Does morphology impact the pronunciation of consonant clusters? Evidence from German

Chiara Celata, Maria Paola Bissiri, Carolin Schmid

Abstract

This study investigates the phonetics-morphology interface by asking whether the acoustic detail plays a role in distinguishing homophone but functionally different phonotactic structures. Word-final consonant clusters included in monomorphemic or bi-morphemic German words (as in *Obst* "fruit" as opposed to *tobst* "(you) rave", where the cluster is the result of morpheme concatenation) were produced by native speakers of Austrian German in a reading task; target words could occur in two different prosodic conditions, that is, phrase-finally (where pre-boundary lengthening was expected to occur) or phrase-internally. Significant variations in the duration of the clusters were found as a function of their lexical or morphological status, the prosodic condition in which they occurred, and the duration of the preceding vowel. Moreover, clusters' status had an impact on how pre-final lengthening was realized. The results are discussed in the framework of the current criticism of established models of phonological representation and speech production.

Keywords: phonetics-morphology interface, consonant clusters, phonotactics, German.

1. *Introduction*

There is currently much debate about which word-level characteristics are systematically encoded in the phonetic and phonological form of words. It has been known for decades that phonological homophones differ systematically at the phonetic level depending on factors such as word frequency or the lexical class to which they pertain: high-frequency words such as *time* are almost 20% shorter than low-frequency homophones such as *thyme* (Gahl, 2008; Jurafsky *et al.*, 2002); systematic acoustic differences signal nouns and verbs in non-

Received: July 2022
Accepted: November 2022

SSL LX (2) 2022, pp. 51-79
doi: 10.4454/ssl.v60i2.329

stress-shifting disyllabic word pairs such as *answer* (verb) versus *answer* (noun) (Sereno and Jongman, 1995).

Recent research is showing that gradient phonetic detail can be used to encode other kinds of paradigmatic contrasts and grammatical functions. In addition to word frequency, factors such as word length, spelling, and pragmatic meaning are encoded in subtle acoustic cues (e.g. Bybee, 2002; Drager, 2011; Cohen-Goldberg, 2015). For example, Martinuzzi and Schertz (2022) have found that multiple subphonemic acoustic differences distinguish *sorry* as apology from *sorry* as attention-seeking, and these differences are consistently used by the listeners when asked to identify the word. Moreover, lexical neighborhood density and the existence of minimal neighbors influence the way in which sounds are produced (e.g. Wright, 2004; Baker and Bradlow, 2009; Gahl *et al.*, 2012; Goldrick *et al.*, 2013; Clopper and Tamati, 2014). For example, English voiceless stops have longer VOT when embedded in words that have a minimal neighbor for voicing (e.g. *teen* vs. *dean*) compared to words without such minimal neighbor (e.g. *table* vs. **dable*) (Baese-Berk and Goldrick, 2009).

All these effects support the idea of interactive mechanisms which link production and perception performance directly to the processing of word-level semantic and functional information. This hypothesis challenges a strictly modular and feedforward view of language processing in which lexical information cannot influence the phonetic implementation directly, bypassing the level of phonological information (Pierrehumbert, 2002).

Along this research line, an important amount of investigation is recently devoted to the phonetics of morphological structure, that is, to the potential impact of morphological complexity on how words are produced, perceived, and eventually spelled. For instance, it has been repeatedly shown that word-final [s] in English varies systematically for duration as a function of the morphological function it expresses (non-morphemic, plural, third person singular, genitive, genitive plural, cliticized *has*, and cliticized *is*) (e.g. Tomaschek *et al.*, 2019). Phonetics is therefore not blind to the morphological component either (e.g. Ben Hedia and Plag, 2017; Plag *et al.*, 2017; Seyfarth *et al.*, 2018; Strycharczuk, 2019; Schlechtweg and Corbett, 2021; Schmitz *et al.*, 2021). The number of paradigmatic neighbors (morphological family) also correlates with the phonetic characteristics of words (Loo *et al.*, 2018).

One question that arises from this type of research is whether complex phonotactic structures that are known to be processed differently by the speakers according to their grammatical function in the words are actually different at the phonetic level or not. This study investigates such an issue by asking whether the acoustic detail plays a role in distinguishing homophone but functionally different word-final consonant clusters in German, a language in which word-final clusters may cue specific morphological functions but are also allowed intra-morphemically. Before describing the study, this introduction succinctly overviews the existing knowledge about consonant clusters in speech production and presents the motivations for extending the scope of the investigation on the grammar-phonetics interface into the domain of consonant clusters.

Consonant clusters are a frequent target of empirical and theoretical analysis in phonetics and phonology, according to a variety of perspectives and methodologies that includes production and perception studies, historical, psycholinguistic/acquisitional and typological approaches.

One topic of traditional investigation concerns the adaptation processes (reductions, assimilations, elisions, insertion of epenthetic vowels, etc.) that clusters undergo especially in connected speech and less controlled styles. Variation in the rate and in the phonetic output of adaptations depends on a multiplicity of phonetic and perceptual factors, such as the principles of coarticulatory resistance (e.g. Recasens, 2018), the degree of gestural cohesion within the syllable (e.g. Pouplier and Goldstein, 2010; Hermes *et al.*, 2013), universal preferences such as those concerning the sonority of segments and their relative perceptibility (e.g. Dziubalska-Kolaczyk, 2015), the frequency and length of words (e.g. Greenberg, 1978), the difference between function and content words (e.g. Zimmerer *et al.*, 2009), and others. While most studies focus on word-internal clusters, phonetic variation in clusters has also been studied as a function of lexical integrity. For instance, it has been shown that the same consonantal sequence undergoes a different amount of phonetic change when it is word-internal compared to when it spans a word boundary (e.g. Ellis and Hardcastle, 2002; Celata *et al.*, 2012); across word boundaries, the relative frequency of the two words also predicts the amount of phonetic assimilation, with more cohesion across lexical boundaries involving high-frequency words (e.g. Bergmann, 2012).

However, less is known about the role of grammatical factors in shaping the phonetics and phonology of consonant clusters. Considering phonology, it is well known that grammatical operations such as morpheme concatenation increase the typology of consonant clusters that are allowed in a language; the result is that in most languages, some clusters only exist across morpheme boundaries (e.g. Engl. /-md/ as in *seemed*). These clusters have been called 'morphonotactic' by some authors (Dressler and Dziubalska-Kołaczyk, 2006) in order to highlight that they come out from morphological operations and as such, they are different from 'phonotactic' clusters that are allowed also intra-morphemically.

