



V-to-V coarticulation induced acoustic and articulatory variability of vowels: The effect of pitch-accent

Andrea Deme^{1,2}, Márton Bartók^{1,2}, Tekla Etelka Grácsi^{3,2}, Tamás Gábor Csapó^{4,2},
Alexandra Markó^{1,2}

¹Eötvös Loránd University, Hungary

²MTA-ELTE „Lendület” Lingual Articulation Research Group, Hungary

³Research Institute for Linguistics HAS, Hungary

⁴Budapest University of Technology and Economics, Hungary

{deme.andrea|marko.alexandra}@btk.elte.hu, bartokmarton@gmail.com,
graczi.tekla.etelka@nytud.mta.hu, csapot@tmit.bme.hu

Abstract

In the present study we analyzed vowel variation induced by carryover V-to-V coarticulation under the effect of pitch-accent as a function of vowel quality (using a minimally constrained intervening consonant to maximize V-to-V effects). We tested if /i/ is more resistant to coarticulation than /u/, and if both vowels show increased coarticulatory resistance in pitch-accented syllables. Our approach was unprecedented in the sense that it involved the analysis of parallel acoustic (F₂) and articulatory (*x*-axis dorsum position) data in a great number of speakers (9 speaker), and real words of Hungarian. To analyze the degree of coarticulation, we adopted the locus equation approach, and fitted linear models on vowel onset and midpoint data, and calculated the differences between coarticulated and non-coarticulated vowels in both domains. To measure variability, we calculated standard deviations of midpoint F₂ values and dorsum positions.

The results showed that accent clearly exerted an effect on the phonetic realization of vowels, but the effect we found was dependent on both the vowel quality, and the domain (articulation/acoustics) at hand. Observation of the patterns we found in parallel acoustic and articulatory data warrants for reconsideration of the term ‘coarticulatory resistance’, and how it should be conceptualized.

Index Terms: vowel variability, articulatory variability, acoustic variability, EMA, prominence, prosodic conditioning of variability, strengthening

1. Introduction

Coarticulation is one of the main sources of segmental variability. Since the seminal work of [1] it is recognized that not only adjacent speech sounds but also transconsonantal vowels have an effect on each other, and the vowels in V₁CV₂ sequences are claimed to be produced with one single underlying diphthongal gesture to which the consonant’s gesture is superimposed.

The extent a segment is susceptible to coarticulation, i.e., the contextual variability it exhibits, is referred to as **coarticulatory resistance** (CR; greater resistance = less variance) [2]. CR in V-to-V coarticulation may be influenced by several factors. In an acoustic study, in 5 speakers [3] demonstrated that vowels show smaller *acoustic* variability, if they are in a **lexically stressed** syllable (vs. unstressed). Further, in an articulatory study, in 6

speakers, [4] confirmed that vowels also show smaller *articulatory* variability (measured at the edge and the first quarter of the vowel) under **sentential** (i.e., higher level) **accent**. To the authors’ knowledge, however, no previous studies attempted to gather parallel acoustic and articulatory data, to address the question, if vowel variability observed in the two domains may be in parity, and show congruent tendencies (which is far from evident, due to the well-known non-linear relationship between articulation and acoustics).

Although inconclusively and in smaller samples, it was also shown that certain **vowel qualities** exhibit greater resistance than others: in Italian, in an articulatory study of 1 speaker, /i/ was found to be more resistant than /a/, and /a/ than /u/ [5]; in German, in an articulatory study of 3 speakers, /i/ was found to be more resistant than /a/ [6]; while in Thai, in an acoustic study of 6 speakers, the high /i/ and /u/ appeared similarly resistant [7] (and [7] claimed that the lower the vowel, the more susceptible it is to V-to-V coarticulation).

Lastly, in an acoustic study of 5 speakers’ nonsense read sequences, [8] demonstrated that **intervening consonants**, which exert a smaller degree of tongue dorsum contact with the palate allow for more V-to-V coarticulation, and he also corroborated that the reduction of the vowel is stronger, if it is in an unstressed syllable (that is, CR decreased in the absence of lexical stress). In its concluding remarks, [8] also pointed out that future work should clarify if these effects hold in more speakers, in real words, and for other languages.

In line with the above, the aim of the present study is (1) to further explore if prominence provokes CR, (2) to further clarify the effect of vowel quality on CR, and (3) to uncover the interrelations of acoustic and articulatory variability of vowels due to carryover V-to-V effects conditioned by pitch-accent. For this purpose, we analyzed V-to-V coarticulatory effects both in the acoustic and the articulatory domains, in real words, but in phonetically well-controlled contexts, in minimally constrained C-context, in the presence/absence of sentence-level accent (+ word stress co-varying with accent) in Hungarian, in a high number of speakers.

