Morphological status and acoustic realization: Findings from New Zealand English

Julia Zimmermann

Heinrich-Heine-University Düsseldorf

julia.homann@uni-duesseldorf.de

Abstract

This paper investigates the acoustic realization of morphemic and non-morphemic S in New Zealand English. A corpus study is reported that examines the role of morphological structure in fricative duration. Multiple linear regression is used to isolate these effects, which are then compared to previous findings on the homophony of morphemic and non-morphemic S in General American English. The results demonstrate the importance of morphological structure in speech production.

Index Terms: phonetic detail; morphological structure; English; homophony

1. Introduction

Recent research on lexeme homophony has shown that seemingly homophonous lexemes actually differ in phonetic details such as duration and vowel quality (e.g. [1], [2]). This poses a challenge to traditional models of speech production which locate frequency information at the level of the phonological form, and which postulate that phonetic processing and the module called 'articulator' do not have access to any information regarding the lexical origin of a sound (e.g. [3], [4]). Leaving stylistic and accentual differences aside, a certain string of phonemes in a given context should therefore always be articulated in the same way according to these models, irrespective of its morphemic status, and only show phonetic variation originating from purely phonetic sources such as speech rate or context.

The findings on lexemes prompt the question of whether similar differences also hold for allegedly homophonous affixes (instead of free lexemes). Early experimental research found some evidence that morphemic and non-morphemic sounds may differ acoustically. Walsh & Parker [5] carried out a production experiment and measured the duration of /s/ in three pairs of monomorphemic words and their homophonous counterparts that contained a final morphemic /s/ (e.g. *lapse* versus *laps*). In two out of three experimental conditions they found a small difference in the means of the two different kinds of /s/, with morphemic, i.e. plural /s/, being on average nine milliseconds longer than non-morphemic. Similarly, Losiewicz [6] investigated the acoustic difference between morphemic, i.e. past tense, /d/ and /t/, and non-morphemic /d/ and /t/ using an experimental setup, and also found durational differences between the two sets of sounds, with past tense /d/ and /t/ being longer than non-morphemic /d/ and /t/. Both of these studies, however, only considered very small data sets and did not control for all potentially confounding covariates that might have influenced the duration of the segments.

More recently, Plag, Homann & Kunter [7] conducted a corpus study to investigate the duration of S (that is [s] or [z]) as non-morphemic instances and as markers of plural, genitive, genitive plural, 3rd person singular and the cliticized forms of *has* and *is* in General American English. They used multiple regression modelling to control for pertinent covariates and found systematic differences in duration between the different kinds of S. However, their results went in the opposite direction of those of Walsh & Parker [5], with non-morphemic S being longer than the morphemic S. Furthermore, within the group of morphemic S, the affixes were found to be systematically longer than the clitics.

Seyfarth et al. [8], like Walsh & Parker [5], find morphemic S to be longer than non-morphemic S when considering homophonous word pairs such as *lacks* and *lax*. They used an experimental setup in which pairs of participants read out naturalistic dialogues that served as carriers for the words under investigation. The effect they find is ascribed to phonetic paradigm uniformity, where "[a] word's phonetic realization is influenced by the articulatory plans of its morphological relatives" [8], i.e. the articulatory plan of the base of the complex word (*lack*) affects the phonetic realization of the complex word (*lacks*), while no such effect is available for the simplex word (*lax*). The authors attribute the differences between their own and Plag, Homann & Kunter's findings to the sample size of the corpus study, which they suspect to have caused an imbalance in terms of the syntactic positions in which the items occur.

These divergent findings and open questions call for further evidence about the nature of durational differences between morphemic and non-morphemic S in English.

2. Morphemic and non-morphemic S in New Zealand English

The present study extends the research on the acoustic properties of affixes by looking at the behavior of S in a different variety of English, namely New Zealand English. Using over 6,900 items from the Quakebox corpus [9], the duration of morphemic and non-morphemic S is investigated in order to test whether New Zealand English shows the same systematic durational differences as found for General American English by [5], [7] or [8], whether it displays a different pattern or whether it displays no difference at all.

