). For the assembly of *frees*, the prediction thus might be that the stem will be lengthened in order to buy time for assembly and coordination of the upcoming suffix /–z/.

However, there are two crucial difficulties for this account as an explanation of our results (see Buz & Jaeger, 2015; Jaeger & Buz, 2017). First, speech planning theories generally take the syllable or phonological word as the unit that is passed to the articulation system (Crompton, 1982; Dell, 1986; Levelt et al., 1999; Levelt & Wheeldon, 1994; Wheeldon & Lahiri, 1997). Because all of the stimuli in our experiment were monosyllabic, each full word would need to be assembled before being articulated, and there would be no opportunity for speakers to lengthen the stem while waiting for assembly. Second, even if segments were the unit of articulation rather than syllables or larger units, this account does not predict that the word-final [s, z] segments themselves should be lengthened, as they were in our experiment. By the time articulation of the final [s, z] begins, phonological encoding of the full word must be complete, and the speaker has no need to slow down production.

4.3. Comparison with previous work

Our finding that the English words inflected with [s, z] had longer stems and suffixes than uninflected words agrees with some existing experimental work (Frazier, 2006; Sugahara & Turk, 2009; Walsh & Parker, 1983). However, it does not reconcile the differences between that work and Plag et al. (2017), who analysed suffix durations in a spontaneous speech corpus. In particular, Plag et al. (2017) found that voiceless [s] suffixes were *shorter* in inflected words compared to uninflected ones; as well as a more complicated pattern of differences within several kinds of voiced [z] suffixes (e.g. plural [z] was longer than third-person singular [z]). These corpus results are not predicted by the paradigm uniformity account, or by any current production theories (Plag et al., 2017, pp. 29–32).

Why did our suffix duration results pattern in the opposite direction as Plag et al. (2017)? In an exploratory analysis, we tested interactions between voicing, word type, and manner. The crucial effect of word type on suffix durations (or on stem durations) did not significantly differ between voiced and voiceless suffixes, either overall, within fricatives, or within stops (all $p >$ 0.24). There was a significant effect of voicing on suffix durations ($\hat{\beta} = 31$ ms difference between voiced and voiceless, $t(36.31) = 3.10$, $p<0.01$), but this effect did not qualitatively alter the results reported in Section 3.2.2, nor did it vary significantly by manner ($p>0.3$). Additionally, we found that a model which included the two- and three-way interactions involving voicing was not significantly better than a model with only word type, manner, their interaction, plus voicing with no interactions (by likelihood ratio test; $p > 0.38$). Two methodological differences were the use of read-aloud versus truly-spontaneous conversational speech, and the analysis of homophones versus non-homophones. However, it is unclear whether either consideration would cause the morphological effect to reverse direction, or to interact with voicing as in Plag et al. (2017).

It is also possible that the unbalanced nature of the corpus data in Plag et al. (2017) influenced the analysis. For example, Hsieh, Leonard, & Swanson (1999) find that in natural speech, plural nouns appear in final position much more often than third-person verbs, and are thus lengthened more often. Other work has pointed out that different parts-of-speech (likely correlated with inflectional status) may systematically occur in different prosodic and segmental contexts (see Section 4.1; Bybee, 2002). Thus, it may be the case that the patterns reported in Plag et al. (2017) reflect such systematic differences in context in natural speech, rather than representational or processing differences. Plag et al. (2017) take the possibility of systematic differences into account and include an appropriate variety of lexical and contextual control variables in their models. However, in order to accurately estimate parameters for correlated variables (e.g. suffix type and syntactic position), it is necessary to have many observations in most cells of the design, which may not have been the case (the analysis selected about 650 tokens at random from the corpus). Further, as the authors acknowledge (p. 14), it is challenging to code and statistically control for the effects of diverse prosodic contexts.

