Research Article

Dominic Schmitz*, Dinah Baer-Henney and Ingo Plag

The duration of word-final /s/ differs across morphological categories in English: evidence from pseudowords

https://doi.org/10.1515/phon-2021-2013
Published online October 25, 2021

Abstract: Previous research suggests that different types of word-final /s/ and /z/ (e.g. non-morphemic vs. plural or clitic morpheme) in English show realisational differences in duration. However, there is disagreement on the nature of these differences, as experimental studies have provided evidence for durational differences of the opposite direction as results from corpus studies (i.e. non-morphemic > plural > clitic /s/). The experimental study reported here focuses on four types of word-final /s/ in English, i.e. non-morphemic, plural, and *is*- and *has*-clitic /s/. We conducted a pseudoword production study with native speakers of Southern British English. The results show that non-morphemic /s/ is significantly longer than plural /s/, which in turn is longer than clitic /s/, while there is no durational difference between the two clitics. This aligns with previous corpus rather than experimental studies. Thus, the morphological category of a word-final /s/ appears to be a robust predictor for its phonetic realisation influencing speech production in such a way that systematic subphonemic differences arise. This finding calls for revisions of current models of speech production in which morphology plays no role in later stages of production.

Keywords: morphology; pseudoword paradigm; speech production; subphonemic difference

***Corresponding author: Dominic Schmitz**, English Language and Linguistics, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany, E-mail: Dominic.Schmitz@uni-duesseldorf.de
Dinah Baer-Henney, Linguistics and Information Science, Heinrich Heine University, Düsseldorf, Germany, E-mail: dinah.baer-henney@uni-duesseldorf.de
Ingo Plag, English Language and Linguistics, Heinrich Heine University, Düsseldorf, Germany, E-mail: ingo.plag@uni-duesseldorf.de

1 Introduction

Recent research on the acoustic properties of seemingly homophonous elements has shown unexpected effects of morphological structure on their phonetic realisation. For words, experimental and corpus studies have found evidence that seemingly homophonous lexemes differ significantly in phonetic details such as vowel quality or length (e.g. Drager 2011; Gahl 2008). For stems, Kemps et al. (2005a, 2005b) found that stems in isolation and when suffixed differ acoustically, and that listeners make use of such phonetic cues in speech perception. For prefixes, Ben Hedia and Plag (2017) and Ben Hedia (2019) showed that the more segmentable *un-* or *in-* are (e.g. *un-* is more segmentable for *undid* than *under*), the longer the duration of their nasals.

On the level of individual segments, several studies have shown that the phonetic realisation of word-final /s/ and /z/ in English (henceforth S) depends on its morphological category. In corpus studies, Zimmermann (2016), Plag et al. (2017), and Tomaschek et al. (2019) found that S is longer in stems such as *pass* (henceforth, non-morphemic S) than in morphemic S cases such as the plural suffix in *pats*, which are in turn longer than auxiliary clitics, as in *Pat's gone*. Experimental studies (e.g. Li et al. 1999; Plag et al. 2019; Seyfarth et al. 2017; Walsh and Parker 1983) also found seemingly identical word-final S to be realised differently depending on its morphological category. However, their results are not as clear as those of the previously mentioned corpus studies. One major drawback of all previous studies is the potentially confounding phonetic realisation effects of the lexical and contextual properties of the items under investigation. Examples of such effects are, for instance, prosodic effects arising from different contexts in which the items of interest appear (e.g. phrase-final lengthening effects, e.g. Klatt 1976; Wightman et al. 1992), uncontrolled lexical frequencies (high frequency words show shorter segment durations, e.g. Lohmann 2018), unbalanced distributions of items across different categories (e.g. analysing pooled data on word-final /s/ and /z/ with only a small number of data points for /s/, e.g. Seyfarth et al. 2017), or differences in informativity (i.e. the predictability of the word in its context or in its paradigm, e.g. Bell et al. 2009; Cohen 2014; Jurafsky et al. 2001; Kuperman et al. 2007; Pluymaekers et al. 2005a; Cohen Priva 2015; Seyfarth 2014; Tang and Shaw 2021; Torreira and Ernestus 2009; Tucker et al. 2019; Zee et al. 2021).

Most importantly, as traditional models of speech production assume that phonetic processing does not have access to information on morphological makeup (e.g. Levelt and Wheeldon 1994; Levelt et al. 1999), morpho-phonetic effects pose a serious challenge, calling for an explanation on how morphological information would come to influence articulation.

The present study addresses realisational differences in individual segments based on different types of word-final S in English. We investigate whether different types of word-final S, i.e. non-morphemic, plural, and *is-* and *has*-clitic S, show differing phonetic realisations in terms of duration. This, for the first time, will be done within a pseudoword paradigm in order to provide further insight into subphonemic realisational differences beyond lexical and contextual properties. We suggest that if systematic differences can also be found within pseudoword paradigms, one can assume realisational differences between seemingly identical segments in morphologically-differing structures to be of a robust nature rather than a by-product of confounding lexical or contextual factors. This would in turn call for a revision of models on the relationship between morphology, phonology, and phonetic realisation.

The paper is structured as follows. In the next section, we will take a closer look at the interplay of morphological structure and the phonetic signal. Section 3 will present our methodology. The analysis and results of our study are presented in Sections 4 and 5, followed by a discussion and conclusion in Section 6.

2 Morphology and phonetic realisation

In English, a number of morphological categories can take the phonological form of /s/ (phonetically realised as [s] or [z]), i.e. plural, genitive, genitive plural, third person singular, as well as the clitics of *is*, *has*, and *us* (as in *let's*). As such, there is nothing in the segmental representation of the morphological categories that accounts for systematic realisational differences on the phonetic level between different S morphemes, or between morphemic and non-morphemic S. Any such difference is therefore unexpected from traditional views on the planning and production of speech segments.

However, there is growing evidence for the presence of morphological information in the phonetic signal (in general, and with regard to word-final S), and this evidence is a challenge for existing theories of morpho-phonology and of speech production. In this section we will first review the empirical evidence for morpho-phonetic effects, zooming in on final S in English. We will then turn to pertinent theories to develop the hypotheses about the morpho-phonetic effects to be tested in this study.

The evidence for the presence of morphological information at the phonetic level emerges mainly from the study of homophonous lexemes, stems and affixes. For homophonous lexemes, Gahl (2008) and Lohmann (2018) investigated acoustic realisations of seemingly homophonous word pairs such as *time* and *thyme*, and found the more frequent member of each pair to be of shorter duration. Further evidence for differing acoustic realisations of supposedly homophonous lexemes was found by Drager (2011). Drager compared realisations of *like* as adverb, verb, discourse particle, and as part of the quotative *be like*. Differences surfaced in several phonetic parameters. Similar effects were found for function words such as *four* and *for* and different uses of words such as *to*, which were investigated by Lavoie (2002) and Jurafsky et al. (2002). Such fine realisational differences indicate that at the phonetic level two or more phonologically homophonous lemmas may differ in their realisation.

Similarly, evidence shows that seemingly homophonous elements below the word level have different phonetic realisations. Kemps et al. (2005a, 2005b) found that in Dutch and English segmentally identical free and bound variants of a base (e.g. *help* without a suffix versus *help* in *helper*) differ acoustically. Sugahara and Turk (2004, 2009) found phonetic differences between the final segments of a mono-morphemic stem as compared to the final segments of the same stem if followed by a suffix, e.g. in *mist rain* versus *missed rain*. The stem had slightly longer rhymes if followed by certain suffixes. Seyfarth et al. (2017) found that for words ending in fricatives the durations of a word's morphological relatives influence the realisation of that word. In their study, stems of multi-morphemic words showed longer durations than similar strings of segments in homophonous mono-morphemic words (e.g. *free* in *frees* vs. *freeze*). They concluded that the durational targets of the multi-morphemic word's relatives influence the word's duration to such an extent that a durational difference between the pertinent multi-morphemic word and its homophonous mono-morphemic counterpart arise. A similar effect of morphological relations influencing duration was found for plurals and their bare stems in a corpus-based study by Engemann and Plag (2021).

For prefixes, Smith et al. (2012) found systematic realisational differences for *dis-* and *mis-* between prefixed and so-called pseudo-prefixed words (e.g. *discolour* vs. *discover*). Prefixed words showed longer durations and longer voice onset times, among other things. Ben Hedia and Plag (2017) and Ben Hedia (2019) showed that the more segmentable a prefix the longer the duration of its nasal.

On the articulatory level, Cho (2001) found evidence for the variability of intergestural timing between identical strings in mono- versus multi-morphemic contexts. In their electropalatographic study, Cho showed that the timing of the gestures for [ti] and [ni] in Korean shows more variation when the sequence is mono-morphemic (/mati/ 'knot' and /pani/ 'name') as compared to the timing of the same gestures in multi-morphemic sequences (/mat-i/ 'the oldest' and /pan-i/ 'class-Nom'), thus indicating that morphological structure is reflected in articulatory gestures, which in turn may lead to correlates in the acoustic signal. Thus, morphology is reflected in the phonetic realisation of otherwise identical strings of segments.

Thus, it seems that there is vast evidence for seemingly homophonous elements, i.e. lexemes, bases and affixes, to differ on the level of speech production. Differences on the level of segments have been reported as well. Previous corpus studies on word-final S in English found realisational differences between non-morphemic, suffix and clitic variants. Zimmermann (2016) on New Zealand English (data from QuakeBox corpus; Walsh et al. 2013), and Plag et al. (2017) as well as Tomaschek et al. (2019) on North American English (data from Buckeye Corpus of Conversational Speech; Pitt et al. 2007) find that non-morphemic S showed longer durations than suffix and clitic S. In turn, suffix S also showed longer durations than clitic S. While these results draw a clear picture of S duration across morphological categories (including the non-morphemic S), they are subject to unbalanced data sets due to the nature of corpora. That is, corpus data may contain a huge number of confounding and moderator variables that experimental data can control for (Gries 2015).

Previous experimental studies, however, have reported less consistent results and show some problematic methods and analyses. Walsh and Parker (1983) carried out a production experiment with three homophonous word pairs (e.g. *Rex* and *wrecks*). They measured the duration of the word-final S in both the mono- and the multi-morphemic word of each pair in three different conditions. Each word was produced by eight to 10 participants. Condition I consisted of an unambiguous context; condition II consisted of a semantically neutral context; Condition III consisted of a semantically anomalous context. While in two of these conditions there was a small difference of 9 ms in the means of the different types of S, there was none in the third condition. Still, they concluded that 'speakers of English systematically lengthen morphemic /s/' (Walsh and Parker 1983: 204). However, their analysed data set was small (110 observations), included a mixture of common and proper nouns, and no phonetic covariates were integrated in their analysis. Further, instead of applying appropriate inferential statistical methods (e.g. t-tests or more advanced methods), the mean durations of the types of S under investigation were compared impressionistically. Therefore, there are several reasons to be sceptical of their results.

In another study, Li et al. (1999) measured S duration in child-directed speech on data originally elicited for another study, on vowel durations in function words (see Swanson and Leonard 1994), which found plural S to be longer than third person singular S. However, as the study originally was not designed for this endeavour, half of all plural items occurred sentence-finally, while almost all third person singular items occurred sentence-medially. The durational difference found between the suffixes may hence have been due to effects of phrase-final lengthening (e.g. Klatt 1976; Wightman et al. 1992) rather than to inherent phonetic differences due to morphological categories.

In a more recent study, Seyfarth et al. (2017) conducted a production experiment to collect data on non-morphemic, plural, and third singular /s/ and /z/ durations. They found the non-morphemic variant to be shorter than the morphemic instances. However, they did not find differences between the voiced and the voiceless allomorphs during their analysis. This may be a worrisome result especially due to the small number of items with voiceless allomorphs (n = 6) as compared to the high number of items with voiced allomorphs (n = 20) in their data.

Most recently, Plag et al. (2019) found plural and genitive plural S to be of different durations. In their study, the genitive plural suffix showed significantly longer durations as compared to the plural suffix. An overview of the durational differences found in the aforementioned experimental studies is given in Table 1.

Table 1: Overview of durational differences of word-final S found in previous studies.

