Morphological status and acoustic realization

Is there a difference between Bra[d] Pitt and a grille[d] cheese omelet, or between Kate Mo[s] and killer robot[s]?

Julia Zimmermann
Heinrich-Heine-Universität
Düsseldorf, Germany
julia.homann@hhu.de

***Abstract*—This paper investigates the acoustic realization of morphemic and non-morphemic S and D in English. Two corpus studies are reported that examine the effect of morphological structure on fricative center of gravity and on stop duration. Multiple linear regression is used to isolate these effects. The results demonstrate the importance of morphological structure in speech production.**

Keywords—phonetic detail; morphological structure; English; homophony

I. 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. [4], [5]). 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. [9], [10]). 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 [18] 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 /s/ being on average nine milliseconds longer than non-morphemic, i.e. plural, /s/. Similarly, Losiewicz [12] 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. Both 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 [14] conducted a corpus study to investigate the duration of English 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*. 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 [18], 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.

These diverging findings call for further evidence about the nature of the effect of morphological status on acoustic realization. The present study extends the research on the acoustic properties of affixes in two dimensions. First, I test whether the phonetic differences between the different kinds of S found by Plag, Homann & Kunter [14] go beyond duration. The acoustic parameter I look at is spectral center of gravity. Second, another group of English affixes and clitics, namely past tense, past participle and adjectival marker -ed, as well as cliticized forms of *had* and *would*, and their non-morphemic counterpart, final [d] and [t], is investigated in order to test whether other homophonous suffixes show the same systematic durational differences as found for S by Plag, Homann & Kunter [14].

II. Morphemic and non-morphemic S

One measure of consonant reduction, besides duration, is the spectral center of gravity. Very generally, it can be understood as the "mean" frequency of a consonant. "[F]or fricatives, the [center of gravity] frequency is inversely related to the size of the cavity in front of the noise source" [17]. Fig. 1 illustrates the difference between the centers of gravity of a [s] as in *wants* [w🕘nts] and a [ʃ] as in *actually* [ækʃli:] by displaying the two sounds' spectrograms. These spectrograms range from 0 to 8000Hz on the y-axis, time is represented on the x-axis, and the intensity of the shaded areas indicates the amplitudes of the component frequencies that make up the sound. Darker shades represent higher amplitudes. The [s] has a higher center of gravity than the [ʃ], which can be seen from the higher location of the darker shaded areas in the

This research was funded by the Deutsche Forschungsgemeinschaft and by Strategischer Forschungsförderfonds der Heinrich-Heine-Universität Düsseldorf, which I gratefully acknowledge.

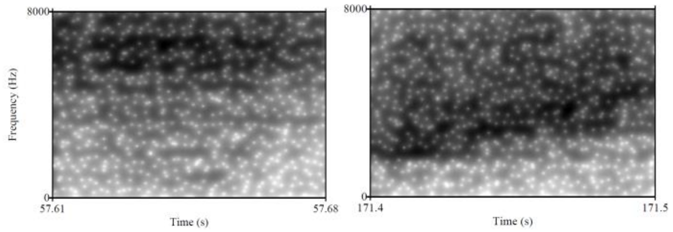

Fig. 1: Spectrograms of a [s] (left) and a [ʃ] (right)

spectrogram. For the [s], they average around 5170Hz, while for the [ʃ] they lie a little lower, at an average of 4000Hz.

If there are differences in the centers of gravity of the different S to be found, this could underpin the notion that on the phonetic level, they are indeed acoustically different sounds.