4.4. Conclusions

We found that English inflected words with [s, z] suffixes had significantly longer stems and suffixes than uninflected words that were segmentally-identical: *frees* is not homophonous with *freeze*. This supports predictions based on a model of phonetic paradigm uniformity, in which the durational targets of a target word's morphological relatives influence the realisation of that word (Frazier, 2006; Hayes, 2000; Steriade, 2000). We found this result based on a large and diverse set of word types, which were balanced for frequency and orthography, and elicited in phonetically-matched conversational speech designed to avoid metalinguistic task effects that have challenged the interpretability of previous work (see e.g. Bermúdez-Otero, 2010; Hanique & Ernestus, 2012; Mousikou et al., 2015; Plag, 2014; Plag et al., 2017).

This finding challenges discrete accounts of language production in which morphological information does not interact with phonetic realisation (Kiparsky, 1982; Levelt et al., 1999). In particular, the phonetic paradigm uniformity account suggests one specific mechanism involving the cross-influence of phonological plans among morphological relatives (cf. Kuperman, Pluymaekers, Ernestus, & Baayen, 2007), and makes straightforward, testable predictions about phonetic realisation (Frazier, 2006; Kaplan, 2016, and see ongoing work by Abby Kaplan). Future work might investigate especially the cross-linguistic validity of these predictions, and further explore the interaction of probabilistic variables with paradigm uniformity effects.

Notes

1. It is an open question whether the phonological forms of all members of a morphological paradigm are invariably assembled from constituent forms (Cohen-Goldberg, 2015; Cohen-Goldberg et al., 2013), or are activated in distinct lexical entries supported by abstracted forms (Blevins, 2006; Blevins, Ackerman, & Malouf, 2017; Hay & Baayen, 2005; Jackendoff & Audring, 2017). The proposal here assumes only that non-target phonological forms of morphological relatives can become active through spreading activations (which may or may not involve activation of constituent representations, e.g. Dell, 1986).
2. One question involves what elements are included in a word's phonetic-phonological form, which may specify phonological segments, stress, prosodic constituency (Levelt et al., 1999), a range of acceptable phonetic realizations (Goldrick et al., 2011; Lavoie, 2002; Seyfarth, 2014), contrastive or non-contrastive sub-segmental detail (Bybee, 2001; Johnson, 2007; Pierrehumbert, 2002), or other features, all of which which may not be represented together as a single integrated representation (Goldrick, 2014). We assume here that a phonological representation includes, at a minimum, segments and prosodic constituency as well as timing relationships (Katz, 2010, 2012).
3. For example, Frazier (2006) found that vowel durations were longer in inflected words than in morphologically-simple homophones. However, the log wordform frequency in the SUBTLEX-US corpus (Brysbaert & New, 2009; Brysbaert, New, & Keuleers, 2012) was significantly greater for the morphologically-simple words ($\mu = 2.90$) compared to the inflected words in that study ($\mu = 2.02$; unpaired $t(32) = 2.48$, $p<0.05$; excluding two inflected words *brayed* and *rued* which have zero frequency in SUBTLEX). Losiewicz (1992) had the same confound (Hanique & Ernestus, 2012); and see also discussion in Sugahara & Turk (2004, 2009).
4. In some cases in which both words were followed by a vowel or semivowel, they had different qualities. For two pairs, the target words were followed by a different segment, but excluding these from the suffix duration analysis did not qualitatively affect the results. Additionally, the target words in 33 of 40 pairs were preceded by the same segment, or else by vowels or semivowels with a different quality.
5. For the suffix durations, pilot results were the same as those reported here; stem durations and other measures were not analysed. Pilot results are reported by Seyfarth, Garellek, Malouf, and Ackerman (2015, oral presentation).

Acknowledgments

We are grateful to Ingo Plag, the Spoken Morphology Research Unit at Heinrich Heine University Düsseldorf, and the audience at the 3rd American International Morphology Meeting for helpful discussion, comments, and advice. We also thank Alexia Pimentel for assistance with recording the experimental participants. Any errors or omissions are ours.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by a National Science Foundation Graduate Research Fellowship (Division of Graduate Education) to the first author under [grant number DGE-1144086]. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Arnold, J. E., & Watson, D. G. (2015). Synthesising meaning and processing approaches to prosody: Performance matters. *Language, Cognition and Neuroscience*, *30*(1–2), 88–102. doi:10.1080/01690965.2013.840733

Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In *Yearbook of morphology* (pp. 109–149).