Study	Findings
Zimmermann (2016), Plag et al. (2017), and Tomaschek et al. (2019)	non-morphemic > plural > clitics
Walsh and Parker (1983)	plural > non-morphemic
Li et al. (1999)	plural > 3rd singular
Seyfarth et al. (2017)	plural > non-morphemic
Plag et al. (2019)	genitive plural > plural

In sum, there is evidence that there may be durational differences between different types of S. However, while results of corpus studies are in line with each other, they might be flawed due to imbalanced data sets. Previous experimental studies, on the other hand, have often relied on small data sets, and lacked phonetic covariates, appropriate statistical methods, or a proper distinction of voiced and voiceless segments. Another crucial difference between corpus and experimental studies is the use of homophones. While all previous experimental studies restricted their data to homophone pairs, corpus studies take into consideration all words. The limitation to homophones and the resulting competition between their representations might be a problem in itself as it appears to be unclear how members of homophone pairs are stored and connected to their respective frequencies (see Section 2.2). In all cases, previous results were subject to potentially confounding effects of the lexical properties (e.g. effects of frequency, e.g. Gahl 2008; Lohmann 2018; effects of storage, e.g. Caselli et al. 2016) and contextual effects (e.g. phrase final lengthening, e.g. Klatt 1976; Wightman et al. 1992) of the items under investigation. Also, so far, no experimental study included clitics in their analysis whereas corpus studies have suggested that clitics show different durations than suffixes.

A study is therefore called for that investigates the durational nature of different types of word-final S in English, preferably an experimental study with carefully controlled data avoiding potentially confounding effects. This paper presents such a study investigating word-final S in English by means of a pseudoword production task. In this task, we elicited three types of word-final S: monomorphemic, plural, and clitic S (with the auxiliaries *is* and *has*). We will address some the issues of previous studies. That is, the use of pseudowords prevents potential lexical effects to confound our findings, while our highly controlled task evades the influence of contextual effects. Even though our data will also contain homophones to a certain extent, the individual members do not have lexical representations. That is, we can rule out effects of competition between homophonous lexical entries due their similar representations. In addition, the use of pseudowords eliminates potential differences in duration due to differences in frequency between the homophones.

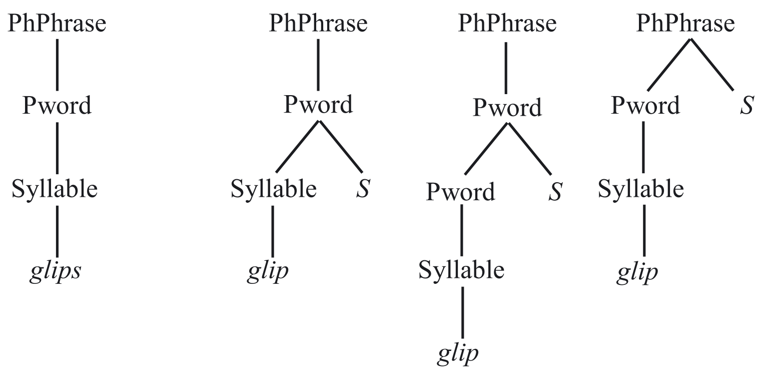
Let us now turn to the question of how morpho-phonetic effects can be explained at the theoretical level. Existing theories make different predictions concerning the possible presence of durational differences between different types of S. We will discuss four approaches here: Feed-forward models of phonology-morphology interaction, Prosodic Morphology, exemplar theory and discriminative learning. One possible source of phonetic differences between different types of word-final S could lie in the prosodic structure.

In standard feed-forward formal theories of morphology–phonology interaction, all types of S, be they morphemic or non-morphemic, are treated in a similar way (e.g. Chomsky and Halle 1968; Kiparsky 1982). In the case of morphological word-final S, a process called 'bracket erasure' is said to remove all morphological information from a pertinent word form once retrieved from the lexicon during the stage of 'lexical phonology' and leaves speech production without an insight into the morphological makeup at the stage of 'post-lexical phonology'. Once retrieved, there is no informational difference between word-final morphemic and non-morphemic types of S. Thus, there is nothing in such a system that could account for realisational differences, e.g. different durations, between phonologically identical suffixes and non-morphemic segments. The realisation of clitics is a post-lexical process to begin with, and thus outside the scope of any prediction by this theory.

In the framework of Prosodic Phonology, there is a complex mapping of morphological structure onto prosodic structure (e.g. Booij 1983; Nespor and Vogel 2007), since prosodic boundaries may correlate with particular phonetic properties, segments at such boundaries may show systematic differences in phonetic implementation (see, for example, Keating 2006). Phonetic differences between two phonologically homophonous affixes could therefore result from a difference in the prosodic structure that goes with the two affixes. In particular, different types of word-final S can be analysed as having different positions in the hierarchical prosodic configuration. These configurations co-determine the degree of integration of an S to the word it belongs to. These different degrees of integration might then emerge as durational differences between types of S in speech production.

Applying Selkirk (1996) approach, non-morphemic S, uncontroversially, is an integral part of the prosodic word, as shown in (1). Goad (1998) analyses plural S as an ‘internal clitic’, which is adjoined to the highest prosodic constituent below the prosodic word, as shown in (2). In Goad (2002), however, plural S is analysed as an ‘affixal clitic’, like third person singular S in Goad et al. (2003) and Goad and White (2019), as shown in (3). The prosodic status of the cliticized auxiliary S is not entirely clear, but presumably it is best analysed as ‘free clitic’, as in (4).

(1)	(2)	(3)	(4)
non-morphemic S	plural S	plural S	clitic S
	‘internal clitic’	‘affixal clitic’	‘free clitic’



The prosodic phonology approach thus posits a structural prosodic difference between non-morphemic S, plural S and clitic S. This prosodic difference might be mirrored in durational differences. It is, however, not so clear what particular phonetic effects this approach would predict, and by which processing mechanism the structural prosodic differences would be translated into different articulations. The most plausible prediction would be that closer integration into the prosodic word would correlate with shorter durations. That is, non-morphemic S should be shortest, clitic S longest, and plural S in between. From the perspective of phrase-final lengthening (e.g. Klatt 1976) one should also expect that clitic S is longest, as it immediately precedes a phrase boundary.

The distinction of lexical and post-lexical processing is also an integral part of established theories in psycholinguistics. According to models of speech production such as the one proposed by Levelt et al. (1999; see Roelofs and Ferreira 2019 for an update), morphemic S would not differ in realisation from corresponding non-morphemic realisations of S. In such models, meanings are stored in the mental lexicon with their forms being represented phonologically. The module called 'articulator' uses these phonological forms for speech production, hence, has no information on the lexical origin of particular segments. As a consequence, in this architecture no systematic differences between different types of S should emerge.

In contrast, exemplar-based models (e.g. Bybee 2001; Gahl and Yu 2006; Goldinger 1998; Pierrehumbert 2001, 2002) have an architecture that would in principle allow for morpho-phonetic effects. In such models, lexemes are linked to a frequency distribution over their phonetic outcomes as experienced by the individual speaker. These distributions are updated with each new experience: experienced subtle subphonemic differences then may result in representations mirroring these properties. While such an account may allow for durational differences between different types of word-final S to emerge from stored phonetic representations, it leaves open the question of how such systematic differences between clouds of exemplars would come about in the first place. The downside of this is that it is also unclear in which direction differences between different types of S should play out.

Finally, there is the discriminative learning approach, which is based on simple but powerful principles of discriminative learning theory (Ramscar and Yarlett 2007; Ramscar et al. 2010; Rescorla 1988; see, for example, Baayen et al. 2011, Baayen et al. 2019; Blevins et al. 2016 for its application to linguistic problems). According to this theory, learning results from exposure to informative relations among events in the individual's environment. Individuals use the associations between these events to create cognitive representations of their environment. Most importantly, associations and their resulting representations are updated constantly on the basis of new experiences. Associations are built between features ('cues', e.g. biphones) and classes or categories ('outcomes', e.g. different types of S) that co-occur in events in which the learner is predicting the outcomes from the cues (Tomaschek et al. 2019: 11). The relation between cues and outcomes is modelled mathematically by the so-called Rescorla–Wagner equations (Rescorla 1988; Rescorla and Wagner 1972; Wagner and Rescorla 1972). Following these equations, an association strength or 'weight' increases every time a cue and an outcome co-occur, while it decreases if a cue occurs without the outcome in a learning event. This results in a continuous recalibration of association strengths, which is a crucial part of discriminative learning.

In recent implementations of discriminative learning, the association weights between semantic representations and phonetic representations have been shown to be predictive of phonetic durations (e.g. Stein and Plag 2021). With regard to final S, Tomaschek et al. (2019) show that the different durations of final S can be understood as following from the extent to which words' phonological and collocational properties can discriminate between the inflectional functions expressed by the S. The input features (cues) for their discriminative network were the words ('lexomes' as pointers to the meaning of the forms) in a five-word window centred on the S-bearing word and the biphones in the phonological forms of these words. These cues are associated with the inflectional functions of the S. Two main measurements emerged as significant predictors of S duration. The so-called 'activation' ('named 'prior' in Tomaschek et al. 2019) is a measure of an outcome's baseline activation, i.e. of how well an outcome is entrenched in the lexicon. The other measure is 'activation diversity', which quantifies the extent to which the cues in the given context also support other targets. The general pattern now is the following: When the uncertainty about the targeted outcome increases, the acoustic duration of S decreases. In other words, stronger support (both from long-term entrenchment and short-term from the context) for a morphological function leads to a longer, i.e. enhanced, acoustic signal.

In sum, the discriminative approach predicts that differences between different types of S may emerge from the associations of form and meaning that the speakers develop as a result of their experience with the pertinent words. But what about pseudowords? It has recently been shown (Chuang et al. 2020) that these associations also play a role for pseudowords. Pseudowords have no representation in the lexicon, but, as these authors show, pseudowords nevertheless resonate with the lexicon due to their formal similarity with existing words. This resonance even influences subtle phonetic details such as duration (Chuang et al. 2020). It is, however, yet unclear what kinds of durational differences can be expected between different types of S in nonce words.

Finally, effects of informativity or predictability (which are also inherently present in discriminative learning approaches) may also play a role (e.g. Cohen Priva 2015; Seyfarth 2014; Zee et al. 2021). Thus, greater predictability of the word in its context has been found to lead to phonetic reduction, i.e. for example, shortening in duration. On the other hand, higher paradigmatic predictability has been shown to correlate with longer duration ('paradigmatic enhancement', Bell et al. 2020; Kuperman et al. 2007). As these informativity effects are necessarily bound to existing words, an experiment that uses pseudowords cannot straightforwardly test these approaches.

Based on the different theories laid out above, different hypotheses about durational differences between different types of S in pseudowords can be set up. They are given in (5) to (7). Hypothesis 1 ('Feed-forward Hypothesis') arises from feed-forward approaches and is in accordance with the prediction that no systematic phonetic differences should be observable between different types of S. Hypothesis 2 ('Prosodic Hypothesis') is derived from prosodic approaches. According to these approaches, a higher degree of prosodic integration should correlate with shorter durations. Hence, non-morphemic S should be shorter than plural S, and plural S should be shorter than clitic S. Finally, exemplar-based approaches and discriminative learning approaches both predict the presence of morpho-phonetic effects, but it is unclear how these differences would play out for the three types of S in the present study. This is encapsulated in Hypothesis 3 ('Emergence Hypothesis'). As it stands, the Emergence Hypothesis is a rather weak hypothesis because, unlike the Prosodic Hypothesis, it does not make any clear predictions concerning the expected pattering of differences. Presently, no exemplar-based computational implementation is available that could be used to explore potential durational effects. But pertinent work is available for the discriminative learning approach.

Tomaschek et al. 2019 showed the feasibility of the discriminative learning approach for modeling the duration of final S. In their analysis, stronger support for a morphological function leads to an enhanced, i.e. longer, acoustic signal. This relationship between network support and duration would also be predicted to hold for the present data set. However, this prediction can only be tested by implementing a discriminative learning model. The present paper has a much more modest aim, however. The present study wants to establish whether there are durational differences also with nonce words, and if so, how these differences play out. Support for the Emergence Hypothesis would pave the way for future studies that test whether the patterning of these differences may emerge via discriminative learning.