If there are indeed the same differences in the durations of the different S to be found as in Plag, Homann & Kunter [7], this would underpin the notion that the acoustic realization of English S is influenced by its morphemic status and furthermore strengthen the corpus-based findings in [7].

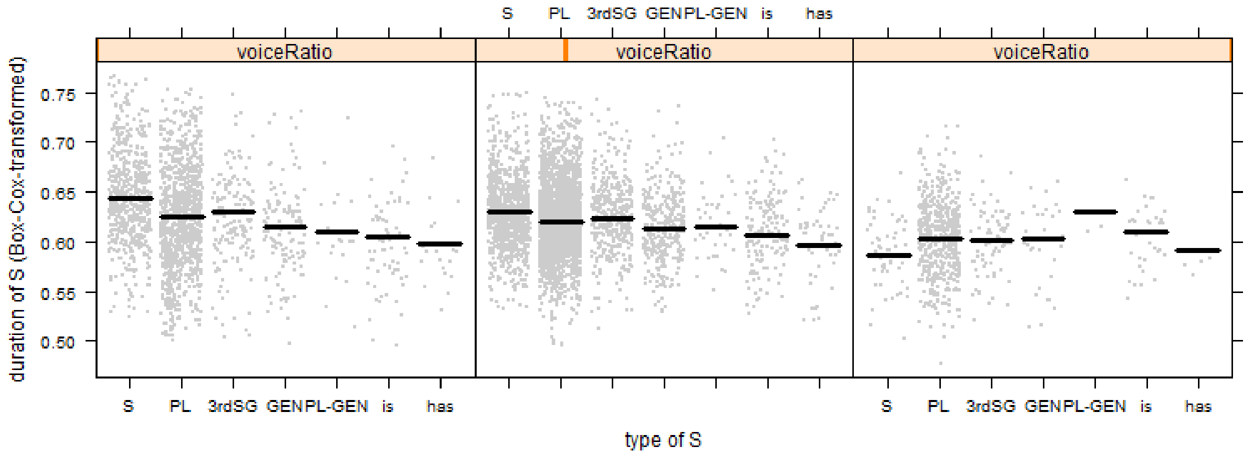

Figure 1: *Interaction plot for estimated box-cox transformed durations of S by amount of voicing and type of S. Model estimates are represented by solid horizontal lines, distribution of actual measured values is represented by grey dots. Panels represent voiceless (0-12% voicing), partially voiced (12-62% voicing) and fully voiced (62-100% voicing) items from left to right.*

2.1. Data

Following Plag, Homann & Kunter [7], non-morphemic instances of word-final S and S as a marker of plural, genitive, genitive plural, 3rd person singular and the cliticized forms of *has* and *is* were included in the analysis. Examples of each type of S in context are given in (1).

(1) non-morphemic: a *series* of aftershocks
(2) plural: there were huge *clouds* of dust
(3) genitive: my *family's* houses were okay
(4) genitive plural: we went to my *parents'* house
(5) 3rd person singular: something *falls* on it
(6) *has*-clitic: all this *mud's* come up
(7) *is*-clitic: the *lift's* broken

Items, i.e. morphemic S and the base it is attached to, or non-morphemic S and the word of which it forms the final segment, were sampled from the Quakebox Corpus [9]. This corpus is a collection of transcribed audio and video recordings of Cantabrians talking about their experiences in two major earthquakes that occurred in Christchurch in 2010 and 2011. The interviews were recorded in the 'QuakeBox', a shipping container which had been converted for use as a transportable recording studio. It was placed in different locations across the city of Christchurch. Audio was sampled at 48kHz stereo, using two Earthworks SR30 microphones (one headset microphone worn by the participant, one ceiling microphone inside the booth) and a USBPre2 microphone amplifier on a laptop computer running Audacity. At the time of data collection for the current study, 85 hours of high-quality recordings containing over 830,000 word tokens produced by 774 speakers were available. To keep the dataset free from potential dialectal differences, only those speakers who had identified themselves as native speakers of New Zealand English with a European background were included in this study (N=368). Using the corpus' automatically aligned phonetic transcriptions, all S-final words that were not followed by an S-initial word were extracted from the relevant recordings. Irregular forms, grammatical categories except for indefinite pronouns, brand and place names and items ending in in [ɪz] or [əz] were excluded manually. Due to the nature of Quakebox, the initial dataset of about 15,000 items was highly imbalanced in terms of type frequencies, with e.g. *house* contributing more than 2,000 items. For balancing purposes, only up to 25 randomly selected tokens were considered per type, leading to a reduced dataset of about 7,600 items. With the help of a Praat [1] script, relevant acoustic measures such as duration and voicing were extracted automatically.