Baayen, R. H., Levelt, W. J., Schreuder, R., & Ernestus, M. (2007). Paradigmatic structure in speech production. In *Proceedings from the annual meeting of the Chicago linguistic society* (Vol. 43, pp. 1–29). Chicago: Chicago Linguistic Society.

Baker, R., Smith, R., & Hawkins, S. (2007, August). *Phonetic differences between mis-and dis-in English prefixed and pseudo-*

prefixed words. Proceedings of the 16h international congress of phonetic sciences, Saarbrucken, Germany.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01
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*(1), 2–111. doi:10.1016/j.jml.2008.06.003
Benua, L. (1997). *Transderivational identity* (Unpublished doctoral dissertation). University of Massachusetts, Amherst, MA.
Berkovits, R. (1993). Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech*, *36*(1), 89–98. doi:10.1177/002383099303600105
Bermúdez-Otero, R. (2010). *Morphologically conditioned phonetics? Not proven*. Talk presented at On Linguistic Interfaces II, Belfast.
Blazej, L. J., & Cohen-Goldberg, A. M. (2014). Can we hear morphological complexity before words are complex?. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/a0038509
Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, *42*(03), 531. doi:10.1017/S0022226706004191
Blevins, J. P., Ackerman, F., & Malouf, R. (2017). Word and paradigm morphology. In J. Audring & F. Masini (Eds.), *Oxford handbook of morphological theory*. Oxford: Oxford University Press.
Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (Version 6.0.12)*.
Brewer, J. B. (2008). *Phonetic reflexes of orthographic characteristics in lexical representation*. ProQuest.
Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. doi:10.3758/BRM.41.4.977
Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997. doi:10.3758/s13428-012-0190-4
Bürki, A., Spinelli, E., & Gaskell, M. G. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language*, *67*(4), 449–467. doi:10.1016/j.jml.2012.08.001
Buz, E., & Jaeger, T. F. (2015). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 1–21. doi:10.1080/23273798.2015.1105984
Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). Amsterdam; Philadelphia: J. Benjamins.
Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language*. Stanford, CA: CSLI Publications.
Bybee, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290. doi:10.1017/S0954394502143018
Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*, *120*(3), 1589–00. doi:10.1121/1.2217135
Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180. doi:10.1016/S0095-4470(02)00085-2
Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332. doi:10.1016/0010-0277(88)90017-0
Cho, T., Lee, Y., & Kim, S. (2014). Prosodic strengthening on the /s/-stop cluster and the phonetic implementation of an allophonic rule in English. *Journal of Phonetics*, *46*, 128–146. doi:10.1016/j.wocn.2014.06.003
Christodoulou, A. (2012). *Variation in word duration and planning* (Unpublished doctoral dissertation). The University of North Carolina, Chapel Hill.
Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement: Contextual and paradigmatic effects on morpheme pronunciation. *Morphology*, *24*(4), 291–323. doi:10.1007/s11525-014-9243-y
Cohen Priva, U. (2012). *Sign and signal: Deriving linguistic generalizations from information utility* (Unpublished doctoral dissertation). Stanford University.
Cohen-Goldberg, A. M. (2015). Towards a theory of multimorphemic word production: The Heterogeneity of Processing Hypothesis. *Language and Cognitive Processes*, *28*(7), 1036–1064. doi:10.1080/01690965.2012.759241
Cohen-Goldberg, A. M., Cholin, J., Miozzo, M., & Rapp, B. (2013). The interface between morphology and phonology: Exploring a morpho-phonological deficit in spoken production. *Cognition*, *127*(2), 270–286. doi:10.1016/j.cognition.2013.01.004
Conwell, E. (2016). Prosodic disambiguation of noun/verb homophones in child-directed speech. *Journal of Child Language*, 1–18. doi:10.1017/S030500091600009X
Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109–162). Berlin: Mouton de Gruyter.
Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. doi:10.1037/0096-1523.28.1.218
Davis, S. (2005). "Capitalistic" versus "militaristic": The paradigm uniformity effect reconsidered. In L. Downing, T. Hall, & R. Raffelsiefen (Eds.), *Paradigms in phonological theory* (pp. 107–121). Oxford: Oxford University Press.
de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. doi:10.3758/s13428-014-0458-y
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283– 321.
Ernestus, M., & Baayen, R. H. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology* (Vol. 8, pp. 27–49). Berlin: de Gruyter Mouton.
Ernestus, M., & Baayen, R. H. (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes*, *22*(1), 1–24. doi:10.1080/01690960500268303
Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135. doi:10.1006/jpho.2000.0114