The distinction between morphonotactic and phonotactic clusters is important because the morphological integrity of clusters has been shown to impact various processing levels. For instance, in strongly inflecting and morphologically rich languages such as Polish or Lithuanian, the morphological status of clusters has been shown to impact the order of acquisition of clusters by children by facilitating the acquisition of complex consonant structures (e.g. Kamandulyte, 2006). In other inflecting languages, such as German and Italian, the morphological status of clusters modulates the precision with which adolescent and adult speakers identify and manipulate them through vowel epenthesis (Celata *et al.*, 2015), the processing of cluster transition probabilities in visual word recognition (Celata, 2020), and the accuracy and speed of compound identification (Sommer-Lolei *et al.*, 2021). However, almost nothing is known about the potential effects of morphological integrity on clusters' phonetic realization.

One pilot study in this domain is by Leykum *et al.* (2015) on Standard Austrian German. Several repetitions of 16 minimal or quasiminimal pairs with intra-morphemic and cross-morphemic word-final clusters (e.g. [nst] in *Kunst* "art" vs. *kannst* "you can") were acoustically analyzed (336 target items in total). No clear effect of the morphological status of the cluster was found on either cluster duration or other parameters (cluster intensity and preceding vowel duration); only the rate of [t] deletion was found to be higher in cross-morphemic clusters than in lexical clusters. The Authors concluded that more investigation is needed in order to fully reject the hypothesis that cross-morphemic clusters are phonetically different from intra-morphemic ones, firstly because the sentence reading task could have provided some redundant information about inflection of the verb forms containing cross-morphemic clusters, and secondarily because the dimension of the dataset is rather small. Another similar study by Leykum and Moosmüller (2021) investigated two varieties of German (Austrian and German) and French. For the two varieties of German, homophone intra-morphemic and cross-morphemic [t]-ending clusters of two, three, and four consonants were analyzed for acoustic duration, relative intensity, and rate of [t] deletion. The results suggested that cross-morphemic clusters (such as in *schafft* "she/he creates") were slightly longer than intra-morphemic ones (as in *Schaft* "shaft"), but the Authors attributed the effect to performance limitations of the speaking task and did not interpret it as conclusive evidence supporting a phonetic difference between the two types of clusters. In word-medial positions, homophone clusters were compared for both varieties of German (e.g. [nkt] in *Akupunktur* "acupuncture" as opposed to *Funkturm* "radio tower") and also for a small number of French items (e.g. [zl] in *islandaise* "Icelandic" as opposed to *dislocation* "dislocation"). A slight effect of cluster type on duration was found in some data subsets defined by other experimental parameters (e.g. in the speech of males compared to females and in nouns compared to adjectives), but the Authors acknowledged that it was not possible to control for the phonological context (i.e., there was unsystematic variation in phonemes surrounding the consonant clusters, in the position of the cluster within the words and in the position of lexical stress with respect to the cluster) and this could have had a blurring effect.

2. *Current study on German word-final lexical and morphological clusters*

2.1. *Hypotheses*

This production study asks if German word-final homophone clusters are acoustically different when they are part of the lexical morpheme or when they arise from morpheme concatenation.

Based on the evidence reviewed above, we hypothesize that the morphological function of word-final clusters is indexed acoustically. In particular, we assume that a clear articulation of the segments composing the cluster is more important in the production of morphological than lexical sequences (e.g. van Son and Pols, 2003; Ben Hedia and Plag, 2017). This might lead to the relative hyperarticulation of consonants composing cross-morphemic clusters (henceforth, morphological clusters or MCs) compared to consonants composing intra-morphemic clusters (henceforth, lexical clusters or LCs); we thus expect MCs to be significantly longer than LCs.

We tested the hypothesis in two different prosodic contexts, namely, phrase-internally and phrase-finally; in the latter context, pre-boundary lengthening was supposed to occur. Pre-boundary lengthening is a manifestation of prosodically conditioned variation in segmental duration for which a large body of evidence has been collected (see Cho, 2016, for a review). By including this factor in the design, we wanted to evaluate if potential differences in the acoustic duration of MCs and LCs vary as a function of intervening sentence-level factors that are known to influence the duration of segments.

2.2. *Materials*

We selected sixteen words, eight of which containing an LC and eight a homophone MC. The two groups were balanced for word length (calculated in number of syllables), quality and quantity of the vowel preceding the cluster (with only one exception, see Table 1 below), and average word frequency. Word frequency was extracted from

German CELEX (*http://web.phonetik.uni-frankfurt.de/simplex.html*). For each word we calculated both its form frequency in the corpus and its 'type' frequency, that is, the cumulated frequency of all inflected forms of that lexical item (i.e., all cases and numbers for nouns, all forms of the verbal paradigm for verbs). The average form frequency was 97 whereas the average 'type' frequency was 216. As can be seen in Table 1, the two groups of items were dissimilar for their average form frequency, which was 194 for the LC items and 0.5 for the MC items; however, they were similar for average 'type' frequency, which was 201 and 231, respectively. To account for these differences, 'type' frequency was included in the statistical design.

In addition, sixteen fillers (half nouns and half verbs) were selected. Ten were disyllables and six were monosyllables; they were varied for word onset (two of them beginning with a vowel, nine with a consonant, two with a CC cluster, and three with a CCC cluster) as well as for their word offset (five beginning with a vowel, five with a consonant, and six with a CC cluster). Their average form frequency was 66 and their 'type' frequency was 158.

For the reading task, each target word was included in two different sentences, corresponding to two prosodic conditions. In the phrase-internal prosodic condition, the speakers had to read a question (*Was hast du gesagt?* "What did you say?") followed by a reply in which the target word was in phrase-internal position (*Ich habe X gesagt* "I said X"). In the phrase-final condition, both the question and the answer were slightly varied such that the target item was placed in phrase-final position: *Was sagst du heute? Heute sage ich X* "What do you say today? Today I say X").

It should be noted that the carrier sentences did not contain any source of additional morphosyntactic information correlated to the nature of the cluster; in particular, verbs were not preceded by their subjects, and nouns were not preceded by an article. This means that the target items were actually produced by the speakers as 'citation forms' within the sentence. This aspect may have had an influence on the outcome of the experiment and will be discussed in the final section.

Cluster	Word with LC	English gloss	Form frequency	Type frequency	Word with MC	English gloss	Form frequency	Type frequency
/nst/	*Wanst*	“paunch”	5	5	*bannst*	“(you) banish”	0	76
/nst/	*Gespenst*	“ghost”	27	43	*verkennst*	“(you) underestimate”	0	63
/pst/	*Obst*	“fruit”	68	68	*tobst*	“(you) rave”	0	42
/lst/	*Wulst*	“bead”	3	4	*nullst*	“(you) zero”	0	2
/ŋkst/	*Angst*	“fear”	596	608	*langst*	“(you) are enough”	0	878
/rnst/	*Ernst*	“gravity”	533	555	*lernst*	“(you) learn”	4	720
/rpst/	*Herbst*	“autumn”	303	313	*färbst*	“(you) paint”	0	53
/lpst/	*Selbst*	“oneself”	16	16	*wölbst*	“(you) bend”	0	16
Average			*194*	*201*			*0.5*	*231*

Table 1. *Experimental stimuli with their form and ‘type’ frequencies.*

2.3. *Methodology and participants*

Fourteen native Standard Austrian German speakers (9 female, 5 male) participated in the reading task. For a phonetic and phonological description of Standard Austrian German (SAG) see Moosmüller (1991) and Moosmüller *et al.* (2015). The speakers were recruited online via social media. Inclusion criteria were that they were born and raised in Vienna and had spent a large part of their life in Vienna, that they had a university entrance diploma from secondary school (i.e., *Matura*) and were between 20 and 40 years of age[1].