In order to validate the automatic segmentations, the segmentations of 240 randomly selected items were checked manually using Praat [10]. Several different measurements for frequential center of gravity of the S in these items were considered as indicators for the reliability of the automatic segmentations, as any non-fricative material included in the S would have an effect on its frequential center of gravity. For each type of frequential measurement, the respective values for the automatic and the manual segmentations were then plotted against each other. Visual inspection of these plots showed clear tendencies where the automatic segmentations deviated most from the manual segmentations. For instance, frequential center of gravity weighted by the absolute spectrum based on the automatic segmentations yielded values from 1,000 to 13,500 Hz, while using the respective manual segmentations yielded values from 3,000 to 10,000 Hz. Therefore, any items with a frequential center of gravity weighted by the absolute spectrum ranging below 3,000 or above 10,000 Hz were excluded from the reduced dataset. The final set contained 7,081 tokens (from 1,879 types) for which the automatic segmentations were considered reliable.

2.2. Results

Linear mixed effects regression with a number of pertinent covariates (such as frequency, speaking rate, phonetic environment, etc.) was used to predict the duration of the S. The distribution of the durations of the S was slightly skewed and thus lacked linearity. This could have yielded unreliable estimates in linear regression, since one of the central assumptions of any linear regression model is a linear relationship between the dependent and independent variables. To alleviate this problem, the durations were Box-Cox transformed ([11], $\lambda = 0.2222222$).

Table 1: *Significance levels of individual contrasts between voiceless types of S as found in this study (Significance codes: '***' p < 0.001, '**' p < 0.01, '*' p < 0.05, '.' p < 0.1). Durations range from longest in leftmost column/top row to shortest in rightmost column/bottom row.*

	S	PL	3SG	GEN	PL-G	is	has
S	///	***	***	***	***	***	***
PL		///		**	.	***	***
3SG			///		*	***	***
GEN				///		*	*
PL-G					///		
is						///	
has							///

Models were fitted starting out with a fully specified model that contained all predictors that could be expected to have an effect on the (transformed) duration of S according to previous research (e.g. [12], [13], [14]). Stepwise exclusion of insignificant predictors, following the same simplification procedure as employed by Plag, Homann & Kunter [7], led to the final model, which showed a significant random effect of Speaker and significant main effects of Speaking Rate, Number Of Consonants in the rhyme of the final syllable of the item, Number Of Syllables in item, Duration Of The Base, Type Of Following Segment (`pause`, `affricate`, `approximant`, `fricative`, `nasal`, `plosive`, `vowel`), Duration Of Following Sound, Duration Of Preceding Sound, number of uses of item in Previous 30s of speech, log of Item Frequency in Quakebox, log of frequential Center Of Gravity by absolute spectrum and an interaction between Type Of S and Amount Of Voicing Of S (i.e., the ratio of voiced frames to total number of frames in the S). All effects go in the expected directions.

Figure 1 displays the average model estimates for the interaction between Voicing and Type Of S. The panels represent voiceless (left panel, 0-12% voicing), partially voiced (middle panel, 12-62% voicing) and fully voiced (right panel, 62-100% voicing) items, while type of S and transformed duration of S can be found on the x- and y-axis, respectively. As can be seen in the two leftmost panels, voiceless and partially voiced non-morphemic S are longer than most other types of S in those two conditions, while suffix S tend to be longer than clitic S. The back-transformed estimated mean durations for the group of voiceless S range from 98ms (*has*) to 139ms (non-morphemic S). The significance levels of all individual contrasts between voiceless types of S can be found in Table 1. The three ranges for the amount of voicing used in Figure 1 are based on the distribution of voicing in the dataset, which displays three main peaks, as illustrated in Figure 2.