(5) Hypothesis 1: Feed-forward Hypothesis
There is no durational difference between word-final non-morphemic S, plural S and auxiliary clitic S .

(6) Hypothesis 2: Prosodic Hypothesis
There are durational difference between different types of word-final S: non-morphemic S is shorter than plural S, plural S is shorter than auxiliary clitic S.

(7) Hypothesis 3: Emergence Hypothesis
There are durational differences between different types of word-final S (non-morphemic, plural and auxiliary clitic).

3 Methods

3.1 Speakers and recordings

Forty native speakers of Southern British English took part in the experiment. Twenty-six of them were female and 14 were male. Their mean age was 28.7 years, ranging from 19 to 58. Eight speakers were bi- or multilingual, and 25 speakers were from London while the other 15 speakers were from other places in South Britain. The participants had no background in linguistics.

The recordings took place at Chandler House, University College London. The acoustic data were recorded on a computer with a Røde NT1 – a microphone using an RME Fireface UC audio interface and sampled at 44.1 kHz, 16 bit.

3.2 Speech material

We adopted Berko-Gleason's (1958) pseudoword paradigm for the production experiment, using a total of 48 pseudowords. Following her reasoning, we assume phonetic effects found in pseudoword paradigms to mirror linguistic reality. Our pseudowords followed the phonotactic constraints of English (Clements and Keyser 1983) and contained a complex onset consisting of a plosive and an approximant (/pl/, /bl/, /kl/, /gl/, /pr/), and either a short vowel (/ɪ/, /ʌ/), a long vowel (/iː/, /uː/), or a diphthong (/aʊ/, /eɪ/) as nucleus. One half of the pseudowords had simple codas (/p/, /t/, /k/, /f/), while the other half had an additional voiceless alveolar fricative (/ps/, /ts/, /ks/, /fs/). The set of coda consonants preceding the S was chosen in such a way that the voiceless realisation of the S allomorphs was elicited. Our study is restricted to the voiceless realisation as clearest results have emerged from literature for voiceless S. Pseudowords with complex codas were used to elicit non-morphemic S, while pseudowords with simple codas were used to elicit morphemic types of S. The pseudowords used in the experiments are given in Table 2.

One issue when constructing pseudowords is their spelling. For vowels, orthographic representations were chosen following the highest phonotactically legal grapheme-phoneme probabilities (Gontijo et al. 2003). The aforementioned coda consonants, however, showed a variety of possible orthographic representations to choose from. That is, /p/ may be represented by <p> or <pp>, /t/ may be represented by <t> or <tt>, /k/ may be represented by <k>, <c>, or <ck>, and /f/ may be represented by <f>, <ph>, or, exceptionally, by <gh>. When combined with a coda-internal /s/, some additional options can be observed: /ks/ may not only be represented as <ks>, <cs> or <cks> but also as <x>, /ps/ may be represented as <ps>, <pps>, and <pse>, and /ts/ may be represented as <ts>, <tts>, and <tz>. The choice of orthographic representation is important for two reasons. First, when comparing two kinds of words, variable representations add another source of variation of unclear consequences and should be avoided. Second, studies on the influence of number of letters on spoken language production have found that increasing the number of letters to represent a single sound may go together with longer durations in speech (e.g. Brewer 2008). Based on these considerations, the following orthographic representations were chosen for all word-final clusters: /ks/ is represented uniformly in spelling as <ks>, /ps/ is represented uniformly as <ps>, /ts/ is represented uniformly as <ts>, and /fs/ is represented uniformly as <fs>.

A second potential problem with the pseudowords constructed for this study is their phonotactics. All our pseudowords are phonotactically legal, and their final consonant clusters (with /s/ as the second consonant) are not uncommon in multi-morphemic words. However, in mono-morphemic words these clusters are rarer, or, in the case of /fs/, even unattested (e.g. in CELEX, Baayen et al. 1995). The different phonotactic probabilities of these clusters could potentially influence the pronunciation of /s/ in our nonce words, especially when spoken in the contexts where these words receive a mono-morphemic interpretation. We have included two measures in our regression models to control for phonotactic probability. First, we included the biphone probability sum (Vitevitch and Luce 2004) as a general measure of phonotactic probability of the whole word-form. Second, we included biphone probability to control for potential transitional effects resulting from having a different consonant preceding the S.

Table 2: Orthographic representation of the complete stimulus set.

	ɪ	iː	uː	ʌ	aʊ	eɪ
items for morphemic S elicitation	glip	pleep	cloop	prup	bloup	glaip
	glit	pleet	cloot	prut	blout	glait
	glik	pleek	clook	pruk	blouk	glaik
	glif	pleef	cloof	pruf	blouf	glaif
items for non-morphemic S elicitation	glips	pleeps	cloops	prups	bloups	glaips
	glits	pleets	cloots	pruts	blouts	glaits
	gliks	pleeks	clooks	pruks	blouks	glaiks
	glifs	pleefs	cloofs	prufs	bloufs	glaifs

To elicit the types of S under investigation, 48 contexts and accompanying questions for S elicitation were created. The verbs[1] directly following the pseudowords in these contexts were chosen in such a way that out of 12 verbs in total, three each started with a voiceless plosive (/pl/, /k/), a vowel (/ɑ/, /iː/, /ʌ/, /eɪ/), a

1 A reviewer pointed us towards a potential influence of the following verb on clitic S in that some verbs used in the experiment are strong verbs (e.g. *write*) while most are weak verbs (e.g. *watch*). To test the possibility that the difference in cue strength between weak verbs and strong verbs might have had an influence on the duration of the clitic S, we did a post-hoc analysis with the difference between weak and strong verbs included as an additional predictor. There was no influence ($p > 0.05$).

nasal (/m/, /n/), and an approximant (/w/, /l/, /r/). Examples are given in (8) to (11) with verbs in bold print (see Appendix A for all contexts). This was done to control for possible coarticulatory effects of these segmental classes with the preceding S.

(8) Every day, the *glips* **plays** with the cloops.

(9) Two days ago, the *glips* **ate** their lunch together.

(10) Tonight, the *glip's* **meeting** the cloop for a drink.

(11) The *glip's* **written** a love letter to the cloop.

To keep priming effects to a minimum, pseudowords were split into two groups. Each group consisted of 24 pseudowords, with 12 pseudowords used for morphemic S elicitation and 12 pseudowords used for non-morphemic S elicitation. This way we ensured that no single participant encountered a phonologically identical pseudoword as both mono- and multi-morphemic, i.e. no participant was to encounter /glɪps/ as both singular and plural/clitic item. Participants were distributed equally across both groups. Each participant was supposed to produce 12 tokens for each of the four types of S (non-morphemic, plural, *is*-clitic, *has*-clitic; 48 tokens overall).

To ensure that each pseudoword was elicited within each context, i.e. with each verb for each type of S, 12 pseudorandomized lists were created. The same 12 lists were used for both groups to keep them comparable. Additionally, types of S were alternated in such a way that no type of S was elicited twice in a row. This was done to keep priming effects to a minimum.

3.3 Procedure

First, participants were introduced to the idea of a recently discovered far away planet. They were told that the inhabitants of this planet at first might appear bizarre, but engage in activities known to the participants, and not to worry about the unfamiliar names of the creatures. Second, the trial structure was explained, i.e. for each slide there would be pictures and names of alien creatures, a short explanation of a situation, and a question relevant to the situation which was to be answered aloud. Participants were then told to proceed in a natural pace and to take as much time as necessary to read and understand the aliens' names as well as the situations. To avoid possible confusion due to the simplicity of the task at hand, participants were made to believe that they were part of a control group of an experiment originally designed for children. Before starting practice trials, participants were reminded to use the aliens' names instead of pronouns when answering. Then, a practice set of four contexts (see Appendix B) was used to familiarize the participants with the experimental procedure itself.

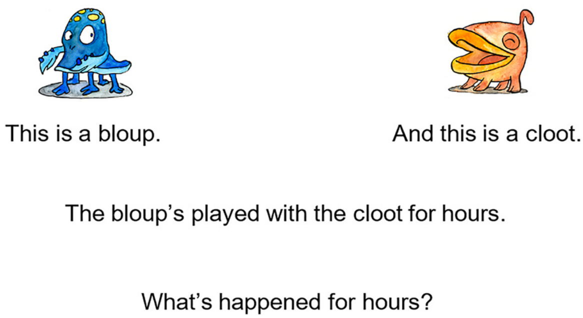


Figure 1: Item, context and question display during the production experiment.

For each trial, the screen proceeded similarly (see Figure 1 as well as examples (12) to (15)): First, the pertinent pseudoword(s) were introduced. In the stimuli testing the plural, one pseudoword (in its plural form) was introduced, while in the other three conditions two different pseudowords were introduced. In either case, two images (van de Vijver and Baer-Henney 2014) representing the pseudowords were used to create familiarity with the items under investigation. In all cases but plural, two images of different creatures were given, while in plural contexts, two images of the same creature were used. The pseudowords and images were paired randomly across lists to rule out possible confounding effects of appearance, e.g. the bouba/kiki effect (e.g. Fort et al. 2015; Köhler 1929). Second, a context was introduced. Third, a question was given to elicit an answer with the pertinent type of S while the context slowly faded out. The fading out of the question forced the participants not to rely on the reading-aloud of the given context. This open format was chosen in order to elicit speech that is as natural as possible. By choosing such an open format one obviously runs the risk of eliciting a large proportion of responses that do not contain the desired forms. This drawback of our design was countered by having a large number of trials and participants. This strategy resulted in a sufficient number of observations. The experiment was carried out in a self-paced fashion; participants were instructed to progress in a contextually appropriate manner and at a speaking rate they considered to be normal.

(12) non-morphemic context

Introduction:	This is a glaits. # And this is a pleeps.
Context:	Every day, the glaits plays with the pleeps.
Question:	What happens every day?
Answer:	The glaits plays with the pleeps.

(13) plural context

Introduction:	This is a glait. # And this is another one.
Context:	Two days ago, the glaits ate their lunch together.
Question:	What happened two days ago?
Answer:	The glaits ate their lunch together.

(14) *is*-clitic context

Introduction:	This is a glait. # And this is a pleep.
Context:	Tonight, the glait's meeting the pleep for a drink.
Question:	What's happening tonight?
Answer:	The glait's meeting the pleep for a drink.

(15) *has*-clitic context

Introduction:	This is a glait. # And this is a pleep.
Context:	The glait's written a love letter to the pleep.
Question:	What's happened?
Answer:	The glait's written a love letter to the pleep.

3.4 Labels and measurements

As a first step, all recordings were manually transcribed on the utterance level. Using the freely available WebMAUS Basic system (Kisler et al. 2017; Schiel 1999), a phonetic transcription and segmentation based on the manual transcription was created. This automated segmentation was then manually checked by six trained annotators using the software Praat (Boersma and Weenink 2020). Boundaries marking the beginning of an item or S were moved to the nearest zero crossing where both spectrogram and waveform indicated the initiation of the gesture for the respective segment, following laid out segmentation criteria based on features of specific sounds as described in the phonetic literature (e.g. Ladefoged 2003). In the case of S, the boundaries were set to the zero crossing closest to the onset and offset of the friction visible in the waveform (see Figure 2). If a pause followed the S, the boundary was set to the point where the friction of the S dropped to silence.

The reliability of the segmentation criteria was verified by trial segmentations, in which it was ensured that all annotators placed boundaries with only very small variations. Each annotator worked on a disjoint set of items; segmentation criteria were regularly re-verified in meetings of the annotators. After the segmentation process, a Praat script was used to extract the item, its phonetic transcription and its duration, as well as the S duration itself. If applicable, the duration of the

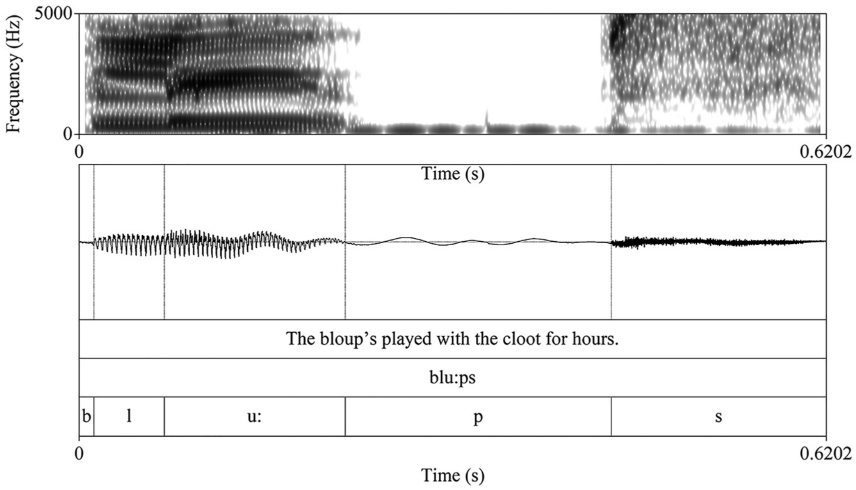


Figure 2: Example acoustic analysis for the item *bloups*.

following pause was also extracted. Additionally, the preceding and the following word were extracted as well.