The recordings were conducted in a sound-attenuated booth at the Acoustics Research Institute of the Austrian Academy of Sciences in Vienna. The recording setup included an Edirol Roland R-44 recorder and an AKG C451 EB microphone. The signal was digitized at 44.1 kHz.

The speakers had to read off the sentences from a sheet of paper. Before the recording started, the speakers were orally instructed on the reading task. They were familiarized with the target words in order to avoid irritations and disfluencies resulting from inflected forms of infrequent verbs or nouns. Moreover, they were instructed to use clear, but not exaggerated speech, to produce the target words in focus, but without inserting a phrase-boundary after the target words in phrase-internal condition, and to not turn the pages while reading. The speakers were encouraged to repeat a sentence in case they mispronounced the target word.

Two sentence lists had been created for the reading task. Speakers were divided into two groups (gender-mixed) according to the sentence list they were asked to read. Both sentence lists consisted of 48 different sentences and included all the 16 filler sentences in both prosodic conditions. Additionally to the filler sentences, one list included 4 LC sentences and 4 non corresponding MC sentences in both prosodic conditions, and the other list included the other 4 LC and 4 MC sentences in both prosodic conditions. Each list was repeated twice so that each speaker read 96 sentences in total (32 realizations of LC and MC clus-

---

[1] Note that one of the female speakers did not fully meet the inclusion criteria as she only lived in Vienna for her studies. She was nevertheless included in the study as she was considered as a SAG speaker by other phonetically trained native SAG speakers.

ters). The order of the sentences was separately randomized (randomization function in excel) for the two repetitions of each list by manually ensuring that identical words in different prosodic conditions were not too close to each other (at least 10 different words intervening). The overall number of experimental items (words with LCs and MCs) was 14 speakers x 8 items x 2 prosodic conditions x 2 repetitions = 448. Due to the exclusion of 7 misproductions, the final dataset included 441 data points.

2.4. *Analysis*

The audio files were annotated in Praat 6.1.40 (Boersma and Weenink, 2021) as in the example shown in Figure 1. For each item, we annotated the vowel (V) and following cluster interval (C) in one tier; vowel quality and the individual consonants composing the cluster were phonetically annotated in SAMPA in another tier.

Concerning the phonetic annotation, in some cases it was impossible to identify the boundary between the [s] and the following [t]; in those cases, the interval was labeled as <st> (as in the example shown in Figure 1).

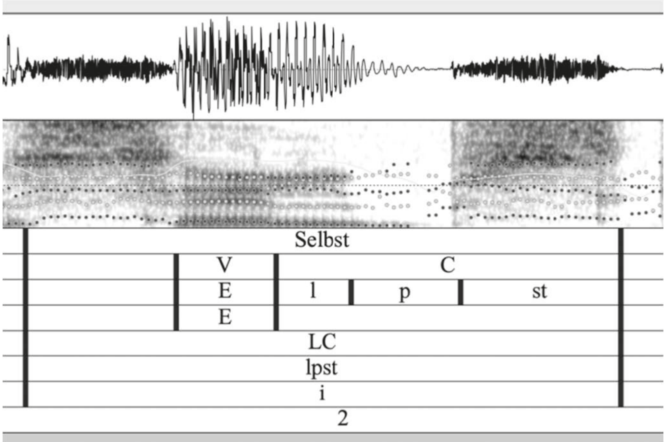

Figure 1. *Example of annotation in Praat.*

This inconsistency was not problematic for the current study since the analysis focused on the duration of the cluster as a whole. Stimuli with a rhotic were produced with a vocalized rhotic, as usual in Standard

Austrian German (Moosmüller *et al.*, 2015). Therefore, they were annotated with a centralized diphthong ([E6] in SAMPA or [ɛɐ] in IPA) in the V slot, followed by the consonant cluster; for instance, in the case of *Herbst*, V corresponded to [E6] and C to [pst]. The rhotic was produced as a consonant (more specifically, as a uvular tap preceded and followed by a vocoid) in only one word in the corpus; the rhotic was labeled as [R] and included in the C interval and the item was excluded from further analysis. Creaky voice periods between glottal stop and vowel (in vowel-initial words such as *Angst* or *Obst*) were attributed to the stop, not to the vowel (unless the vowel was entirely creaky voiced, which occurred a few times).

The annotation included the following additional information: the type of the cluster (either LC or MC), the phonological transcription of the cluster, the prosodic condition ('i' for the phrase-internal condition and 'f' for the phrase-final one), and the repetition (1 or 2). Concerning the prosodic condition, there were nine instances in the corpus, in which two of the speakers produced a pause after the target word in the phrase internal condition; these were labeled as <i-f> and excluded from further analyses.

For this study we focused on the duration of the whole cluster; further phonetic parameters (duration of individual segments, non durational phonetic indices) will be analyzed in the prosecution of the research. The duration of the preceding vowel was also included in the analysis, according to the statistical design described below. The two durational indices were extracted automatically via a Praat script.

A regression model with cluster duration as the dependent variable was run in R (lmer function). Fixed factors were cluster Status (LC vs. MC), Prosody (phrase-internal vs. phrase-final), Vowel Duration (continuous variable), and Type Frequency (logarithmically scaled). Random slopes and intercepts for Speaker and Word were included. The model with the most complex structure was run[2], but it did not converge and was affected by singularity. This starting model

---

[2] The model had the following formula: lmer(CDuration ~ Status * Prosody * VDuration * log.TypeFreq + (1 + Status | Speaker) + (1 + Prosody | Speaker) + (1 + VDuration | Speaker) + (1 + Status | Word) + (1 + Prosody | Word) + (1 + VDuration | Word)).

was then simplified by means of the step function of the R package lmerTest (version 3.1.3) by Kuznetsova *et al.* (2017), removing non-significant factors and interactions, however the resulting model did not converge. Our approach was to find the best fit by removing the least possible from the random effects in the starting model. After dropping by-Word random slope for VDuration from the starting model and running the step function, we obtained a model with a singularity error. The same occurred after dropping by-Speaker random slope for vowel duration as well. In order to minimally reduce the random structure, we then run the step function on two different starting models: one in which by-Speaker random slope for cluster status was dropped[3], and another one in which by-Speaker random slope for prosody was dropped instead[4]. We ended up with two different resulting models that were almost equivalent in terms of goodness of fit (r^2, AIC values) and had very similar residuals; they differed slightly for the complexity of their random structures and for some significance in the fixed effects and interactions. Both resulting models did not include the logarithmic Type Frequency, which was dropped by the step function as non-significant. In what follows, we report the results of both, in order to provide the most informative picture and then discuss major and minor effects in the most grounded way. Both models are also fully shown in the *Appendix*.