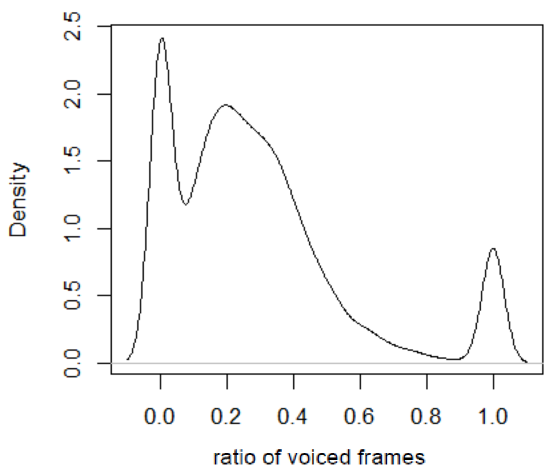

Figure 2: *Density distribution of amount of voicing in S*

Table 2: *Significance levels of individual contrasts between voiceless types of S as found by Plag, Homann & Kunter [7] (Significance codes and duration range identical to Table 1).*

	S	PL	3SG	GEN	PL-G	is	has
S	///	**	*	***	**	***	***
PL		///				*	*
3SG			///			*	*
GEN				///			
PL-G					///		
is						///	
has							///

3. Discussion

The study presented in this paper provides evidence for the existence of correlates of morphological structure in the acoustic signal. The duration of S in New Zealand English is dependent on morphological status. These findings clearly pattern with those by Plag, Homann & Kunter [7] for American English. In fact, the contrasts between the different kinds of voiceless morphemic S are even more pronounced in New Zealand English than they are in American English (cf. Table 2), with four additional significant contrasts in the former compared to the latter. The estimated durational difference between non-morphemic S and *has*-clitic S in this study (41ms) is also very close to the one observed by Plag, Homann & Kunter (38ms). Altogether, this study was able to replicate their results using a dataset more than ten times the size of the dataset of the original study. In these dimensions, potential imbalances in terms of syntactic position of the items, as suspected by Seyfarth et al. [8] about Plag, Homann & Kunter's dataset, should not be an issue.

At a very general level, these findings can be interpreted as support for the idea that there is morphological information in the phonetic signal, i.e. in postlexical stages of speech production. This goes against the assumptions of standard feed-forward formal theories of morphology–phonology interaction (e.g. [12], [16]). In these models, allomorphy is determined at a particular phonological cycle inside the lexicon, and at the level of underlying representations. Once the correct underlying form is derived, the morphological boundary of the respective cycle is erased (a process called 'bracket erasure', see [12], [16]) and the form leaves the lexicon. All further phonological processes are relegated to another module called 'postlexical phonology' and later to the articulatory component, neither of which have access to morphological information. According to my findings, it is possible to trace information about the structural status of a sound in the acoustic signal. Thus, the observed differences between the different TYPES OF S call into question the distinction between lexical and post-lexical phonology [16], which in turn would have important implications for theoretical mechanisms like bracket erasure and cyclic application of morpho-phonological rules.

At the theoretical level, these findings further challenge standard assumptions in models of speech production. Well-established models of speech production and the mental lexicon seem unable to accommodate my findings. Levelt, Roelofs & Meyer [17], for example, assume that pre-programmed gestures, which are stored in a syllabary, are executed by the articulator for the discrete syllables and segments of a language, which are phonologically represented. However, the articulator cannot provide a pre-programmed gesture for each syllable of a language if different meanings cause differences in these gestures. It is problematic that in such models, morphologically dependent sub-phonemic detail is not part of these representations. Such detail would need to be accounted for by purely phonetic factors that influence articulatory implementation such as speech rate [3]. For the duration of S, such an account is ruled out, as the effect of the type of S persists besides purely phonetic influences.

To summarize, both phonological theory and extant psycholinguistic models fail to provide a convincing explanation for the existence of morphological structure in the acoustic signal that I find in my data.