A. Data

With regard to the center of gravity of non-morphemic S and of the different morphemic S's I used the same set of 644 items that was used by Plag, Homann & Kunter. They had sampled up to 100 tokens consisting of bases/hosts and their respective S for each kind of morphemic/clitic S (or less, where 100 were not available) and 199 tokens of words ending in non-morphemic S from the Buckeye Corpus [13]. This corpus comprises about 300,000 words from 40 long-time local residents of Columbus, Ohio, who were recorded conversing freely with an interviewer for about one hour each. In addition to the raw speech files, the Buckeye Corpus offers time-aligned written and phonetic transcriptions of the interviews. Plag, Homann & Kunter [14] had checked Buckeye's (partly automatic) phonetic annotations manually for each item and adjusted them where necessary. Boundaries marking the beginning of an item or of an S were moved to the zero crossing that was closest to the point where both spectrogram and waveform indicated the initiation of the gesture for the respective segment, i.e. in the case of S, the boundaries were set to the zero crossing closest to the onset of the friction visible in the waveform. Boundaries marking the end of an item and thus the end of an S were moved to the zero crossing closest to the point where the initiation of the gesture for the following segment became visible in both spectrogram and waveform. In cases with no following segment, the boundary was set to the point where the friction of the S dropped to silence.

With the help of a Praat [1] script, I extracted the spectral Center Of Gravity of the middle portion of the S, disregarding 20% each at the beginning and at the end. The beginning and the end were excluded from the analyses to exclude any potential coarticulation with neighboring sounds that could have an influence on the center of gravity. Weighting of the center of gravity was done by the absolute spectrum. Other relevant acoustic measures such as duration and voicing were extracted automatically as well.

B. Results

Linear mixed effects regression with a number of pertinent covariates (such as speaker gender, different lexical frequency measures, speaking rate, phonetic environment, etc.) was used to predict the Center Of Gravity as a function of the Type Of S. Models were fitted starting out with a fully specified model that contained all predictors that could be expected to have an effect on the phonetic details of S according to previous research (e.g. [6], [8], [14], [15], [16]). This includes, for example, that female speakers can generally be expected to have higher voices and thus higher centers of gravity in their fricatives than male speakers. Highly frequent items tend to be more reduced than infrequent items. Likewise, faster speech usually shows more reduction than slower speech. In both cases, more reduction could equal lower centers of gravity. Stepwise exclusion of insignificant predictors then led to the final model. A predictor was considered insignificant if it passed three tests. First, its t-statistics had to yield a t-value greater than 2 (or less than –2) when included in the model. Second, the Akaike information criterion (AIC) of the model including the factor had to be lower than the AIC of the model without it. Third, an ANOVA comparing the model including the factor to a model without it had to yield a p-value lower than 0.05, thus showing that the inclusion of the factor did significantly improve the fit of the model. A variable under consideration was only retained in the model if it passed all three tests. The final model showed a significant random effect of Speaker and significant main effects of Speaker Gender, Voicing Of S, Duration Of S, Duration Of Base, Number Of Syllables In Base and of Type Of Following Segment (`pause, affricate, approximant, fricative, nasal, plosive, vowel`).

Type Of S was found to have a significant effect on the Center Of Gravity as well. The major contrast lies between the *has*-clitic and all other kinds of S, with the *has*-clitic having a significantly lower center of gravity than genitive S (p=0.024), plural S (p=0.0213), plural-genitive S (p=0.0114) and *is*-clitic (p=0.0154), and a marginally significantly lower center of gravity than 3rd person singular S (p=0.0632) and non-morphemic S (p=0.0802). Fig. 2 displays the model estimates for the different types of S, with the different Types Of S on the x-axis and Center Of Gravity on the y-axis.

III. Morphemic and non-morphemic D

Another set of English morphemes that is standardly assumed to share some of the same allomorphs at the phonological level are past tense, past participle and adjectival marker *-ed*, as well as cliticized forms of *had* and *would*. In this part of the study, I focus on the duration of the allomorph D (that is [d] or [t]) and its non-morphemic counterpart D, since this is the form that the clitics can take as well. Tapped or fully omitted instances of D were not considered in the analyses.