3. *Results*

We start by reviewing the results obtained by the model with the (slightly) higher r^2 and lower AIC value (model 1, henceforth). Model 1's random structure was also slightly more complex than that of model 2, including by-Speaker random slope for Prosody and random

---

[3] Starting formula for first model: lmer(CDuration ~ Status * Prosody * VDuration * log.TypeFreq + (1 + Prosody | Speaker) + (1 + VDuration | Speaker) + (1 + Status | Word) + (1 + Prosody | Word)).

[4] Starting formula for second model: lmer(CDuration ~ Status * Prosody * VDuration * log.TypeFreq + (1 + Status | Speaker) + (1 + VDuration | Speaker) + (1 + Status | Word) + (1 + Prosody | Word)).

intercept for Word. According to model 1 (graphed in Figure 2), there was a significant and expected effect of Prosody (clusters in phrase-final position were longer than clusters in phrase-internal position) and a very small but significant effect of VDuration (the longer the vowel, the shorter the cluster).

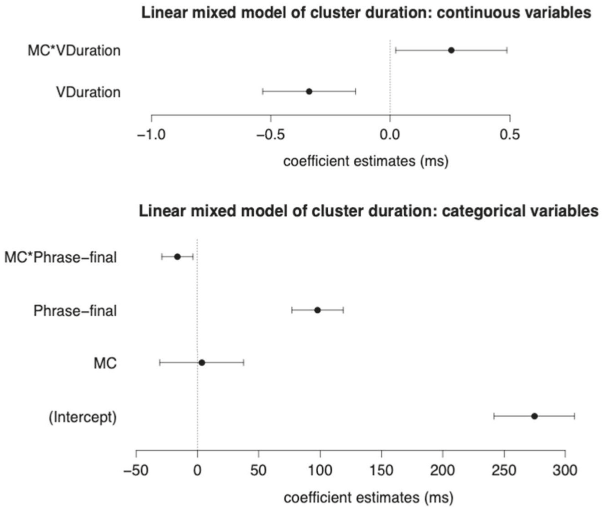

Figure 2. *Model 1's effects (top: continuous variables; bottom: categorical variables).*

Cluster Status (LC vs. MC) was included in the model, which estimated MCs 3.5 ms longer than LCs, but the effect was not significant. There were, however, two significant interactions of cluster Status with Prosody and with VDuration. These are illustrated in Figure 3.

The significant interaction between cluster Status and Prosody and corresponding Least Square Means showed that LCs were 31 ms shorter than MC in phrase-internal position, and this difference was significant with $p < 0.01$; in phrase-final position, the difference amounted to only 15 ms, and was not significant ($p \approx 0.248$) (Figure 3). Cluster duration was negatively correlated to vowel duration, however, the interaction between cluster Status and VDuration showed that this effect was significantly weaker for MCs (Figure 4).

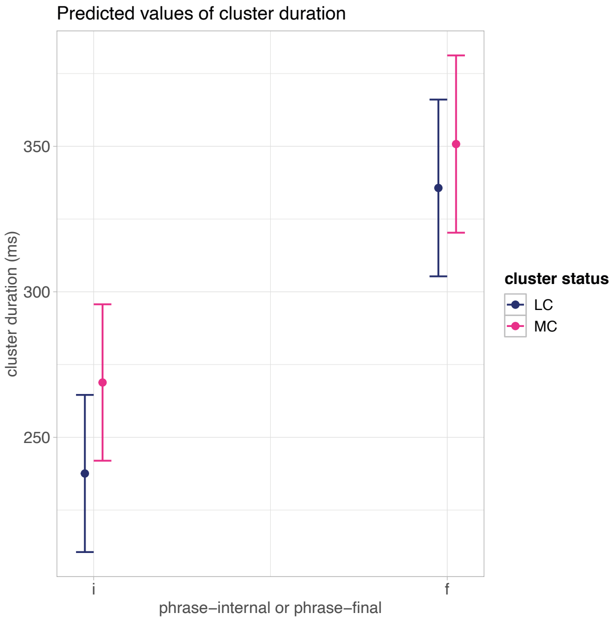

Figure 3. *Model 1's interactions: cluster duration as a function of cluster Status and Prosody.*

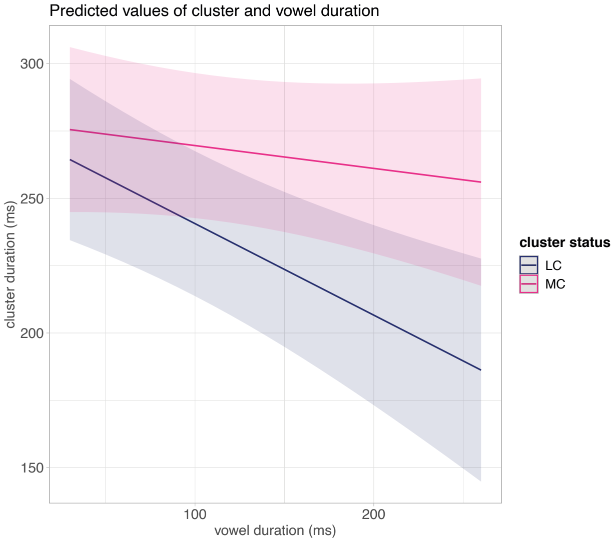

Figure 4. *Model 1's interactions: cluster duration as a function of cluster Status and VDuration.*

We now turn to model 2, which had a slightly higher AIC and lower r^2 (thus, a slightly worse fit compared to model 1), and a simpler random structure, only including random intercepts for Word and Speaker. According to model 2 (graphed in Figure 5), there was a significant effect of cluster Status, with MCs showing a longer duration than LCs (estimated difference: 29 ms).

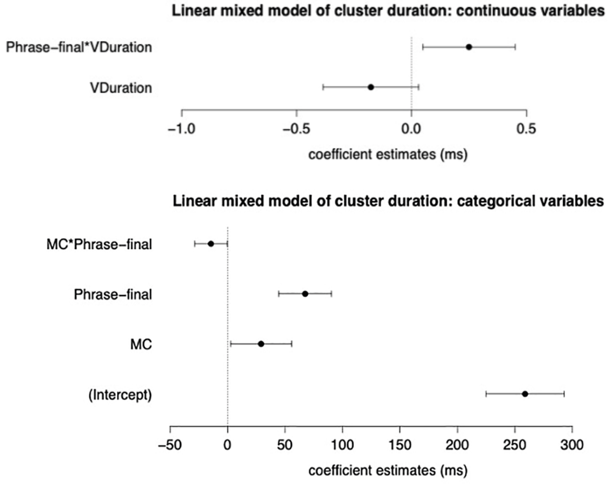

Figure 5. *Model 2's effects (top: continuous variables; bottom: categorical variables).*

As in model 1, there was also a significant effect of Prosody (phrase-final > phrase-internal); by contrast, the effect of VDuration only approached significance ($p \approx 0.0964$). There were two significant interactions, and only one concerned cluster Status.