4. Conclusion

This paper has systematically investigated the relationship between morphemic status and phonetic implementation of homophonous affixes and their non-morphemic counterparts. This was done using natural conversation data. The analysis has yielded important evidence on the question of affix homonymy, revealing that phonologically homophonous bound morphemes can be phonetically distinct, and that morphemic and non-morphemic S may differ as well. This is unpredicted by current linguistic and psycholinguistic theories of lexicon and grammar. Further studies are certainly called for to be able to develop new models of the mental lexicon and of the relationships between morphology, phonology and phonetic implementation.

Furthermore, additional research is needed to address the many questions the present study raises. If there are indeed systematic differences between the different types of S in speech production, one would also like to know whether language users are influenced by these differences in perception. The difference in mean estimated duration between plural and *has*-clitic S amounts to 41ms. This difference lies well above the threshold for differentiating two fricative sounds that only differ in duration (e.g. [18], [19]) and translates to the average plural S being more than 1.4 times as long as the average *has*-clitic S.

5. Acknowledgements

This research was funded by the Deutsche Forschungs-gemeinschaft (Forschergruppe 2373 'Spoken Morphology', TP3 VAR, PL 151/8-1 and ZP PL 151/7-1), which I gratefully acknowledge.

I would also like to thank Jen Hay and everyone at UC's NZILBB in Christchurch for their support in carrying out this study.

6. References

[1] Drager, K. Sociophonetic variation and the lemma. Journal of Phonetics 39(4):694–707. 2011.

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

[3] Levelt, W. J. M. Speaking: From intention to articulation. Cambridge, MA: The MIT Press. 1989.

[4] Levelt, W. J. M. and Wheeldon, L. R. Do speakers have access to a mental syllabary? Cognition 50(1):239–269. 1994.

[5] Walsh, T. and Parker, F. The duration of morphemic and non-morphemic /s/ in English. Journal of Phonetics 11(2):201–206. 1983.

[6] Losiewicz, B. L. The effect of frequency on linguistic morphology. Ph.D. dissertation, University of Texas, Austin. 1992.

[7] Plag, I., Homann J. and Kunter, G. Homophony and morphology: The acoustics of word-final S in English. Journal of Linguistics, Available on CJO 2015 doi:10.1017/S0022226715000183. 2015.

[8] Seyfarth, S., Garellek, M., Malouf, R., and Ackerman, F. Acoustic differences in morphologically-distinct homophones. Talk presented at the 3rd American International Morphology Meeting. Amherst, MA. 2015.

[9] Walsh, L., Hay, J., Bent, D., Grant, L., King, J., Millar, P., Papp, V. and Watson, K. The UC QuakeBox Project: creation of a community-focused research archive. New Zealand English Journal, 27:20–32. 2013.

[10] Boersma, P. and Weenink, D. J. M. Praat: Doing phonetics by computer (Version 4.3.14) http://www.praat.org. 2013.

[11] Box, G. E. P. and Cox, D. R.. An analysis of transformations. Journal of the Royal Statistical Society, Series B 26(2):211–252. 1964.

[12] Hanique, I., Ernestus, M. and Schuppler, B. Informal speech processes can be categorical in nature, even if they affect many different words. The Journal of the Acoustical Society of America 133(3):1644–1655. 2013.

[13] Pluymaekers, M., Ernestus, M. and Baayen, R. H. Lexical frequency and acoustic reduction in spoken Dutch. Journal of the Acoustical Society of America 118(4):2561–2569. 2005.

[14] Pluymaekers, M., Ernestus, M., Baayen, R. H. and Booij, G. Morphological effects in fine phonetic detail: The case of Dutch -igheid. In Cécile Fougeron (ed.), Laboratory phonology 10 (Phonology and Phonetics), 511–531. Berlin and New York: Mouton de Gruyter. 2010.

[15] Chomsky, N. and Halle, M. The sound pattern of English. New York: Harper and Row. 1968.

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

[17] Levelt, W. J. M., Roelofs, A. and Meyer, A. S. A theory of lexical access in speech production. Behavioral and Brain Sciences 22(01):1–38. 1999.

[18] Klatt, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America 59(5):1208–1221. 1976.

[19] Shatzman, K.B., & McQueen, J.M. Segment duration as a cue to word boundaries in spoken-word recognition. Perception & Psychophysics, 68(1):1-16. 2006.