3.5 Pre-processing

A part of the 1,920 (40 participants * 48 utterances) recorded data points had to be excluded from analysis for one or more of the following reasons. If an utterance did not include a word-final S, this utterance was discarded ($n = 599$). A high number of failures to produce final S was expected especially with the clitics since participants could use a different tense form, or the full form of the auxiliary. It was also expected that participants would produce wrong pronunciations (including those with the final S) of the newly encountered written word-forms, as the participants had to retrieve them from short-term memory after the fading out of the context. Additionally, utterances containing stutter or hesitation ($n = 29$), or replacement of pseudowords by pronouns ($n = 15$) were excluded as well. Some utterances were ungrammatical ($n = 9$), while other utterances contained pseudowords that were not part of the original set of pseudowords ($n = 8$). Cases where the interpretation of the final S was ambiguous presented another problem ($n = 114$). An example of such a case is given in (16) where a *has*-clitic was expected. Note that two pseudowords without a non-morphemic word-final S were introduced, while either a non-morphemic S or *has*-clitic S was produced for the item under investigation, and most likely a non-morphemic word-final S for the second pseudoword. As for regular inflected verbs there was no way to decide which type of S had been produced in such cases, such utterances were discarded.

(16) Introduction: This is a glait. # And this is a pleep.
Context: The glait's attended concerts with the pleep many times.
Question: What's happened many times?
Answer: The glaits attended many concerts with the pleeps many times.

After exclusions, 1,146 data points (approx. 60%) remained in the final data set. The final data set as well as the analysis and results discussed in the following sections can be found at https://osf.io/j4wxc/.

4 Analysis

4.1 Covariates

The set of covariates chosen for the present study is similar to that of other studies on phonetic effects of morphological structure (e.g. Hanique et al. 2013; Plag et al. 2017; Pluymaekers et al. 2005b, 2010). In the following, we first describe covariates used as fixed effects before we turn to variables used as random effects.

BASEDURLOG. Indicating a more local speaking rate (e.g. Plag et al. 2017), base duration was measured as well. Base duration in this case is equal to the summed duration of all word-internal segments preceding the S under investigation. That is, the stem of multi-morphemic items and the segmental string without the final S of mono-morphemic items is henceforth considered the base. We log-transformed and centred the base duration and called this variable BASEDURLOG.

BIPHONEPROB. For the reasons outlined in Section 3.2 we included the probability of the final biphones /fs/ (0), /ks/ (0.00427), /ps/ (0.00058) and /ts/ (0.00072) in mono-morphemic words as a covariate. BIPHONEPROB was computed on the basis of the transcriptions of all mono-morphemic words in CELEX (Baayen et al. 1995).

BIPHONEPROBSUM & BIPHONEPROBSUMBIN. A potential factor influencing the duration of a word in running speech is its predictability in context. The more predictable, the shorter the duration (e.g. Bell et al. 2009; Pluymaekers et al. 2005a; Torreira and Ernestus 2009). Such a word bigram frequency, however, is not applicable to pseudowords for obvious reasons. Instead, the summed biphone probability was used analogously as a comparable measure. The summed biphone probability for each pseudoword and its phonological variants was calculated by the Phonotactic Probability Calculator (Vitevitch and Luce 2004). Additionally, a binary covariate based on the summed biphone probability was created. The threshold for low versus high summed biphone probability for biphoneProbSumBin was the mean of the continuous covariate. That is, all values below the mean were considered to be low, while all values above the mean were taken as high.

folSeg & folType. To account for potential effects of the following word on the duration of S (e.g. Klatt 1976; Umeda 1977), these were included in regard to their onset segment adjacent to the word-final S. This was included in its phonological representation in folSeg (i.e. k for onset of *cooked*) as well as in its segmental class by folType (i.e. approximant APP for *listen*, fricative F for *find*, nasal N for *know*, plosive P for *cook*, vowel V for *eat*).

gender / location / monoMultilingual. Participants' gender and whether they had grown up in London or elsewhere in South Britain (location) were included as well as they may influence phonetic realisations. Additionally, participants who were early bilinguals (i.e. the L2 was acquired as a pre-school child) were categorized as multilingual, while all other participants were categorized as monolingual in monoMultilingual.[2]

neighbourhoodDensity & neighbourhoodFrequency. Neighbourhood densities and frequencies were included as covariates as the number of neighbours may influence phonetic reduction (e.g. Gahl et al. 2012). Both neighbourhood measures were taken from the CLEARPOND database (Marian et al. 2012). That is, neighbourhood-Density describes the number of words differing in one segment from the item in question (Marian et al. 2012: 3), while neighbourhoodFrequency describes the mean frequency (per million) of these neighbouring words.

pauseDur & pauseBin. In order to account for final-lengthening effects, all stretches of silence between the offset of the word-final S and the onset of the following word were measured. Silence of 50 ms and above was considered as pause (Lee and Oh 1999; see also Zvonik and Cummins 2003, and Krivokapić 2007, on short pause durations in-between short phrases). The closure durations of following plosives were taken into account by subtracting the mean closure duration of the pertinent plosive (mean values for /p, t, k/ adopted from Yao 2007)

2 Psycholinguistic experiments are standardly done with monolingual speakers (mostly of English, and mostly in the U.S.). In the multicultural context of a large European city like London, experiments with student populations necessarily involve speakers that are multilingual (with varying degrees of competence). To control for this potential confound, we added the variable monoMultilingual. While there are studies of phonetic duration in bilingual speech (e.g. Lee and Iverson 2012; Mack 1982) the effect of mono-/multilingualism on the duration on word- final S has not been explored yet.

from the measured stretch of silence. It was considered a pause only if the resulting duration was above the aforementioned threshold. Pause measurements were included as the continuous variable PAUSE as well as the binary variable PAUSEBIN (with the levels pause and no_pause).

PREC. It has been shown that the consonant preceding word-final S may influence the duration of word-final /s/ (e.g. Umeda 1977: 853). In particular, Umeda (1977: 853) finds that /s/ becomes shorter after plosives, and longer after the fricative /θ/ (and this presumably also holds for /s/ after the fricative /f/). We therefore included the consonant preceding the final /s/ as a covariate, PREC.

SPEAKINGRATE. As speaking rate is a self-evident variable affecting segment durations, this was controlled for. Speaking rate was computed as the number of syllables in an utterance divided by the duration of the utterance. For the statistical analysis, SPEAKINGRATE was centred (Afshartous and Preston 2011; Robinson and Schumacker 2009; Winter 2019). The computation was done automatically in Praat (de Jong and Wempe 2008). This way of computing speaking rate is similar to those utilized in previous studies (e.g. Plag et al. 2017).

ITEM & TRANSCRIPTION. Pseudowords were sometimes produced with varying segmental make-up. We therefore included both the orthographic representation of the pseudoword, and a phonological transcription of the word as spoken as two variables. These covariates were labelled ITEM and TRANSCRIPTION.

LIST & SLIDENUMBER. To account for possible durational differences due to priming and similar effects, the list number (1–12) and the point of occurrence during the experiment of the individual item were also included.

SPEAKER / AGE. SPEAKER ID was included to account for inter-speaker differences in production. AGE was included as well as they may show an influence on phonetic realisations.

4.2 Collinearity

One issue to address when fitting a model to a multitude of similar covariates is collinearity (e.g. Tomaschek et al. 2018). To avoid such issues, covariates were tested for correlation using the languageR package (Baayen and Shafaei-Bajestan 2019).

Correlations were checked between ITEM and TRANSCRIPTION (*rho* = 0.82, $p < 0.001$, Spearman), PAUSEDUR and PAUSEBIN (*rho* = 0.87, $p < 0.001$, Spearman), NEIGHBOURHOODDENSITY and NEIGHBOURHOODFREQUENCY (*rho* = 0.86, $p < 0.001$, Spearman), BIPHONEPROBSUM and BIPHONEPROBSUMBIN (*rho* = 0.87, $p < 0.001$, Spearman), SPEAKINGRATE and baseDur (rho = −0.33, $p = 0$, Pearson), and for folSeg and folType ($rho = -0.74$, $p < 0.001$, Spearman).

Given that all of the pairwise correlations except speakingRate and BaseDur were significant, the following procedure was adopted to avoid collinearity. For each pair of variables with a correlation of $rho > 0.5$, two linear mixed effects models, each containing only one of two variables, were created, and compared with a log-likelihood test. Each of these models contained the log-transformed S duration as dependent variable, one of the highly correlated variables as fixed effect, and speaker as random intercept. This allowed us to decide which of the covariates under discussion was a stronger predictor for our dependent variable. This covariate was then kept while the other one was no longer used. The same procedure was adopted to select between biphoneProb and preC. These procedures led to the exclusion of item (in favour of transcription), pauseDur (in favour of pauseDurBin), neighbourhoodFrequency (in favour of neighbourhoodDensity), biphoneProbSum (in favour of biphoneProbSumBin), folSeg (in favour of folType), and biphoneProb (in favour of preC).

4.3 Statistical analysis

Differences in consonant duration may play out as differences in absolute duration or as differences in relative duration (e.g. with gemination: Ben Hedia 2019; Oh and Redford 2012; Ridouane and Hallé 2017). Some previous analyses of the duration of S (Plag et al. 2017) have therefore looked at both absolute and relative duration, and the present paper will also present these two types of analyses. In the first analysis (Section 5.1) we used absolute duration of S as the dependent variable, whereas in the second analysis (Section 5.2), the duration of S relative to the duration of the whole word is used as the dependent variable. Relative duration (i.e. the variable proportionOfS) was calculated by dividing the absolute duration of the S by the duration of the whole word.

In order to analyse our data, models were fitted using linear mixed-effects regression in R (R Core Team 2019) using RStudio (RStudio Team 2018) and as implemented by lme4 (Bates et al. 2015), lmerTest (Kuznetsova et al. 2017), and LMERConvenienceFunctions (Tremblay and Ransijin 2015).

The dependent variable, duration of S, was log-transformed and centred following standard procedures to reduce the potentially harmful effect of skewed distributions in linear regression models (Winter 2019). The name of this variable is sDurLog. proportionOfS did not have a skewed distribution and no transformation was necessary.

Following the standard backward stepwise selection process (e.g. Baayen 2008), the first models containing the explanatory variable typeOfS (with levels nm = non-morphemic; pl = plural; is = *is*-clitic; has = *has*-clitic) alongside all covariates provided in Section 4.1. (with the exception of those excluded in 4.2) were included, plus two-way interactions of all covariates with the explanatory variable typeOfS. Random intercepts were included for transcription, list, slideNumber, speaker, and age. Following the 'keep it maximal' policy of Barr et al. (2013), we initially also included a random slope for typeOfS by speaker.

This full model was then continuously reduced through step-wise exclusion of non-significant factors using the 'step' function in R introduced by the lmerTest package (Kuznetsova et al. 2017). This function starts with the backward elimination of random-effect terms, followed by the backward elimination of fixed-effect terms.

At the last stage of the model fitting process, the final model needed trimming of the residuals (e.g. Baayen and Milin 2010). We removed data points with residuals larger than 2.5 standard deviations to ensure a satisfactory residual distribution. This resulted in a loss of nine data points (0.8%) and led to a satisfactory distribution of the residuals.

4.4 Overview of the data

An overview of all variables and their distribution is given in Tables 3 and 4.