A. Data

Items were sampled from the Buckeye Corpus [13]. Using its POS-tagged orthographic transcription, 40 tokens of each morphemic D and 120 non-morphemic D's were randomly extracted. If less than 40 tokens were available for a certain kind of morphemic D, all available tokens were extracted. With regard to adjectival D, 40 items each were sampled for attributive and predicative use. The overall set of extracted items amounted to 359. Of these items, 18 were excluded

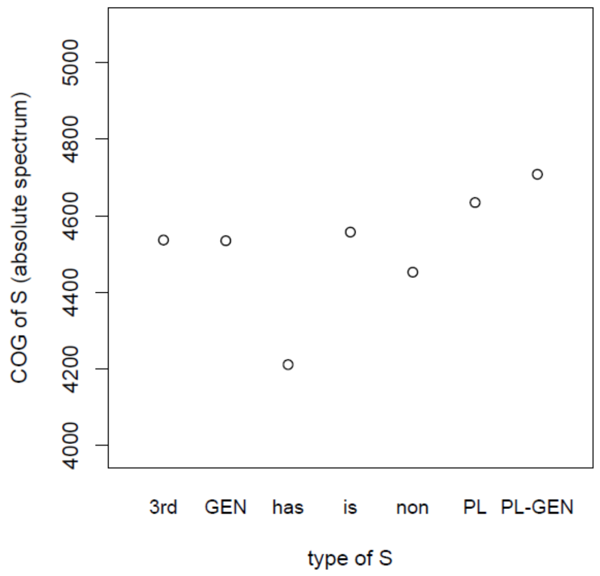

Fig. 2: Estimated center of gravity by type of S

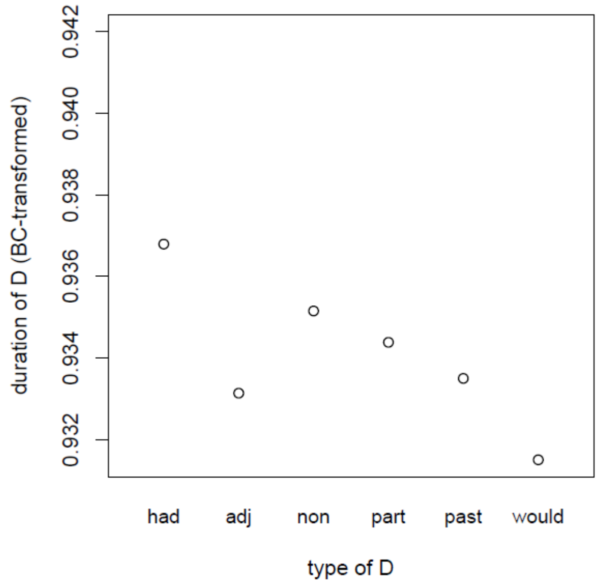

Fig. 3: Estimates of Box-Cox transformed durations of D by type of D

because the final D was not unambiguously attributable to the word in question due to assimilation to an initial stop in the following word. The final set of 341 tokens represents 260 types. Buckeye's (partly automatic) phonetic annotations were checked manually for each item and adjusted where necessary according to the same procedure as described above for S. With the help of a Praat [1] script, relevant acoustic measures such as duration and voicing were extracted automatically.

B. *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 complete obstruction of the D's. This particular interval was chosen because it is contained in all D's, while release and aspiration show more variation. The presence or absence of these additional phases was coded as a covariate though, and also included in the statistical model.

The distribution of the durations of the D'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 ([2], $\lambda = 0.02020202$).

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 D according to previous research (e.g. [6], [15], [16]). Stepwise exclusion of insignificant predictors, following the same simplification procedure as for S described above, led to the final model, which showed a significant random effect of SPEAKER and significant main effects of SPEAKING RATE, the NUMBER OF CONSONANTS in the rhyme of the final syllable of the item, the TYPE OF FOLLOWING SEGMENT (`pause`, `affricate`, `approximant`, `fricative`, `nasal`, `plosive`, `vowel`), VOICING of D and the presence of ASPIRATION in D, which all go in the expected directions. Fig. 3 displays the model estimates for the different TYPES OF D.