As in model 1, the Least Squares Means calculated on model 2 show that the difference between MCs and LCs is bigger in phrase-internal (estimate: 29 ms; significant at $p < 0.05$) than in phrase-final position (estimate: -15 ms; not significant); see Figure 6.

Moreover, and differently from model 1, the interaction between Prosody and VDuration showed a significant trend reversal in phrase-final position: phrase-finally the correlation between cluster duration and VDuration is significantly more positive than phrase internally (Figure 7).

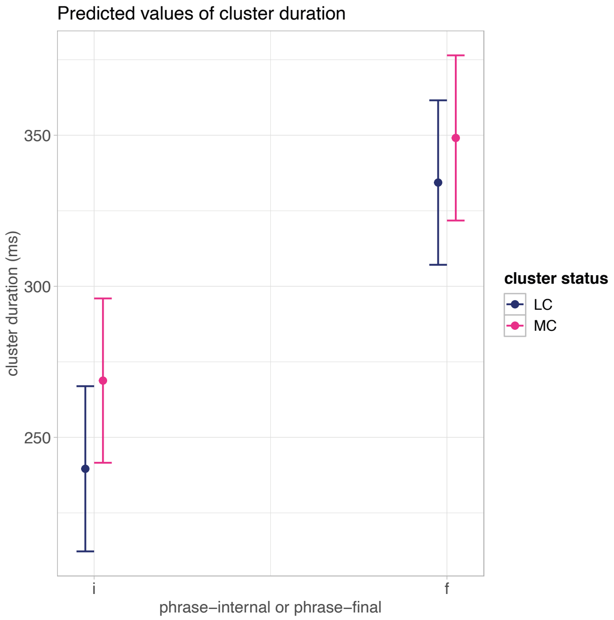

Figure 6. *Model 2's interactions: cluster duration as a function of cluster Status and Prosody.*

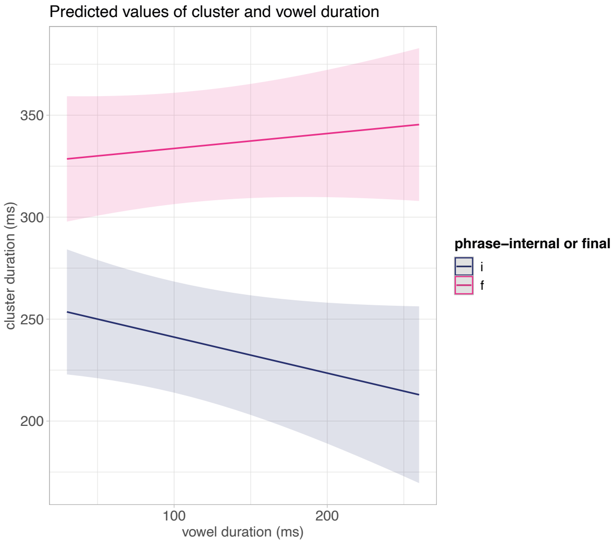

Figure 7. *Model 2's interactions: cluster duration as a function of VDuration and Prosody.*

4. *Discussion*

This study tested the hypothesis that word-final homophone clusters in German are longer when the segments composing the cluster pertain to two different morphemes (as in verb form *tobst* "(you) rave") compared to when the word form is monomorphemic and the cluster does not span any morpheme boundary (as in noun *Obst* "fruit"). The effect was expected based on the hypothesis that, although the clusters are phonologically the same in both conditions, in the case of morpheme concatenation the clusters carry specific morphological meaning that might require the consonants to be more neatly articulated or less reduced in speech (Dressler and Dziubalska-Kołaczyk, 2006; Leykum and Moosmüller, 2021). Therefore, according to this view, the fine phonetic detail associated with cluster production would be directly predicted by word-level characteristics (Pierrehumbert, 2002; Tomaschek *et al.*, 2019) that are not encoded at the level of phonological specification in the strict sense. In this study we limited our analysis to cluster duration; nevertheless, other phonetic parameters, both durational and non-durational, could be equally assumed to change. This further hypothesis will have to be tested in future work.

We used a reading task in which native Austrian German speakers produced different target items in carrier sentences of different syntactic-prosodic shapes. Target items were either a noun (in the lexical condition: LCs) or a verbal form (in the morphological conditions: MCs); the two subsets of items were balanced for average word length (calculated in number of syllables) and very similar for other relevant phonological characteristics, such as the quality and phonological length of the preceding vowel. The two subsets were not balanced for form frequency (2nd singular inflected verb forms being less frequent than nouns in their nominative singular form in a reference corpus), however, they were balanced for average type frequency; moreover, frequency was consistently found to play no role in predicting cluster duration across different statistical models.

Given that the carrier sentences had to be semantically neutral and allow the target item to be collocated in specific prosodic positions, target items were included as ‘citation forms’ in the carrier sentences; e.g. *Was sagst du heute? Heute sage ich X*, or *Was hast du gesagt? Ich habe X gesagt* (“What do you say today? Today I say X” and “What did you say? I said X”, respectively). This means that the sentences did not include any additional or redundant morphosyntactic information (such as subject pronouns for verbs or articles for nouns), which might in principle have a different impact on the two subsets of items (Hanique and Ernestus, 2012). In particular, being German a non-pro-drop language, the production of verbal forms without their subject might have resulted in a less natural task than the production of bare nouns. However, it should also be considered that subject pronouns are not obligatorily pre-verbal in German. This said, nouns were presented in their nominative singular forms, which can be considered to be the ‘default’ citation form, whereas verbs were presented in their 2nd singular present indicative form, which can hardly be considered the ‘default’ citation form for verbs. So the question remains of whether verb items, that contain MC clusters, were perceived as less natural in the context of the sentence reading task, which might lead to overall slower reading and longer segments.

In order to deal with this problem, we tested whether vowel duration was influenced by cluster status. We reasoned that stressed vowel duration should be the first and foremost cue of potential speech rate variations across items. We therefore ran regression models with VDuration as the dependent variable and according to the same procedure explained above for the analysis of cluster duration. The results (see *Appendix*, models 3 and 4) showed that vowel duration only depended on prosodic condition (longer duration phrase-finally). In other words, we did not find any evidence of longer vowels in items with MCs, as one might have thought if the production of these items had been slowed down or made less fluid by the presence of isolated verbal forms.