Fourakis, M., & Iverson, G. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, *41*, 140–149. doi:10.1159/000261720

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. *Yearbook of Morphology*, 165–183. doi:10.1007/978-94-011-2516-1_10

Frazier, M. (2006). Output-output faithfulness to moraic structure: Evidence from American English. In C. Davis, A. R. Deal, & Y. Zabbal (Eds.), *Proceedings of the 36th annual meeting of the North East linguistics society* (pp. 1–14). Amherst: GLSA Publications.

Gafos, A. I. (2003). Dynamics in grammar: Comment on Ladd and Ernestus & Baayen. In *Laboratory phonology 8*.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496. doi:10.1353/lan.0.0035

Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics*, *49*, 96–116. doi:10.1016/j.wocn.2014.12.002

Gahl, S., & Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, *89*, 162–178. doi:10.1016/j.jml.2015.12.006

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806. doi:10.1016/j.jml.2011.11.006

Goad, H., White, L., & Steele, J. (2003). Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations?. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, *48*(2), 243–263.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*(7–8), 817–855. doi:10.1080/01690960600824112

Goldrick, M. (2014). Phonological processing: The retrieval and encoding of word form information in speech production. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 228–244). Oxford: Oxford University Press.

Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*(1), 58–72. doi:10.1016/j.cognition.2011.05.006

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. doi:10.1080/01690960500181332

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62*(2), 113–134. doi:10.1016/j.jml.2009.11.008

Goldrick, M., Keshet, J., Gustafson, E., Heller, J., & Needle, J. (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition*, *149*, 31–39. doi:10.1016/j.cognition.2016.01.002

Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (Vol. 1, pp. 1–36). New York, NY: Academic Press.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, *3*(1), 1–22. doi:10.1017/S0954394500000429

Guy, G. R., Hay, J., & Walker, A. (2008). *Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English*. Talk presented at the Laboratory Phonology 11 conference, Victoria University, Wellington, New Zealand.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30. doi:10.18637/jss.v059.i09

Hall, T. A. (2001). The distribution of superheavy syllables in Modern English. *Folia Linguistica*, *35*(3–4), 399–442. doi:10.1515/flin.2001.35.3-4.399

Hall, T. A. (2002). Against extrasyllabic consonants in German and English. *Phonology*, *19*(01). doi:10.1017/S0952675702004268

Hanique, I., & Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio*, *11*(2), 147–164. doi:10.1418/38783

Hanique, I., Ernestus, M., & Schuppler, B. (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *Journal of the Acoustical Society of America*, *133*, 1644. doi:10.1121/1.4790352

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, *39*(6), 1041–1070. doi:10.1515/ling.2001.041

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

Hay, J., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, *9*(7), 342–348. doi:10.1016/j.tics.2005.04.002

Hayes, B. (2000). Gradient well-formedness in optimality theory. *Optimality Theory: Phonology, Syntax, and Acquisition*, *88*, 120.

Hofhuis, E., Gussenhoven, C., & Rietveld, A. (1995). Final lengthening at prosodic boundaries in Dutch. In E. Elenius & P. Branderud (Eds.), *Proceedings of the 13th international congress of phonetic sciences* (Vol. 1, pp. 154–155). Stockholm, Sweden: Stockholm University.

Hsieh, L., Leonard, L. B., & Swanson, L. (1999). Some differences between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration. *Journal of Child Language*, *26*(03), 531–543.

Jackendoff, R., & Audring, J. (2017). Morphology in the parallel architecture. In J. Audring & F. Masini (Eds.), *The Oxford handbook of morphological theory*. Oxford: Oxford University Press.