As can be seen, the *had*-clitic seems especially long compared to the other D's. The contrast between *would* and *had* turns out to be significant with p=0.0072, as well as the contrast between *would* and non-morphemic D with p=0.034, while the contrast between *had* and adjectival D proves marginally significant (p=0.0567). The back-transformed mean estimated durations range from 39ms (*would*) to 52ms (*had*). These results mirror Plag, Homann & Kunter's [13] findings for S in so far as there are systematic duration differences between the morphemes. However, the distribution of these differences does not pattern with those for S: durations of *has* and *is* clitics were found to be at the same (i.e. shorter) end of the scale, while I find *had* and *would* clitics to be the longest and one of the shortest D's, respectively.

IV. DISCUSSION: S AND D

Both studies presented in this paper provide evidence for the existence of correlates of morphological structure in the acoustic signal. Both the duration of D and the spectral properties of S are dependent on morphological status.

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. [3], [7]). In these models, allomorphy is determined at a particular phonological cycle inside the lexicon, and at the level of underlying representations. Once the right underlying form is derived, the morphological boundary of the respective cycle is erased (a process called 'bracket erasure', see [3], [7]) 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 and D call into question the distinction between lexical and post-lexical phonology [7], 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 [11], 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 [9]. For the center of gravity of S and for the duration of D, such an account is ruled out, as the effect of the type of S/D 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.

V. CONCLUSION

This paper has systematically investigated the relationship between morphemic status and phonetic implementation of homophonous affixes and their non-morphemic counterpart. 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 and D may differ, too. This is unpredicted by current linguistic and psycholinguistic theories of the lexicon and grammar. Further studies are certainly called for to replicate the observed effects, and to develop new models of the mental lexicon and of the relationships between morphology, phonology and phonetic implementation.

Furthermore, more research is needed to address the many questions the present study raises. If there are indeed systematic differences between the different types of S and D 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 *had*-clitic and *would*-clitic D amounts to 13ms. Although small, this difference could potentially be perceptible, given that it translates to the average *had*-clitic being 33% longer than the average *would*-clitic. It is also plausible that listeners might make use of the different spectral properties of the different types of S.

REFERENCES

[1] Boersma, Paul & David J. M. Weenink. 2013. *Praat: Doing phonetics by computer* (Version 4.3.14) (2005). http://www.praat.org.

[2] Box, George E. P. & David R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society*, Series B 26.2, 211–252.

[3] Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

[4] Drager, Katie. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39.4, 694–707.

[5] Gahl, Susanne. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84.3, 474–496.

[6] Hanique, Iris, Mirjam Ernestus & Barbara Schuppler. 2013. 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.

[7] 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

[8] Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59.5, 1208–1221.

[9] Levelt, William J. M. 1989. *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

[10] Levelt, William J. M. & Linda R. Wheeldon. 1994. Do speakers have access to a mental syllabary? *Cognition* 50.1, 239–269.

[11] Levelt, William J. M., Ardi Roelofs & Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.01, 1–38.

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

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

[14] Plag, Ingo, Julia Homann & Gero Kunter. 2015. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, Available on CJO 2015 doi:10.1017/ S0022226715000183

[15] Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118.4, 2561–2569.

[16] Pluymaekers, Mark, Mirjam Ernestus, R. Harald Baayen & Geert Booij. 2010. Morphological effects in fine phonetic detail: The case of Dutch -igheid. In C´ecile Fougeron (ed.), *Laboratory phonology 10 (Phonology and Phonetics)*, 511–531. Berlin and New York: Mouton de Gruyter.

[17] Van Son, Rob J. J. H. & Louis C. W. Pols. 1996. An acoustic profile of consonant reduction. *Proceedings of ICSLP'96, Philadelphia, USA*, 1529–1532.

[18] Walsh, Thomas & Frank Parker. 1983. The duration of morphemic and non-morphemic /s/ in English. *Journal of Phonetics* 11.2, 201–206.