In any case, further research will have to ascertain more precisely, via more ecological speech production paradigms, the potential role of unnatural wording in shaping the acoustic effects of extra-phonological variables, such as those taken into account here. While it is of first and foremost importance to keep the phonetic context perfectly balanced across the experimental subsets to be compared, it is also true that unnatural contexts of word elicitation might in principle obscure the speakers' recovering of the semantic and morphological meaning of forms to an extent that could vary across grammatical categories and that should be accurately quantified.

Given these premises, we think that this study has provided evidence in support of the role of acoustic detail in signaling the clusters' morphological status in German. To get the most comprehensive and faceted picture, we presented the results obtained from two non-identical but equally valid regression models, given that the effects under investigation could be very subtle, and the intervening variables potentially very numerous.

Model 2, which had a slightly less complex random structure, found an effect of cluster status on cluster duration (Figure 5). Here, the difference was estimated at 29 ms. The direction of the difference was as expected: lexical clusters (LCs) were found to be shorter than morphological ones (MCs).

Moreover, cluster status appeared to be shaped by the effect of other variables. Both models reported a strong effect of the prosodic condition: as expected based on pre-boundary lengthening, clusters in phrase-final position were consistently longer than clusters in phrase-internal position. However, model 1 showed that pre-boundary lengthening blurred the difference between LCs and MCs, whose duration was statistically equivalent in phrase-final position (Figure 3). According to model 2, the difference between LCs and MCs was significant in phrase-internal, but not in phrase-final position (Figure 6). These results are consistent with the fact that the effect of cluster status on cluster duration is very subtle and any other durational variation can obfuscate it. More specifically, the models show that MCs show proportionally less preboundary lengthening than LCs. This might be explained by the fact that MCs already tend to be longer and there might be a ceiling effect that hinders stretching the duration of the cluster beyond a given threshold. In any case, it is possible to conclude from these data that the cluster's morphological or lexical status may have an impact on how sentence-level durational variations are realized in production.

Our data also revealed that cluster duration and preceding vowel duration were inversely correlated (although in the second model the correlation only approached significance). In model 1, this effect was significantly weaker for MCs (Figure 4), thus pointing to a further difference between the two types of clusters. We hypothesize that the effect is due to the fact that in LCs, the vowel and the following cluster belong to the same morpheme and this might increase their articulatory cohesion, thus enhancing durational compensation effects compared to when a morpheme boundary intervenes (as in MCs). This is however only speculative and would require specific testing. By contrast, model 2 suggested that the correlation between vowel and cluster duration is affected by the prosodic condition: in phrase-final position, the duration of clusters and preceding vowels are more positively correlated (Figure 7), and this effect can be interpreted as a consequence of pre-boundary lengthening which applies uniformly across segments in sentence-final words.

In conclusion, this study supports a general view according to which word-level functional information can be encoded in the acoustic detail in a statistically significant way. Evidence is provided for a small dataset of cluster-final German words and more extensive investigation would be needed to unravel the role of additional factors that could shape the nature of the phonetics-morphology interface in this particular context, such as production variables associated with specific word elicitation paradigms or the degree of individual variability and the speakers' experience with the language. Further analysis of individual consonants and of non-durational acoustic parameters can also help specify the patterns of variation.

Acknowledgments

We are grateful to the participants in the experiment and to the Acoustics Research Institute, Austrian Academy of Sciences, for providing the necessary facilities and funds for the recordings. We also thank the participants of the MPP2 workshop (Düsseldorf, 7-9 February, 2022) and two anonymous reviewers for their helpful suggestions and criticisms to an earlier version of this work.

Author contributions

Conceptualization: C. Celata; design of the experiment: C. Celata and M.P. Bissiri, discussed with C. Schmid; recordings: C. Schmid; data pre-processing and annotation: C. Celata; acoustic phonetic analysis: C. Celata and M.P. Bissiri; statistical analysis: M.P. Bissiri; visualization: M.P. Bissiri. Writing – original draft preparation: C. Celata (§§ 1, 2.1, 2.2, 2.4, 4), M.P. Bissiri (§ 3) and C. Schmid (§ 2.3); writing – review and editing: all authors. All authors have read and agreed to the published version of the manuscript.

References

Baese-Berk, M. and Goldrick, M. (2009), *Mechanisms of interaction in speech production*, in «Language and Cognitive Processes», 24, 4, pp. 527-554.

Baker, R.E. and Bradlow, A.R. (2009), *Variability in word duration as a function of probability, speech style, and prosody*, in «Language and Speech», 52, pp. 391-413.

Ben Hedia, S. and Plag, I. (2017), *Gemination and degemination in English prefixation: Phonetic evidence for morphological organization*, in «Journal of Phonetics», 62, pp. 34-49.

Bergman, P. (2012), *Articulatory reduction and assimilation in* n#g-*clusters in complex words in German*, in Hoole, P., Bombien, L., Pouplier, M., Mooshammer, C. and Kühnert, B. (2012, eds.), *Consonant Clusters and Structural Complexity*, Mouton de Gruyter, Berlin / Boston, pp. 311-344.

Boersma, P. and Weenink, P. (2021), *Praat: Doing phonetics by computer* [available online at *https://www.fon.hum.uva.nl/praat/*].

Bybee, J. (2002), *Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change*, in «Language Variation and Change», 14, pp. 261-290.

Celata, C. (2020), *Bottom-up probabilistic information in visual word recognition*, in «Language Sciences», 78, p. 101267.

Celata, C., Calamai, S., Ricci, I. and Bertini, C. (2012), *Nasal place assimilation between phonetics and phonology. An articulatory study of Italian nasal-to-velar clusters*, in «Journal of Phonetics», 41, 2, pp. 88-100.

Celata, C., Korecky-Kröll, K., Ricci, I. and Dressler, W.U. (2015), *Phonotactic processing and morpheme boundaries: Word-final* /Cst/ *clusters in German*, in «Italian Journal of Linguistics», 27, 1, pp. 85-110.

Cho, T. (2016), *Prosodic boundary strengthening in the phonetics-prosody interface*, in «Language and Linguistics Compass», 10, 3, pp. 120-141.

Clopper, C.G. and Tamati, T.N. (2014), *Effects of local lexical competition and regional dialect on vowel production*, in «The Journal of the Acoustical Society of America», 136, 1, pp. 1-4.

Cohen-Golberg, A.M. (2015), *Abstract and lexically specific information in sound patterns: Evidence from* /r/*-sandhi in rhotic and non-rhotic varieties of English*, in «Language and Speech», 58, 4, pp. 522-548.

Drager, K. (2011), *Sociophonetic variation and the lemma*, in «Journal of Phonetics», 39, 4, pp. 694-707.

Dressler, W.U. and Dziubalska-Kołaczyk, K. (2006), *Proposing morphonotactics*, in «Wiener Linguistische Gazette», 73, pp. 1-19.