Jaeger, T. F., & Buz, E. (2017). Signal reduction and linguistic encoding. In E. M. Fernández & H. S. Cairns (Eds.), *Handbook of psycholinguistics*. Hoboken, NJ: Wiley-Blackwell.

Jassem, W., & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, *17*, 317–325.

Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In M. Sole, P. Beddor, & M. Ohala (Eds.), Experimental Approaches to Phonology. In *Honor of John Ohala* (pp. 25–40). Oxford University Press.

Kaplan, A. (2016). *What phonology knows about the lexicon*. Yale University. (Talk presented at Yale University, New Haven, CT).

Katz, J. (2010). *Compression effects, perceptual asymmetries, and the grammar of timing* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, *40*(3), 390–402. doi:10.1016/j.wocn.2012.02.004

Keating, P. A. (2006). Phonetic encoding of prosodic structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167–186). New York, NY: Psychology Press.

Kemps, R. J. J. K., Wurm, L. H., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, *20*(1–2), 43–73. doi:10.1080/01690960444000223

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics*, *43*, 47–56. doi:10.1016/j.wocn.2014.02.002

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

Kirov, C., & Wilson, C. (2013). Bayesian speech production: Evidence from latency and hyperarticulation. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35, pp. 788–793).

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221. doi:10.1121/1.380986

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America*, *121*(4), 2261–2271. doi:10.1121/1.2537393

Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). *A study of the non-standard English of Negro and Puerto Rican speakers in New York City* (Tech. Rep. Nos. Coorperative Research Project 3288, Vol. I). Columbia University.

Lavoie, L. (2002). Some influences on the realization of for and four in American English. *Journal of the International Phonetic Association*, *32*(2), 175–202. doi:10.1017/S0025100302001032

Lee-Kim, S. I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, *4*(2), 475–511. doi:10.1515/lp-2013-0015

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, *51*(6), 2018–2024. doi:10.1121/1.1913062

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. doi:10.18637/jss.v069.i01

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

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269. doi:10.1016/0010-0277(94)90030-2

Li, A., Shi, R., & Hua, W. (2010). Prosodic cues to noun and verb categories in infant-directed Mandarin speech. *Speech Prosody, 100088*, 1–4.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*(3), 172–187. doi:10.1177/002383096300600306

Losiewicz, B. (1992). *The effect of frequency on linguistic morphology* (Unpublished doctoral dissertation). University of Texas at Austin.

McCarthy, J. J. (2005). Optimal paradigms. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in phonological theory*. New York: Oxford University Press.

McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*(1), 44–66. doi:10.1080/01690960801998236

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

Munhall, K. G., Fowler, C. A., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, *20*, 225–239.

Neu, H. (1980). Ranking of constraints on /t, d/ deletion in American English: A statistical analysis. In W. Labov (Ed.), *Locating language in time and space* (Vol. 1, pp. 37–54). Cambridge, MA: Academic Press.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, *54*(5), 1235–1247. doi:10.1121/1.1914393

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 539.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology*, *7*, 101–139.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics*, *26* (2), 209–228.

Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, *53*(1), 181–216. doi:10.1017/S0022226715000183

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*(2–4), 146–159. doi:10.1159/000090095

Pluymaekers, M., Ernestus, M., Baayen, R. H., & Booij, G. (2010). Morphological effects on fine phonetic detail: The case of Dutch –igheid. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology* (Vol. 10, pp. 511–532). Berlin: de Gruyter Mouton.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, *17*, 257–282.

Pouplier, M., & Goldstein, L. (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes*, *25*, 616–649. doi:10.1080/01690960903395380

R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raffelsiefen, R. (2005). Paradigm uniformity effects versus boundary effects. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in phonological theory*. New York: Oxford University Press.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499. doi:10.1037/0033-295X.107.3.460

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

Rose, D., Hume, E., & Hay, J. (2015). *Morphological predictability shapes the phonetic realization of morphemes*. University of Otago. (Talk presented at the Linguistic Society of New Zealand Conference. Dunedin, New Zealand.)

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. doi:10.1016/S0010-0277(03)00139-2

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Hillsdale, NJ: Erlbaum.