Table 3: Summary of the dependent variable and numerical predictors in the final data set.

Dependent variable	**Mean**	**St. dev.**	**Min**	**Max**
sDurLog	0.002	0.388	−1.201	1.098
Numerical predictors	Mean	St. Dev.	Min	Max
speakingRate	−0.000	0.899	2.250	3.540
baseDurLog	0.072	0.194	0.000	3.559
pauseDur	0.072	0.193	0.000	3.559
neighbourhoodFrequency	27.345	84.645	0.000	412.027
biphoneProbSum	0.013	0.007	0.005	0.031
biphoneProb	0.001	0.002	0.000	0.004
age	28.740	9.743	19.000	58.000

Table 4: Summary of categorical predictors and the dependent variable in the final data set.

Categorical predictors	**Levels**			
ITEM	48			
TRANSCRIPTION	67			
NEIGHBOURHOODDENSITY	0: 419 1: 238 2: 165 3:107 4: 14 5: 114 6: 32 7: 30			
PAUSEBIN	no: 777	yes: 342		
BIPHONEPROBSUMBIN	low: 856	high: 263		
LIST	24			
SLIDENUMBER	48			
PREC	f: 273 k: 292 p: 281 t: 273			
FOLSEG	18			
FOLTYPE	APP: 299 F: 12 N: 230 P: 300 V: 278			
SPEAKER	40			
GENDER	2			
LOCATION	London: 636		elsewhere: 483	
MONOMULTILINGUAL	monolingual: 871		multilingual: 248	
Explanatory variable	Levels			
TYPEOFS	nm: 308	pl: 373	is: 284	has: 154

5 Results

5.1 Absolute duration

Figure 3 shows the distribution of the observed durations of non-morphemic, plural, *is*- and *has*-clitic S. On average, non-morphemic S duration is 134 ms, which is about 13 ms longer than plural S with a mean duration of 121 ms. The mean duration of the *is*-clitic is 103 ms and the mean duration of the *has*-clitic is 94 ms.

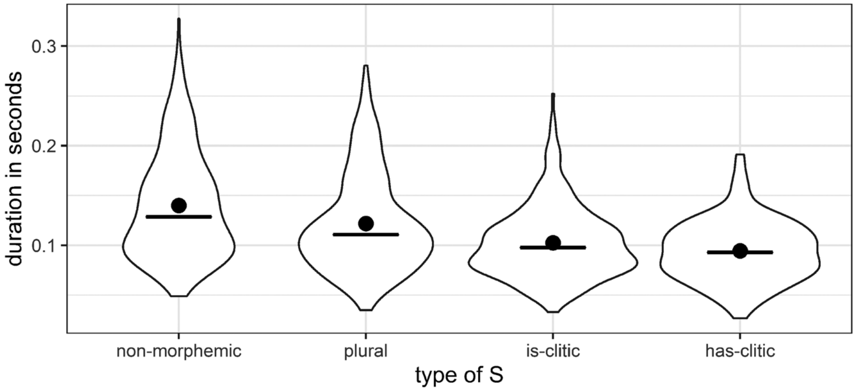


Figure 3: Observed durations of non-morphemic, plural, *is*- and *has*-clitic S. The dot represents the mean, the horizontal line indicates the median. The violin shapes represent rotated density plots describing the distribution of the data.

Multivariate analyses as described in the previous sections were then conducted to control for the many potentially intervening influences of the described covariates mentioned in Section 4.1. In our final model, fitted according to the procedure described above, we found main effects of type of S (typeOfS), speaking rate (speakingRate), base duration (baseDurLog), pause (pauseBin), biphone probability sum (biphoneProbSumBin), preceding consonant (preC), following segmental type (folType), and mono-/multilingualism (monoMultilingual). None of the interactions were significant.

Regarding the random effects, only speaker-specific random intercepts turned out to significantly improve the model fit. The *p*-values for the analysis of variance of the final model are given in Table 5.

Table 5: *p*-Values of fixed effects in the final model, fitted to the log-transformed durations of S.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (>F)
typeOfS	5.312	1.771	3	1,089.66	33.338	0.000
speakingRate	0.230	0.230	1	1,117.09	4.324	0.038
baseDurLog	9.466	9.466	1	1,079.58	178.220	0.000
pauseBin	6.970	6.970	1	1,110.28	131.235	0.000
biphoneProbSumBin	0.398	0.398	1	1,082.26	7.492	0.006
preC	0.623	0.208	3	1,080.29	3.910	0.009
folType	2.677	0.669	4	1,081.55	12.598	0.000
monoMultilingual	0.345	0.345	1	37.37	6.498	0.015

The marginal *R*-squared value of the model is 0.46, that is, fixed effects explain 46 percent of the variation in our data. The variance explained by the entire model is 61 percent as obtained by the conditional *R*-squared value of 0.61 (for marginal and conditional *R*-squared value computation see Nakagawa et al. 2017; values were computed with the MuMIn package, Barton 2019).

The estimates of the final model and their *p*-values are given in Table 6. The reference levels for the categorical predictors are: for typeOfS it is non-morphemic S, for pauseBin it is no-pause, for biphoneProbSumBin it is low, for preC it is t, for folType it is approximant, and for monoMultilingual it is monolingual. All coefficients can be interpreted as changes relative to these reference levels.

Table 6: Fixed-effect coefficients and *p*-values as computed by the final model (mixed-effects model fitted to the log-transformed and centred durations of S).

	Estimate	Std. error	df	*t*-value	Pr (>|t|)
(Intercept)	−1.321	0.068	550.378	−19.498	0.000
typeOfSpl	−0.114	0.019	1,094.00	−6.062	0.000
typeOfSis	−0.178	0.020	1,096.00	−8.839	0.000
typeOfShas	−0.196	0.024	1,091.00	−8.14	0.000
speakingRate	−0.021	0.010	1,117.00	−2.079	0.038
baseDurLog	0.586	0.044	1,080.00	13.35	0.000
pauseBinpause	0.206	0.018	1,110.00	11.456	0.000
biphoneProbSumBinhigh	0.047	0.017	1,082.00	2.737	0.006
preCf	0.061	0.020	1,081.00	−3.044	0.003
preCk	0.055	0.020	1,082.00	−0.303	0.006
preCp	0.050	0.020	1,079.00	2.522	0.012
folTypeF	0.012	0.070	1,084.00	0.171	0.864
folTypeN	−0.036	0.021	1,079.00	−1.764	0.078
folTypeP	−0.045	0.019	1,080.00	−2.384	0.017
folTypeV	−0.136	0.020	1,082.00	−6.85	0.000
monoMultilingualmultilingual	−0.152	0.059	37.37	−2.549	0.015

Effect size of individual predictors was checked by fitting models that lacked a particular predictor, and comparing their marginal *R*-squared values to those of the final model. The results are reflected in the hierarchy given in (17). The decrease in *R*-squared is greatest when removing baseDurLog, followed by pauseBin, and so forth. Overall, the morphological status of an S appears to be a strong predictor of its acoustic duration.

(17) baseDurLog >> pauseBin >> typeOfS >> monoMultilingual >>
folType >> speakingRate >> biphoneProbSumBin >> PreC

Figure 4 shows the effect of the numerical variables included in the final model on S duration. The estimated values of the dependent variable and the base duration are back-transformed into seconds. Speaking rate and base duration show effects in the expected direction. With faster speech, S becomes shorter (panel A), while longer base durations also come with longer S durations (panel B).

The partial effects of the categorical variables included in the final model are illustrated in Figure 5. S duration is longer if the S is followed by a pause (panel A), which can be interpreted as a clear case of phrase-final lengthening (e.g. Cooper and Danly 1981). Higher biphone probability sum leads to longer S durations (panel B). There is also an effect of the preceding consonant: the plosive /t/ is followed by significantly shorter S durations than are /k/ and /f/ (panel C). S

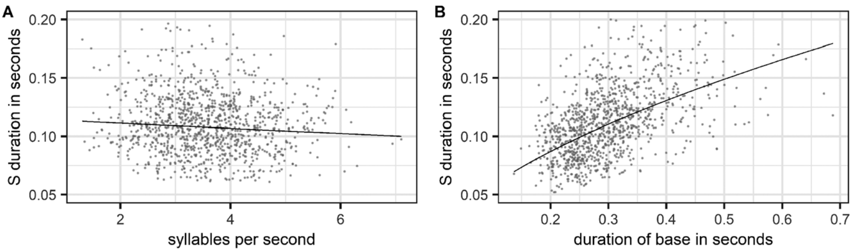


Figure 4: Partial effects of the numerical variables included in the final model, fitted to the log-transformed values of duration of S.

duration is significantly shorter when followed by a vowel, while all other differences between following consonants are minor in nature (panel D). Lastly, monolingual speakers produce longer S durations than multilingual speakers (panel E).

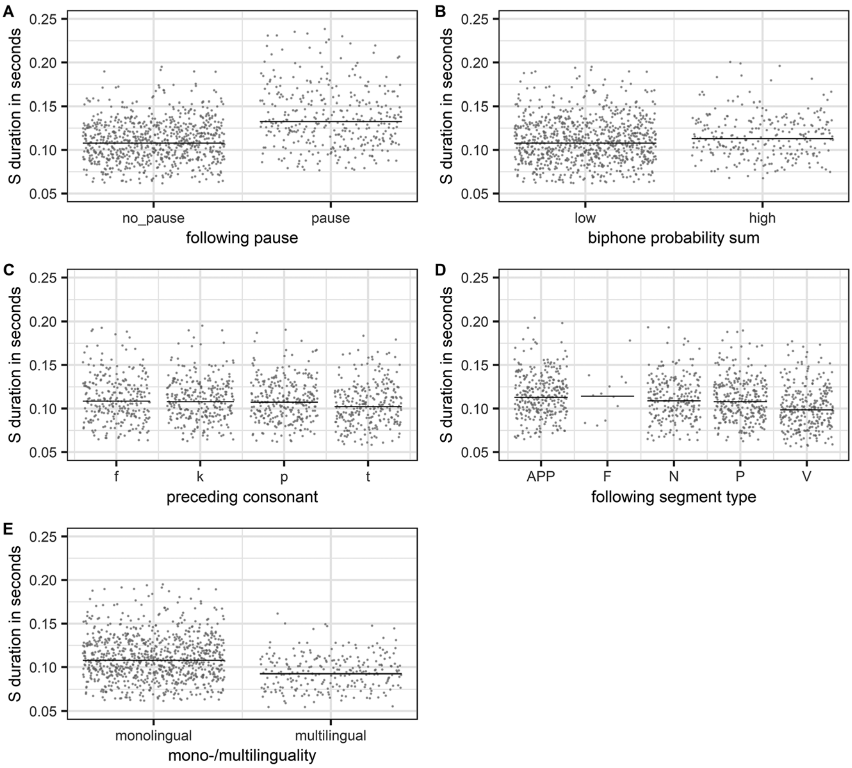


Figure 5: Partial effects of the categorical variables included in the final model, fitted to the log-transformed values of duration of S.

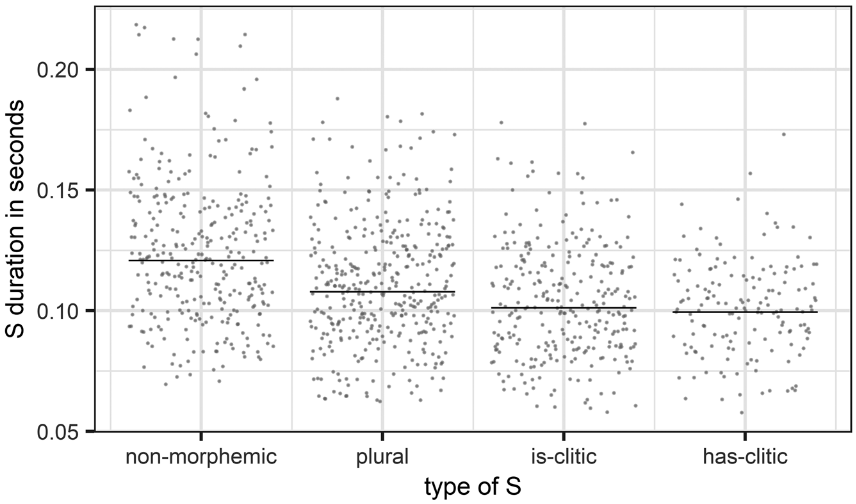


Figure 6: Partial effect of TYPEOFS in the final model, fitted to the log-transformed values of duration of S.