Dziubalska-Kołaczyk, K. (2015), *Are frequent, early and easy clusters also unmarked?*, in «Italian Journal of Linguistics», 27, pp. 29-44.

Ellis, L. and Hardcastle, W. (2002), *Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data*, in «Journal of Phonetics», 30, pp. 373-396.

Gahl, S. (2008), *Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech*, in «Language», 84, 3, pp. 474-496.

Gahl, S., Yao, Y. and Johnson, K. (2012), *Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech*, in «Journal of Memory and Language», 66, pp. 789-806.

Goldrick, M., Vaughn, C. and Murphy, A. (2013), *The effects of lexical neighbors on stop consonant articulation*, in «The Journal of the Acoustical Society of America, 134, 2, pp. EL172-EL177.

Greenberg, J.H. (1978), *Some generalizations concerning initial and final consonant clusters*, in Greenberg, J.H. (1978, ed.), *Universals in Human Language*. Vol. 2: *Phonology*, Stanford University Press, Stanford, pp. 243-280.

Hanique, I. and Ernestus, M. (2012), *The role of morphology in acoustic reduction*, in «Lingue e linguaggio», 2, pp. 147-164.

Hermes, A., Mücke, D. and Grice, M. (2013), *Gestural coordination of Italian word initial clusters: The case of 'impure* s*'*, in «Phonology», 20, 1, pp. 1-25.

Jurafsky, D., Bell, A. and Girand, C. (2002), *The role of the lemma in form variation*, in Gussenhoven, C. and Warner, N. (2002, eds.), *Laboratory Phonology*. Vol. 7, Mouton de Gruyter, Berlin / New York, pp. 3-34.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H. (2017), *lmerTest package: Tests in linear mixed effects models*, in «Journal of Statistical Software», 82, pp. 1-26.

Leykum, H., Moosmüller, S. and Dressler, W.U. (2015), *Word-final (mor)phonotactic consonant clusters in standard Austrian German*, in The Scottish Consortium for ICPhS 2015 (2015, ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, University of Glasgow, Glasgow (UK), n. 701.

Leykum, H. and Moosmüller, S. (2021), *Morphonotactics in speech production*, in Dressler, W.U., Calderone, B., Sommer-Lolei, S. and Korecky-Kröll, K. (2021, eds.), *Experimental, Acquisitional and Corpus Linguistic Approaches to the Study of Morphonotactics*, Austrian Academy of Sciences Press, Vienna, pp. 53-76.

Loo, K., Jarvikivi, J., Tomaschek, F., Tucker, B.V. and Baayen, R.H. (2018), *Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects*, in «Morphology», 28, pp. 71-97.

Martinuzzi, C. and Scherz, J. (2022), *Sorry, not sorry: The independent role of multiple phonetic cues in signaling the difference between two word meanings*, in «Language and Speech», 65, 1, pp. 143-172.

Moosmüller, S. (1991), *Hochsprache und Dialekt in Österreich. Soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien, Graz, Salzburg und Innsbruck*, Böhlau, Köln / Weimar.

Moosmüller, S., Schmid, C. and Brandstätter, J. (2015), *Standard Austrian German*, in «Journal of the International Phonetic Association», 45, 3, pp. 339-348.

Pierrehumbert, J.B. (2002), *Word-specific phonetics*, in Gussenhoven, C. and Warner, N. (2002, eds.), *Laboratory Phonology*. Vol. 7, Mouton de Gruyter, Berlin / New York, pp. 101-140.

Plag, I., Homann, J. and Kunter, G. (2017), *Homophony and morphology: The acoustics of word-final S in English*, in «Journal of Linguistics», 53, pp. 181-216.

Pouplier, M. and Goldstein, L. (2010), *Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production*, in «Language and Cognitive Processes», 25, 5, pp. 616-649.

Recasens, D. (2018), *The Production of Consonant Clusters. Implications for Phonology and Sound Change*, Mouton de Gruyter, Berlin / Boston.

Schlechtweg, M. and Corbett, G.G. (2021), *The duration of word-final* s *in English: A comparison of regular plural and pluralia-tantum nouns*, in «Morphology», 31, pp. 383-407.

Schmitz, D., Baer-Henney, D. and Plag, I. (2021), *The duration of word-final* /s/ *differs across morphological categories in English: Evidence from pseudowords*, in «Phonetica», 78, pp. 571-616.

Sereno, J.A. and Jongman, A. (1995), *Acoustic correlates of grammatical class*, in «Language and Speech», 38, 1, pp. 57-76.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F. and Malouf, R. (2018), *Acoustic differences in morphologically-distinct homophones*, in «Language, Cognition and Neuroscience», 33, 1, pp. 32-49.

Sommer-Lolei, S., Korecky-Kröll, K., Christiner, M. and Dressler, W.U. (2021), *Acquisition and processing of (mor)phonotactic consonant clusters in German*, in Dressler, W.U., Calderone, B., Sommer-Lolei, S. and Korecky-Kröll, K. (2021, eds.), *Experimental, Acquisitional and Corpus Linguistic Approaches to the Study of Morphonotactics*, Austrian Academy of Sciences Press, Vienna, pp. 77-100.

van Son, R.J.J.H. and Pols, L.C.W. (2003), *Information structure and efficiency in speech production*, in *8th European Conference on Speech*

Communication and Technology, EUROSPEECH 2003 – INTERSPEECH 2003 (Geneva, Switzerland, September 1-4, 2003), pp. 769-772 [doi: 10.21437/Eurospeech.2003-63].

Strycharczuk, P. (2019), *Phonetic detail and phonetic gradience in morphological processes*, in *Oxford Research Encyclopedia of Linguistics*, Oxford University Press, Oxford (UK).

Tomaschek, F., Plag, I., Baayen, H. and Ernestus, M. (2019), *Phonetic effects of morphology and context: Modeling the duration of word-final* S *in English with naïve discriminative learning*, in «Journal of Linguistics», 57, 1, pp. 123-161.

Wright, R.A. (2004), *Factors of lexical competition in vowel articulation*, in Local, J.J., Ogden, R. and Temple, R. (2004, eds.), *Laboratory Phonology*. Vol. 6, Cambridge University Press, Cambridge (UK), pp. 26-50.

Zimmerer, F., Scharinger, M. and Reetz, H. (2011), *When BEAT becomes HOUSE: Factors of word final /t/-deletion in German*, in «Speech Communication», 53, 6, pp. 941-954.