Schuppler, B., van Dommelen, W. A., 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 sub-phonemic level. *Journal of Phonetics*, *40*(4), 595–607. doi:10.1016/j.wocn.2012.05.004

Scobbie, J. M., Turk, A., & Hewlett, N. (1999). *Morphemes, phonetics and lexical items: The case of the Scottish vowel length rule*. Proceedings of the XIVth international congress of phonetic sciences, San Francisco. (Vol. 2, pp. 1617–1620).

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155. doi:10.1016/j.cognition.2014.06.013

Seyfarth, S., Garellek, M., Malouf, R., & Ackerman, F. (2015). *Acoustic differences in morphologically-distinct homophones*. Talk presented at the 3rd American International Morphology Meeting, Amherst, MA.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: The effects of speaking rate and coda composition on intra- and inter-articulatory timing. *Journal of Phonetics*, *29*, 89–107. doi:10.1006/jpho.2001.0133

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of phrase-final lengthening in English. In *Proceedings of the 16th international congress on acoustics* (Vol. 2, pp. 1235–1236).

Smith, R., Baker, R., & Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics*, *40*(5), 689–705. doi:10.1016/j.wocn.2012.04.002

Song, J. Y., Demuth, K., Evans, K., & Shattuck-Hufnagel, S. (2013). Durational cues to fricative codas in 2-year-olds' American English: Voicing and morphemic factors. *Journal of the Acoustical Society of America*, *133*(5), 2931–2946. doi:10.1121/1.4795772

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*, 291–311.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the Lexicon* (pp. 313–334). Cambridge: Cambridge University Press.

Strycharczuk, P., & Scobbie, J. M. (2015). *Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English*. Proceedings of the 18th International congress of phonetic sciences, Glasgow.

Strycharczuk, P., & Scobbie, J. M. (2017). Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics*, *59*, 76–91. doi:10.1016/j.wocn.2016.09.003

Sugahara, M., & Turk, A. (2004). Phonetic reflexes of morphological boundaries at a normal speech rate. In B. Bel & I. Marlien (Eds.), *Speech prosody 2004* (pp. 353–356).

Sugahara, M., & Turk, A. (2009). Durational correlates of English sublexical constituent structure. *Phonology*, *26*(03), 477–524. doi:10.1017/S0952675709990248

Tauberer, J., & Evanini, K. (2009). Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of north American English. In *Interspeech* (pp. 2211–2214). Citeseer.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*, 397–440. doi:10.1006/jpho.2000.0123

van Son, R. J. J. H., & Pols, L. C. W. (2003). How efficient is speech? In *Proceedings of the 25th institute of phonetic sciences* (Vol. 25, pp. 171–184).

Walsh, T., & Parker, F. (1983). The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics*, *11*, 201–206.

Warner, N. (2012). Methods for studying spontaneous speech. In A. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 621–633). Oxford: Oxford University Press.

Warner, N., Good, E., Jongman, A., & Sereno, J. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics*, *34*(2), 285–293. doi:10.1016/j.wocn.2004.11.003

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*(2), 251–276. doi:10.1016/S0095-4470(03)00032-9

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing* (Vol. 46, pp. 85–98). Cham: Springer International Publishing. doi:10.1007/978-3-319-12961-7_5

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381. doi:10.1006/jmla.1997.2517

White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, *38*(3), 459–471. doi:10.1016/j.wocn.2010.05.002

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*(3), 1707–1717. doi:10.1121/1.402450

Winter, B., & Roettger, T. (2011). The nature of incomplete neutralization in German. *Grazer Linguistische Studien*, *76*, 55–76.

Withgott, M. (1982). *Segmental evidence for phonological constituents* (Unpublished doctoral 45 dissertation). University of Texas, Austin.

Zerkle, S. A., Rosa, E. C., & Arnold, J. E. (2017). Thematic role predictability and planning affect word duration. *Laboratory Phonology*, *8*, 17. doi:10.5334/labphon.98

Zuraw, K., & Peperkamp, S. (2015). Aspiration and the gradient structure of English prefixed words. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 0382, 1–5). Glasgow, UK: University of Glasgow.