Table 7: Multiple comparisons of means of duration of S (Tukey contrasts). Significant codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

			Estimate	Std. Error	*z*-value	Pr (>|z|)	
Plural	–	non-morphemic	−0.114	0.019	−6.062	<0.001	***
is-clitic	–	non-morphemic	−0.188	0.020	−8.839	<0.001	***
has-clitic	–	non-morphemic	−0.196	0.024	−8.140	<0.001	***
is-clitic	–	plural	−0.064	0.019	−3.294	0.005	**
has-clitic	–	plural	−0.082	0.023	−3.503	0.003	**
has-clitic	–	*is*-clitic	−0.018	0.023	−0.766	0.868	

The effect of the variable of interest, i.e. TYPEOFS, is plotted in Figure 6. As above, the values of the dependent variable are back-transformed into seconds.

We can see that there are durational differences between the different types of S. The results of pair-wise comparisons of the predicted means using Tukey contrasts (as implemented by the multcomp package for R, Hothorn et al. 2008) are summarized in Table 7.

Based on the Tukey tests, the comparison of the different types of S yields the significant contrasts shown in Table 8. If we look at the different durations given in Table 9, the following hierarchy emerges: non-morphemic > plural > *is*-/*has*-clitic.

To summarize, the durational differences between non-morphemic and all other types of S, as well as the durational difference between plural and the clitics

Table 8: Significant contrasts in duration between different types of S. Significant codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

	nm	pl	*is*	*has*
non-morphemic	n.a.	***	***	***
Plural		n.a.	**	**
is-clitic			n.a.	
has-clitic				n.a.

Table 9: S durations as estimated by the final model using non-centred data. All values are back-transformed to seconds. Values given are estimated for items without following pause, high biphone sum probability, monolingual speakers, and across all preceding and following segment types.

typeOfS	Mean
non-morphemic	0.224
Plural	0.200
is-clitic	0.187
has-clitic	0.184

are significant, while there is no significant durational difference between the two clitics. Non-morphemic S is longest in duration, followed by plural S, which in turn is followed by clitic S.

5.2 Relative duration

The results for relative duration are very similar to those of absolute duration. The *p*-values for the analysis of variance of the final model are given in Table 10. Table 11 shows the coefficients for the final model. All effects go in the same direction as in the analysis of absolute duration. The only predictors that have lost significance when compared to the model for absolute duration are preC and speakingRate.

Table 10: *p*-values of fixed effects in the final model, fitted to the relative durations of S.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (>F)
typeOfS	0.161	0.054	3	1,070.68	25.510	0.000
pauseBin	0.186	0.186	1	1,101.26	88.518	0.000
biphoneProbSumBin	0.015	0.015	1	36.32	6.917	0.012
folType	0.071	0.018	4	1,063.31	8.389	0.000
monoMultilingual	0.010	0.010	1	37.81	4.561	0.039

Table 11: Fixed-effect coefficients and p-values as computed by the final model (mixed-effects model fitted to the relative durations of S).

	Estimate	Std. Error	df	t-value	Pr (>|t|)
(Intercept)	0.299	0.007	89.73	45.827	0.000
typeOfSpl	−0.019	0.004	1,085.00	−5.157	0.000
typeOfSis	−0.031	0.004	1,070.00	−7.651	0.000
typeOfShas	−0.035	0.005	1,067.00	−7.260	0.000
pauseBinpause	0.033	0.004	1,101.00	9.408	0.000
biphoneProbSumBinhigh	0.013	0.005	36.32	2.630	0.012
folTypeF	0.001	0.014	1,068.00	0.086	0.931
folTypeN	−0.006	0.004	1,061.00	−1.409	0.159
folTypeP	−0.007	0.004	1,056.00	−1.708	0.088
folTypeV	−0.022	0.004	1,063.00	−5.568	0.000
monoMultilingualmultilingual	−0.024	0.011	37.81	−2.136	0.039

The differences in the means show the same pattern as in the analysis of absolute duration, as can be seen in Table 12.

Table 12: Multiple comparisons of means of relative duration of S (Tukey contrasts). Significant codes: '***' $p < 0.001$, '**' $p < 0.01$, '*' $p < 0.05$.

			Estimate	Std. error	z-value	Pr (>|z|)	
Plural	–	non-morphemic	−0.019	0.004	−5.157	<0.001	***
is-clitic	–	non-morphemic	−0.031	0.004	−7.651	<0.001	***
has-clitic	–	non-morphemic	−0.035	0.005	−7.260	<0.001	***
is-clitic	–	plural	−0.011	0.004	−2.936	0.017	*
has-clitic	–	plural	−0.015	0.005	−3.300	0.005	**
has-clitic	–	*is*-clitic	−0.004	0.005	−0.854	0.827	

The analysis of relative duration thus is fully consistent with the results for absolute duration.

6 Discussion

Following in the footsteps of previous studies on durational differences between different types of S, we tested whether the morphological category of word-final S has an influence on its acoustic duration in speech production. In order to avoid imbalanced data as in the case of corpus studies, we used a production experiment, i.e. speech material elicited by the means of highly controlled contexts of a production task. For the first time in this context, pseudowords instead of real words were used to minimize potentially confounding lexical effects. We found that there are significant durational differences between non-morphemic and morphemic types of word-final S, with morphemic types of S being significantly shorter in duration than non-morphemic S. Also, there are significant durational differences between the plural suffix and the *is*- and *has*-clitic S, with plural S being significantly longer than clitic S and with no significant difference between the two clitics. Hence, type of S emerged as a strong, significant predictor of segmental duration.

The differences between different types of S in the present study are completely in line with previous studies that were based on speech corpora, and on different varieties of English (Zimmermann 2016 on New Zealand English; Plag et al. 2017; Tomaschek et al. 2019 on North American English; this study on British English). In those studies the same pattern of differences was found. Turning to previous experimental studies, we find differing results. The results of both prior experimental studies (Seyfarth et al. 2017; Walsh and Parker 1983) are subject to potentially confounding effects of the lexical and contextual properties of the items under investigation. Their finding of non-morphemic S being shorter than morphemic S may well be an artefact of such properties. The items used in the present study, however, are much less prone to be subject to such effects as they are pseudowords with no established representations in the speakers' mental lexicons. We cannot compare our results on the duration of clitic S to previously reported ones by other experimental studies, as none of the previously conducted experimental studies investigated clitic S production.

No previous studies have used pseudowords either, so before turning to the theoretical interpretation of the results of the present study, a few words are in order on whether using pseudowords might have had an undesired impact on our results. While the use of pseudowords in phonetic experiments comes with a number of benefits (see Section 3.2), it also raises some questions. First, there is the issue of phonotactic probability raised in Section 3.2. Two measures concerned with phonotactics (one describing the phonotactic probability of the whole word, the other taking into consideration the consonant preceding the word-final S) were included in our statistical analysis to address this issue. It turned out that phonotactic probability influences the productions of our pseudowords, as it does for real words. Crucially, there was no interaction between the type of S and the consonant preceding it in mono-morphemic words. This means that speakers produced these clusters in the same way, no matter whether the cluster occurred in the mono-morphemic words, or whether the cluster straddled the morphemic boundary between the stem and the S. The main effects of the phonotactic variables turned out to be rather weak, and, crucially, were properly controlled for in the regression analysis. In sum, the phonotactics of the final cluster does not seem to have unduly influenced the results.

Second, there might have been a problem with another aspect of the phonological structure of the pseudowords in the experiment, i.e. long-distance agreement of phonological features (Coetzee 2005, 2008). Such effects of the Obligatory Contour Principle (OCP: Coetzee 2005) might have arisen with pseudowords such as *pleep* (in which initial /p/ and final /p/ share all features) or *glik* (in which the initial and final sounds share the dorsal feature). Following the findings by Coetzee (2008), we coded a new variable to test this effect post-hoc as an additional covariate and as an interacting term of TYPEOFS with the following levels: not well-formed for pseudowords in which the initial and final consonant share all features (n = 836), moderately well-formed for pseudowords in which the initial and final consonant share the dorsal feature (n = 147), and well-formed for all remaining pseudowords (n = 145). There was no significant main effect of this variable on the duration of S, nor a significant interaction with TYPEOFS. OCP effects thus cannot explain our results.

Third, after having carried out the experiments, it came to our attention that some of our pseudowords have real word relatives that are spelled differently but are phonologically identical. That is, *pleet(s)* corresponds to *pleat(s)*, *glits* corresponds to *glitz* (and no word corresponding to *glit*), and *glik* corresponds to the surname *Glick* (and no surname corresponding to *gliks*), whereas *glif(s)* corresponds to *glyph(s)*, which has a very low frequency and thus may constitute a nonce word for most of our participants. These words might have unduly influenced our results and should perhaps not have been included into the statistical analysis. To check whether these items had any influence on the results, we created a data set containing all data but the four potentially offending items. Fitting the final model (as done in Section 4.3) to this new dataset resulted in basically the same findings, i.e. TYPEOFS was still a significant predictor for S duration showing the same significant differences between non-morphemic, plural, and clitic items as presented in Table 8.

It has recently been shown that the notion of pseudoword is problematic in a more general way. The notion of pseudoword itself is usually based on the idea of the lexicon as a community construct. When talking about the mental lexicon, however, it is clear that what is an existing word and what is an unknown pseudoword is a matter of the individual speaker's mental lexicon. All participants in our experiment denied knowing any of the pseudowords used in this experiment when asked afterwards. At the community level, Google frequencies of pseudowords have been shown to be a robust predictor of reaction times in lexical decision tasks (e.g. Hendrix and Sun 2020). To test whether Google frequency had an effect on our results, the covariate GOOGLEFREQ was created containing the number of

Google search hits for each pseudoword. The addition of this covariate as either fixed effect or interacting term to typeOfS resulted in its exclusion during the model simplification procedure.

Finally, we can turn to the theoretical implications of our results. What do these results mean for the three hypotheses that we tested? The Feed-forward Hypothesis states that there is no durational difference between word-final non-morphemic S, plural S and auxiliary clitic S. This hypothesis is rejected as we have provided carefully controlled evidence that shows that the duration of S varies by morphological category. This is an effect that present feed-forward models cannot accommodate, unless they would be refined in such a way that post-lexical processes can arise from certain kinds of lexical information. At present, no such refinement is available.

The Prosodic Hypothesis states that there are durational differences between different types of word-final S, with non-morphemic S being shorter than plural S, and plural S being shorter than the auxiliary clitic. While it is true that there are durational differences between the categories, the differences we observed pattern in the opposite direction. We found that the more integrated the S is with the stem, the longer its duration. The Prosodic Hypothesis is correct in positing that the two auxiliary clitics should show no difference in duration. Overall, however, the Prosodic Hypothesis must be rejected, as the prosodic structure does not explain the most important patterning of the data.

Finally, the Emergence Hypothesis states that there are durational differences between the different types of word-final S under investigation. The fact that we find such differences means that these differences might emerge through the mechanisms posited by the theories underlying this hypothesis.

As mentioned above, Tomaschek et al. (2019) found that stronger support for a morphological function leads to a longer duration, i.e. as for our findings, non-morphemic S showed the longest duration, auxiliary clitic S showed the shortest durations, and plural suffix S duration was in-between. This effect seems to run counter to the predictions of information-theoretic accounts and probabilistic theories, according to which words and segments are realised shorter when they are less informative (Aylett and Turk 2004; Cohen Priva 2015; Jaeger 2010). However, the enhancement effects are in line with studies showing that duration increases with increasing paradigmatic certainty (Bell et al. 2020; Cohen 2014; Kuperman et al. 2007; Tucker et al. 2019). For instance, Kuperman and colleagues found that the duration of a given interfix in Dutch compounds increases with increasing probability of this interfix (as against its competitors) in the left constituent family of the compound.