Chiara Celata
Dipartimento di Studi Umanistici
Università degli studi di Urbino 'Carlo Bo'
Via San Girolamo 8
61029 Urbino (Italy)
chiara.celata@uniurb.it

Maria Paola Bissiri
Dipartimento di Scienze Umane e dell'Innovazione per il Territorio
Università degli studi dell'Insubria
Via M.E. Bossi 5
22100 Como (Italy)
mariapaola.bissiri@uninsubria.it

Carolin Schmid
Acoustics Research Institute
Austrian Academy of Sciences
Wohllebengasse 12-14
1040 Wien (Austria)
carolin.schmid@oeaw.ac.at

Appendix

Model 1 (dependent variable: cluster duration)

```
Formula: CDuration ~ Status + Prosody + VDuration + (1 + Prosody | Speaker) + (1 |
Word) + Status:Prosody + Status:VDuration
  Data: MCLC_data
REML criterion at convergence: 4427.2

Scaled residuals:
Min                 1Q          Median      3Q          Max
-3.2438             -0.5432     -0.0416     0.4595      4.2242

Random effects:
  Groups            Name        Variance    Std.Dev.    Corr
  Word              (Intercept) 546.5       23.38
  Speaker           (Intercept) 1544.3      39.30
            Prosodyf            1294.1      35.97       -0.20
  Residual                      1072.0      32.74
Number of obs: 441, groups:  Word, 16; Speaker, 14

Fixed effects:
                    Estimate    Std. Error  df          t value   Pr(>|t|)
(Intercept)         274.60007   16.78752    50.29779    16.357    < 2e-16 ***
StatusMC            3.46935     17.48236    47.84196    0.198     0.843535
Prosodyf            98.11649    10.67669    15.98134    9.190     8.87e-08 ***
VDuration           -0.33995    0.09893     401.49882   -3.436    0.000652 ***
StatusMC:Prosodyf   -16.18212   6.44798     398.95933   -2.510    0.012481 *
StatusMC:VDuration  0.25514     0.11821     405.63644   2.158     0.031480 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
                    (Intr)      SttsMC      Prsdyf      VDurtn    StMC:P
StatusMC            -0.499
Prosodyf            -0.091      -0.002
VDuration           -0.575      0.405       -0.135
SttsMC:Prsd         -0.024      -0.031      -0.309      0.200
SttsMC:VDrt         0.368       -0.698      0.087       -0.640    -0.208

> extractAIC(MCLC_4_best.lmer)
[1]   11.000 4468.873
> r.squaredGLMM(MCLC_4_best.lmer)
       R2m         R2c
[1,] 0.3699896 0.8080329
```

Model 2 (dependent variable: cluster duration)

```
Formula: CDuration ~ Status + Prosody + VDuration + (1 | Speaker) + (1 | Word) + Status:Prosody + Prosody:VDuration
  Data: MCLC_data
REML criterion at convergence: 4498.8

Scaled residuals:
Min                 1Q          Median      3Q          Max
-2.9001             -0.5873     -0.0577     0.5554      4.7488

Random effects:
  Groups   Name          Variance    Std.Dev.
  Word     (Intercept)   622         24.94
  Speaker  (Intercept)   1439        37.94
  Residual               1369        37.00
Number of obs: 441, groups:  Word, 16; Speaker, 14

Fixed effects:
                    Estimate    Std. Error  df          t value     Pr(>|t|)
(Intercept)         258.8417    17.3314     56.9587     14.935      < 2e-16 ***
StatusMC            29.1877     13.4756     14.5305     2.166       0.0474 *
Prosodyf            67.5279     11.6807     410.2530    5.781       1.47e-08 ***
VDuration           -0.1766     0.1060      405.7591    -1.666      0.0964 .
StatusMC:Prosodyf   -14.4193    7.1960      411.1688    -2.004      0.0457 *
Prosodyf:VDuration  0.2498      0.1025      412.7105    2.438      0.0152 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
                    (Intr)      SttsMC      Prsdyf      VDurtn      StMC:P
StatusMC            -0.331
Prosodyf            -0.357      0.037
VDuration           -0.597      -0.093      0.495
SttsMC:Prsd         0.009       -0.268      -0.176      0.154
Prsdyf:VDrt         0.379       0.056       -0.899      -0.637      -0.142

> extractAIC(MCLC_5_best.lmer)
[1]   9.000  4534.719
> r.squaredGLMM(MCLC_5_best.lmer)
        R2m         R2c
[1,] 0.3689976  0.7481635
```

Model 3 (dependent variable: vowel duration)

Formula of the starting model: VDuration ~ Status * Prosody * log.TypeFreq + (1 + Status | Speaker) + (1 | Word)

Formula of best model obtained by means of the step function of the R package lmerTest (Kuznetsova *et al.*, 2017): VDuration ~ Prosody + (1 + Status | Speaker) + (1 | Word)

```
  Data: MCLC_data
REML criterion at convergence: 4003.5

Scaled residuals:
Min             1Q          Median      3Q          Max
-3.6801         -0.6352     -0.0513     0.5009      4.0918

Random effects:
  Groups        Name        Variance    Std.Dev.    Corr
  Word          (Intercept) 633.4       25.17
  Speaker       (Intercept) 246.3       15.69
         StatusMC           177.4       13.32       0.19
  Residual                  404.4       20.11
Number of obs: 441, groups:  Word, 16; Speaker, 14

Fixed effects:
                Estimate    Std. Error  df          t value   Pr(>|t|)
(Intercept)     103.641     8.037       25.921      12.90     8.75e-13 ***
Prosodyf         10.643     1.932       400.374      5.51     6.44e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
                (Intr)
Prosodyf        -0.118

> extractAIC(MCLC_VDur4_best.lmer)
[1]   7.000 4026.669
> r.squaredGLMM(MCLC_VDur4_best.lmer)
       R2m        R2c
[1,] 0.01968074  0.7194624
```

Model 4 (dependent variable: vowel duration)

Formula of the starting model: VDuration ~ Status * Prosody * log.TypeFreq + (1 + Prosody | Speaker) + (1 | Word)

Formula of best model obtained by means of the step function of the R package lmerTest (Kuznetsova *et al.*, 2017): VDuration ~ Prosody + (1 + Prosody | Speaker) + (1 | Word)

```
  Data: MCLC_data
REML criterion at convergence: 4010.4

Scaled residuals:
Min            1Q          Median     3Q         Max
-3.2782        -0.5965     -0.0795    0.4587     4.0876

Random effects:
  Groups       Name         Variance   Std.Dev.   Corr
  Word         (Intercept)  622.1      24.94
  Speaker      (Intercept)  221.5      14.88
          Prosodyf          122.9      11.08      0.50
  Residual                  416.1      20.40
Number of obs: 441, groups:  Word, 16; Speaker, 14

Fixed effects:
               Estimate    Std. Error  df         t value   Pr(>|t|)
(Intercept)    103.978      7.522      24.086     13.824    5.98e-13 ***
Prosodyf       10.774      3.553       13.088     3.033     0.00955 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
               (Intr)
Prosodyf       0.151

> extractAIC(MCLC_VDur5_best.lmer)
[1]  7.000 4034.548
> r.squaredGLMM(MCLC_VDur5_best.lmer)
       R2m        R2c
[1,] 0.02033013 0.709054
```