How can these two seemingly opposite frequency effects be reconciled? This question is addressed in a study by Schmitz et al. (2021), in which the authors implemented a linear discriminative model (Baayen et al. 2019; Chuang et al. 2020) and used the measurements derived from the discriminative network to predict the duration of word-final S, using the data on non-morphemic and plural S from the present study. It turns out that the two opposite effects reside in different processing domains. According to Schmitz et al.'s results, the enhancement effect arises from the semantic activation of related words, with more diverse activation going together with shorter durations (see also Stein and Plag 2021; Tomaschek et al. 2019). In contrast, the syntagmatic morphology-related reduction effect arises at the phonotactic and articulatory level, where more certainty (i.e. more support for the articulatory transitions) goes together with shorter articulations.

Overall, it seems that simplistic approaches can neither explain the existence, nor the patterning of the durational differences we find attested. The Feed-forward Hypothesis is rejected because durational differences were in fact observed. The Prosodic Hypothesis is rejected because the observed durational differences pattern in a direction that is opposite to the one predicted. The Emergence Hypothesis is supported by our findings as it proposes that durational differences of some nature should emerge between different types of S.

The complexities of speech production are enormous, and none of the existing approaches has satisfactory answers to the many questions this complexity raises. Even the empirically most adequate approach, discriminative learning, includes a black box. While there are correlations between association weights and acoustic durations, it is unclear how effects of phonological certainty and semantic activation translate into articulatory gestures that result in durational differences. We still find this approach currently most promising, as all other applicable approaches fail to account for findings such as those presented in this paper.

The results of the present study may bring up further questions. First, assuming the durational differences found here and in previous studies are indeed systematic, one would also like to know whether language users are able to perceive them. This automatically leads to questions of whether all differences are perceptible or only some of them given our knowledge on the perception of differences in fricative durations, i.e. that the threshold for perceptible durational differences appears to be at 25 ms (e.g. Klatt and Cooper 1975). Secondly, if the durational differences are perceptible, another question naturally suggests itself: do users of a language not only perceive but also make use of such differences, e.g. to aid comprehension by predicting potential upcoming words? These questions call for highly controlled perception and comprehension studies.

Let us conclude. This paper is the first to use pseudowords to investigate durational differences in productions of different types of word-final S in English. In accordance with previous results from speech corpus studies, we found that non-morphemic S is longer than plural S, which in turn is longer than auxiliary clitic S. By using pseudowords, and by using carefully controlled stimuli, we demonstrated that durational differences between different types of S are of a robust nature rather than a by-product of confounding factors. This means that similar previous results probably did not arise from confounding effects of lexical properties or unbalanced corpus-based data sets. We conclude that differences in S durations are due to the processing of the morphological information encoded in the pertinent type of S. In other words, morphological information may influence speech production in such a way that systematic subphonemic differences arise. This calls for revisions in current models of the relationship between morphology, phonology, and phonetic realisation.

Acknowledgements: The authors are grateful to the members of the DFG Research Unit FOR2373 and the audience of several conferences (19th International Morphology Meeting, February 2020, Vienna; LabPhon 17, July 2020, Vancouver; UK Cognitive Linguistics Conference 2020, July 2020, Birmingham; 16. Phonetik und Phonologie Tagung, September 2020, Trier; Interfaces of Phonetics, May 2021, Oldenburg) for valuable input, to James White and Andrew Nevins allowing the production experiment to take place at Chandler House, University College London, and to Andrew Clark for his technical support. The usual disclaimers apply.
Research funding: This research was funded by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 'Spoken Morphology', grant PL 151/7-2 'Central project', and grants BA 6523/1-1 and PL 151/9-1 'Final S in English: The role of acoustic detail in morphological learning'), which we gratefully acknowledge.
Author contribution: Dominic Schmitz, Ingo Plag, and Dinah Baer-Henney conceived of the presented idea and planned the experiment. Dominic Schmitz carried out the experiment and, with Ingo Plag, performed the statistical analysis with input from Dinah Baer-Henney. Dominic Schmitz wrote the manuscript; it was proofread by all authors. All authors provided critical feedback and helped shape the research, analysis, and manuscript.
Statement of ethics: The research performed in this paper has ethic approval of the ethics committee of the Linguistic Society of Germany and of the University College London (LING-2018-8-01). All participants signed a written informed consent form before participating in the production study and were provided with detailed information sheets.
Conflict of interest statement: The authors have no conflicts of interest to declare.

Appendix A

Contexts and questions used in the production task sorted by onset segment of the verb following the word-final S, and the type of word-final S. The pseudowords cloot/cloots and glaik/glaiks are used as examples.

1. Approximant onset verbs

1a. write
non-morphemic
Context: The cloots writes a letter to the glaiks every month.
Question: What happens every month?
plural
Context: Last week, the cloots wrote a letter to their mother.
Question: What happened last week?
is-clitic
Context: The cloot's writing a letter to the glaik.
Question: What's happening?
has-clitic
Context: The cloot's written a love letter to the glaik.
Question: What's happened?
1b. listen
non-morphemic
Context: Every day, the cloots listens to the glaik's singing.
Question: What happens every day?
plural
Context: Last week, the cloots listened to each other's songs.
Question: What happened last week?
is-clitic
Context: The cloot's listening to the glaik sing.
Question: What's happening?
has-clitic
Context: The glaik's a famous singer. The cloot's listened to all of his songs.
Question: What's happened?
1c. watch
non-morphemic
Context: Every night, the cloots watches the glaiks' TV series.
Question: What happens every night?

plural
Context: Yesterday, the cloots watched TV together.
Question: What happened yesterday?
is-clitic
Context: The cloot's watching the glaik play football.
Question: What's happening?
has-clitic
Context: The glaik's a famous football player. The cloot's his biggest fan. He's watched all of the glaik's matches.
Question: What's happened?

2. Nasal onset verbs

2a. move
non-morphemic
Context: They're good friends and want to live close to each other. Therefore, the cloots moves into a new home.
Question: What happens?
plural
Context: Last year, the cloots moved into a new home.
Question: What happened last year?
is-clitic
Context: The cloot's moving in with the glaik.
Question: What's happening?
has-clitic
Context: The cloot's moved in with the glaik.
Question: What's happened?
2b. meet
non-morphemic
Context: Every Saturday, the cloots meets the glaiks for a drink.
Question: What happens every Saturday?
plural
Context: Last week, the cloots met for a drink.
Question: What happened last week?
is-clitic
Context: Tonight, the cloot's meeting the glaik for a drink.
Question: What's happening tonight?
has-clitic
Context: One year ago, the cloot's met the glaik for the first time.

Question: What's happened one year ago?

2c. knit

non-morphemic

Context: Every night, the cloots knits a blanket for the glaiks.

Question: What happens every night?

plural

Context: Last week, the cloots knitted a blanket together.

Question: What happened last week?

is-clitic

Context: The cloot's knitting a hat for the glaik's birthday.

Question: What's happening?

has-clitic

Context: The cloot's knitted 10 scarfs for the glaik last winter.

Question: What's happened last winter?

3. Plosive onset verbs

3a. play

non-morphemic

Context: Every day, the cloots plays with the glaiks.

Question: What happens every day?

plural

Context: Last week, the cloots played a game.

Question: What happened last week?

is-clitic

Context: The cloot's playing with the glaik.

Question: What's happening?

has-clitic

Context: The cloot's played with the glaik for hours.

Question: What's happened for hours?

3b. call

non-morphemic

Context: Every night, the cloots calls the glaiks for a nice chat.

Question: What happens every night?

plural

Context: Yesterday, the cloots called each other to talk about their day.

Question: What happened yesterday?

is-clitic

Context: The cloot's calling the glaik to talk about their evening plans.

Question: What's happening?
has-clitic
Context: The cloot's calling the glaik, but the glaik does not answer the phone. The cloot's called the glaik several times by now.
Question: What's happened several times now?

3c. cook
non-morphemic
Context: Every Sunday, the cloots cooks lunch for the glaiks.
Question: What happens every Sunday?
plural
Context: Every Friday, the cloots cook dinner together.
Question: What happens every Friday?
is-clitic
Context: The cloot's cooking dinner for the glaik.
Question: What's happening?
has-clitic
Context: The cloot's a great cook. The cloot's cooked lunch for the glaik for many years.
Question: What's happened for many years?

4. Vowel onset verbs

4a. ask
non-morphemic
Context: Every Friday, the cloots asks the glaiks about his weekend.
Question: What happens every Friday night?
plural
Context: Last Friday, the cloots asked each other about their weekend.
Question: What happened last Friday?
is-clitic
Context: The cloot's asking the glaik about his weekend.
Question: What's happening?
has-clitic
Context: They just met. The cloot's a curious thing. He's asked the glaik many questions in the past couple hours.
Question: What's happened in the past couple hours?

4b. eat
non-morphemic
Context: The cloots eats breakfast with the glaiks every day.

Question: What happens every day?
plural
Context: Two days ago, the cloots ate their lunch together.
Question: What happened two days ago?
is-clitic
Context: The cloot's eating cake with the glaik.
Question: What's happening?
has-clitic
Context: They are having lunch together. The cloot's really hungry. He's eaten the glaik's lunch as well.
Question: What's happened?

4c. attend
non-morphemic
Context: Tonight, the cloots attends the glaiks' party.
Question: What happens tonight?
plural
Context: Yesterday, the cloots attended a ball together.
Question: What happened yesterday?
is-clitic
Context: Tomorrow, the cloot's attending the glaik's party.
Question: What happens tomorrow?
has-clitic
Context: They're big music fans. The cloot's attended concerts with the glaik many times.
Question: What's happened many times?

Appendix B

Practice material used in the production task. The pseudowords lope/lopes and feap/feaps were used in the practice trials.

non-morphemic
Context: The feaps is on holiday, therefore the lopes misses him a lot.
Question: What's happening?
plural
Context: Two weeks ago, the feaps convinced their best friend to join their sports team.
Question: What happened two weeks ago?
is-clitic
Context: The lope's late. He's missing his appointment with the feap.

Question: What's happening?
has-clitic
Context: The feap's convinced the lope many times to play a game with him.
Question: What's happened in the past couple hours?

References

Afshartous, David & Richard A. Preston. 2011. Key results of interaction models with centering. *Journal of Statistics Education* 19. 1–24.

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A function explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47. 31–56.

Baayen, R. Harald. 2008. *Analysing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. Harald & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3. 12–28.

Baayen, R. Harald & Elnaz Shafaei-Bajestan. 2019. languageR [R package]. Version 1.5.0. https://CRAN.R-project.org/package=languageR (accessed August 2019).

Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P. Blevins. 2019. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019. 1–39.

Baayen, R. Harald, Petar Milin, Dusica Filipović Durdević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naïve discriminative learning. *Psychological Review* 118. 438–482.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. *The CELEX lexical database (CD-ROM). Linguistic data consortium*. Philadelphia, PA: University of Pennsylvania.

Barr, J. Dale, Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 3. 255–278.

Barton, Kamil. 2019. MuMIn: Multi-model inference [R package]. Version 1.43.6. https://CRAN.R-project.org/package=MuMIn (accessed August 2019).

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.

Bell, Melanie J., Sonia Ben Hedia & Ingo Plag. 2020. How morphological structure affects phonetic realization in English compound nouns. *Morphology* 31. 1–34.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60. 92–111.

Ben Hedia, Sonia. 2019. *Gemination and degemination in English affixation. Investigating the interplay between morphology, phonology and phonetics*, vol. 8. Studies in Laboratory Phonology. Berlin: Language Science Press.

Ben Hedia, Sonia & Ingo Plag. 2017. Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics* 62. 34–49.

Berko-Gleason, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.
Blevins, James P., Farrell Ackerman & Robert Malouf. 2016. Morphology as an adaptive discriminative system. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 271–301. Amsterdam & Philadelphia: John Benjamins.
Boersma, Paul & David Weenink. 2020. Praat:doing phonetics by computer [Computer program]. Version 6.0.49. http://www.praat.org/ (accessed March 2019).
Booij, Geert E. 1983. Principles and parameters in prosodic phonology. *Linguistics* 21. 249–280.
Brewer, Jordan. 2008. *Phonetic reflexes of orthographic characteristics in lexical representation*. The University of Arizona PhD Thesis.
Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
Caselli, Naomi K., Michael K. Caselli & Ariel M. Cohen-Goldberg. 2016. Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology* 69. 434–454.
Cho, Taehong. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica* 58. 129–162.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*, vol. 1. New York: Harper and Row.
Chuang, Yu-Ying, Marie Lena Vollmer, Elnaz Shafaei-Bajestan, Susanne Gahl, Peter Hendrix & R. Harald Baayen. 2020. The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods* 49. 945–976.
Clements, George N. & Samuel Jay Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.
Coetzee, Andries W. 2005. The obligatory contour principle in the perception of English. In Sónia Frota, Marina Vigario & Maria João Freitas (eds.), *Prosodies*, 223–245. New York: Mouton de Gruyter.
Coetzee, Andries W. 2008. Grammar is both categorical and gradient. In Steve Parker (ed.), *Phonological argumentation*, 9–42. Oakville, CT: Equinox Pub. Ltd.
Cohen, Clara. 2014. *Combining structure and usage patterns in morpheme production: Probabilistic effects of sentence context and inflectional paradigms*. Berkeley: University of California PhD Dissertation.
Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6. 243–278.
Cooper, William E. & Martha Danly. 1981. Segmental and temporal aspects of utterance-final lengthening. *Phonetica* 38. 106–115.
de Jong, Nivja & Ton Wempe. 2008. Praat script syllable nuclei [Praat script]. https://sites.google.com/site/speechrate/Home/praat-script-syllable-nuclei-v2 (accessed November 2019).
Drager, Katie K. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39. 694–707.
Engemann, Marie & Ingo Plag. 2021. Phonetic reduction and paradigm uniformity effects in spontaneous speech. *The Mental Lexicon* 16. 166–199.
Fort, Mathilde, Alexander Martin & Sharon Peperkamp. 2015. Consonants are more important than vowels in the Bouba-kiki effect. *Language and Speech* 5. 247–266.
Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84. 474–496.
Gahl, Susanne & Alan C. L. Yu. 2006. Special issue on exemplar-based models in linguistics. *The Linguistic Review* 23. 213–216.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66. 789–806.

Goad, Heather. 1998. Plurals in SLI: Prosodic deficit or morphological deficit? *Language Acquisition* 7. 247–284.

Goad, Heather. 2002. Markedness in right-edge syllabification: Parallels across populations. *Canadian Journal of Linguistics* 47. 151–186.

Goad, Heather & Lydia White. 2019. Prosodic effects on L2 grammars. *Linguistic Approaches to Bilingualism* 9. 769–808.

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

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105. 251–279.

Gontijo, Possidonia F. D., Isa Gontijo & Richard Shillcock. 2003. Grapheme-phoneme probabilities in British English. *Behavior Research Methods, Instruments, & Computers* 35. 136–157.

Gries, Stefan Th. 2015. The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10. 95–125.

Hanique, Iris, Mirjam Ernestus & Barbara Schuppler. 2013. Informal speech processes can be categorical in nature, even if they affect many different words. *Journal of the Acoustical Society of America* 133. 1644–1655.

Hendrix, Peter & Ching Chu Sun. 2020. A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47. 157–183.

Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50. 346–363.

Jaeger, Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61. 23–62.

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan Bybee, J Paul & Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of the lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in laboratory phonology*, 7, 3–34. Berlin, New York: De Gruyter Mouton.

Keating, Patricia A. 2006. Phonetic encoding of prosodic structure. In Jonathan Harrington & Marija Tabain (eds.), *Speech production: Models, phonetic processes, and techniques*. New York & East Sussex: Psychology Press.

Kemps, Rachèl J. J. K., Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005a. Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition* 33. 430–446.

Kemps, Rachèl J. J. K., Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005b. Prosodic cues for morphological complexity in Dutch and English. *Language & Cognitive Processes* 20. 43–73.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In In-Seok Yang (ed.), *Linguistics in the morning calm: Selected papers from SICOL*, 3–91. Seoul: Hanshin.

Kisler, Thomas, Use D. Reichel & Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326–347.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustic Society of America* 59. 1208–1221.

Klatt, Dennis H. & William E. Cooper. 1975. Perception of segment duration in sentence contexts. In Antonie Cohen & Sibout G. Nooteboom (eds.), *Structure and process in speech perception*, 69–89. Berlin: Springer.

Köhler, Wolfgang. 1929. *Gestalt psychology*. New York, NY: Liveright.

Krivokapić, Jelena. 2007. Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics* 35. 162–179.

Kuperman, Victor, Mark Pluymaekers, Mirjam Ernestues & R. Harald Baayen. 2007. Morphological predictability and acoustic salience of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121. 2261–2271.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82. 1–26.

Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell.

Lavoie, Lisa. 2002. Some influences on the realisation of for and four in American English. *Journal of the International Phonetic Association* 32. 175–202.

Lee, Sue Ann S. & Gregory K. Iverson. 2012. Stop consonant productions of Korean-English bilingual children. *Bilingualism: Language and Cognition* 15. 275–287.

Lee, Sangho & Yung-Hwan Oh. 1999. Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems. *Speech Communications* 28. 283–300.

Levelt, Willem J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–38.

Levelt, Willem J. M. & Linda R. Wheeldon. 1994. Do speakers have access to a mental syllabary? *Cognition* 50. 239–269.

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

Lohmann, Arne. 2018. Time and thyme are NOT homophonous: A closer look at gahl's work on the lemma frequency effect including a reanalysis. *Language* 94. e180–e190.

Mack, Molly. 1982. Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America* 71. 173–178.

Marian, Viorica, James Bartolotti, Sarah Chabal & Anthony Shook. 2012. CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7. e43230.

Nakagawa, Shinichi, Paul C. D. Johnson & Holger Schielzeth. 2017. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface* 14. 1–11.

Nespor, Marina & Irene Vogel. 2007. *Prosodic Phonology*. Berlin, Boston: De Gruyter Mouton.

Oh, Grace & Melissa A. Redford. 2012. The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40. 82–91.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee & Paul J. Hopper (eds.), *Typological studies in language*, vol. 45, *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam and Philadelphia: John Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in laboratory phonolog*, 7, 101–140. Berlin, New York: De Gruyter Mouton.

Pitt, Mark A., Leslie Dilley, Keith Johnson, Scott Kiesling, William D. Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. *Buckeye corpus of conversational speech*, 2nd release. Columbus, OH: Department of Psychology, Ohio State University.

Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53. 181–216.

Plag, Ingo, Arne Lohmann, Sonia Ben Hedia & Julia Zimmermann. 2019. *An <s> is an <s'>, or is it? Plural and genitive-plural are not homophonous*, To appear in Lívia Körtvélyessy & Pavol Stekauer (eds.), *Complex words*. Cambridge: Cambridge University Press.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005a. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62. 146–159.

Pluymaekers, M., M. Ernestus & R. H. Baayen. 2005b. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of American* 118. 2564–2569.

Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects in fine phonetic detail: The case of Dutch -igheid. In Cécile Fougeron, Barbara Kuehnert, Mariapaola D'Imperio & Nathalie Vallee (ed.), *Papers in laboratory phonology*, vol. 10, 511–531. Berlin & New York: Mouton de Gruyter.

R Core Team. 2019. R: A language and environment for statistical computing. https://www.R-project.org.

Ramscar, Michael & Daniel Yarlett. 2007. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science* 31. 927–960.

Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34. 909–957.

Rescorla, Robert A. 1988. Pavlovian conditioning. It's not what you think it is. *American Psychologist* 43. 151–160.

Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlocian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In William F. Prokasy & Abraham H. Black (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton Century Crofts.

Ridouane, Rachid & Pierre A. Hallé. 2017. Word-initial geminates: From production to perception. In Haruo Kubozono (ed.), *The phonetics and phonology of geminate consonants*, vol. 2 (*Oxford studies in phonology and phonetics*), 66–84. Oxford, UK: Oxford University Press.

Roelofs, Ardi & Victor S. Ferreira. 2019. The architecture of speaking. In Peter Hagoort (ed.), *Human language: From genes and brains to behavior*, 35–50. Cambridge: MIT Press.

Robinson, Cecil & Randall E. Schumacker. 2009. Interaction effects: Centering, variance inflation factor, and interpretation issues. *Multiple Linear Regression Viewpoints* 35. 6–11.

Rstudio Team. 2018. Rstudio: Integrated Development Environment for R. http://www.rstudio.com.

Schiel, Florian. 1999. Automatic phonetic transcription of non-prompted speech. In *Proceedings of the ICPhS*, 607–610.

Schmitz, Dominic, Ingo Plag, Dinah Baer-Henney & Simon David Stein. 2021. Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology* 12. 1–20.

Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–213. New York & East Sussex: Lawrence Erlbaum.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133. 140–155.

Seyfarth, Scott, Marc Garallek, Gwendolyn Gillingham, Farrell Ackermann & Robert Malouf. 2017. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33. 1–18.

Smith, Rachel H., Rachel Baker & Sarah Hawkins. 2012. Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics* 40. 689–705.

Stein, Simon David & Ingo Plag. 2021. Morpho-phonetic effects in speech production: Modeling the acoustic duration of English derived words with linear discriminative learning. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.678712.

Sugahara, Mariko & Alice Turk. 2004. Phonetic reflexes of morphological boundaries at a normal speech rate. In Bernard Bel & Isabelle Marlien (eds.), *Speech prosody*, 353–356. Groningen: University of Groningen.

Sugahara, Mariko & Alice Turk. 2009. Durational correlates of English sublexical constituent structure. *Phonology* 26. 477–524.

Swanson, Lori A. & Laurence B. Leonard. 1994. Duration of function-word vowels in mother's speech to young children. *Journal of Speech & Hearing Research* 37. 1394–1405.

Tang, Kevin & Jason A. Shaw. 2021. Prosody leaks into the memories of words. *Cognition* 210. 104601.

Tomaschek, Fabian, Peter Hendrix & R. Harald Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249–267.

Tomaschek, Fabian, Ingo Plag, R. Harald Baayen & Mirjam Ernestus. 2019. Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics* 57. 1–39.

Torreira, Fransisco & Mirjam Ernestus. 2009. Probabilistic effects on French [t] duration. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, 448–451.

Tremblay, Antoine & Johannes Ransijin. 2015. LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models [R package]. https://CRAN.R-project.org/package=LMERConvenienceFunctions (accessed August 2019).

Tucker, Ben V., Michelle Sims & R. Harald Baayen. 2019. Opposing forces on acoustic duration. https://doi.org/10.31234/osf.io/jc97w. https://psyarxiv.com/jc97w.

Umeda, Noriko. 1977. Consonant duration in American English. *Journal of the Acoustical Society of America* 61. 846–858.

van de Vijver, Ruben & Dinah Baer-Henney. 2014. Developing biases. *Frontiers in Psychology* 5. Article 634. https://doi.org/10.3389/fpsyg.2014.00634.

Vitevitch, Michael S. & Paul A. Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36. 481–487.

Wagner, Allan R. & Robert A. Rescorla. 1972. Inhibition in pavlovian conditioning: Application of a theory. In Robert A. Boakes & M. S. Halliay (eds.), *Inhibition and learning*, 301–336. New York: Academic Press.

Walsh, Liam, Jen Hay, Derek Bent, Liz Grant, Jeanette King, Paul Millar, Viktoria Papp & Kevin Watson. 2013. The UC QuakeBox project: Creation of a community-focused research archive. *New Zealand English Journal* 27. 20–32.

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

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental duration in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91. 1707–1717.

Winter, Bodo. 2019. *Statistics for linguists: An introduction using R*. New York: Routledge.

Yao, Yao. 2007. Closure duration and VOT of word-initial voiceless plosives in English in spontaneous speech. *UC Berkeley PhonLab Annual Report* 3. 183–225.

Zee, Tim, Louis Ten Bosch, Ingo Plag & Mirjam Ernestus. 2021. Paradigmatic relations interact during the production of complex words: Evidence from variable plurals in Dutch. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.720017.

Zimmermann, Julia. 2016. Morphological status and acoustic realisation: Findings from NZE. In Christopher Carignanand & Michael D. Tyler (eds.), *Proceedings of the sixteenth Australasian international conference on speech science and technology*, 201–204. Parramatta.

Zvonik, Elena & Fred Cummins. 2003. *The effect of surrounding phrase lengths on pause duration*, 777–780. Geneva: Proceedings of Eurospeech.