2. Methods

We recorded 9 Hungarian adult female speakers producing /uhu/C_{alv}/u/, /ihu/C_{alv}/u/, /ihi/C_{alv}/i/, and /uhi/C_{alv}/i/ in real Hungarian words embedded in sentences, three times (recorded in a randomized order). To minimize anticipatory coarticulatory effects from the ensuing segments, target vowels (on which the

carryover V-to-V coarticulatory effect of the transconsonantal vowel was analyzed; indicated by boldface in the examples above) were followed by an alveolar consonant, and a vowel which was in its quality identical with the target vowel. The target vowel was preceded by the glottal fricative /h/, as it is underspecified for oral configuration, and thus interferes the least with the single diphthongal gesture of the V₁ and V₂ vowel segments, and maximizes V-to-V effects [8]. This setting included two contexts, asymmetrical (asymm; /ihu/C_{alv}/u/ and /uhi/C_{alv}/i/), and symmetrical (symm; /uhu/C_{alv}/u/ and /ihi/C_{alv}/i/), which were expected to show and to not show carryover V-to-V coarticulatory effects, respectively. Furthermore, we created two accent conditions, 'V₁hV₂C_{alv}V₂ and V₁#hV₂C_{alv}V₂, where sentence-level (pitch-)accent and accompanying word stress fall either on V₁ or on V₂ (= the target), while the other vowel was unaccented (e.g., *A la.pu hu.szun.kat temetett maga alá* ‘the burdock covered twenty of us’; *Pu.hu.lunk* ‘we are getting weak’; where underline indicates the V₂ target vowel, and dots indicate syllable boundaries). After the exclusion of mispronounced tokens, we analyzed 212 items in total.

Parallel tongue movement and audio recordings were carried out in a sound-treated room using a Carstens AG501 EMA magnetometer system and a head-mounted omnidirectional condenser microphone. We tracked the movement of the upper and lower lips, the jaw, and the tongue, using 4 sensors on the tongue: tip, blade, and 2 on the tongue body (TBO1, TBO2, from tip to root) (see [9] for a similar sensor configuration).

2.1. Acoustic analysis

In the acoustic domain, we measured F₁ and F₂ of the target vowel at the left edge (median of first 10%; F_{2onset}) and in the temporal midpoint (median of middle 10%; F_{2mid}) in Praat [10] automatically, using the Burg algorithm. Building on the locus equation approach, to gauge the **degree of coarticulation**, first, we fitted linear models on F_{2mid} and F_{2onset} [2], as a function of context and condition, for each vowel separately. Then, we also calculated the difference of F_{2onset}s of coarticulated (asymm) and non-coarticulated (symm) instances (to get data comparable to [7], the articulatory data of [4], and the articulatory data of the present study). **Vowel variability** was quantified by the standard deviation of F_{2mid} values calculated for the 3 repetitions of the same token by the same speaker.

2.2. Articulatory analysis

Head movement and bite plane correction of the positional data were carried out by the Carstens software; data were centered around the incisors to get a coordinate system where the more negative the value, the more back the sensor is positioned. For 3D-2D conversion of position data (i.e., to obtain “midsagittal section”), and the production of Emu-compatible ssff tracks, we used the custom made converter of the IfL Phonetik, University of Cologne. Segmental labelling of the audio signal was carried out semi-automatically using the BAS web services G2P [11] and MAUS [12]. For data extraction, we used Emu [13].

First, we measured horizontal (x-axis) displacement of the TBO1 and TBO2 sensors in the left edge (median of first 10%) and in the temporal midpoint (median of middle 10%) of the target vowel. Then, we calculated *horizontal dorsum position* as the mean of TBO1 and TBO2 x-values for each token in the vowel onset and vowel midpoint, to be analyzed as the dependent variable, since the main interest here lies in V-to-V coarticulatory patterns observed in the overall tongue body configuration, rather than a single point on the tongue.

To parametrize the **degree of coarticulation**, we first fitted linear models on midpoint and onset dorsum positions as a function of context and condition, for each vowel separately, just as we did with acoustic data. Then, we also calculated the difference of the dorsum x-values in the coarticulated (asymm) and non-coarticulated (symm) instances, as measured in the vowel onset (similarly to the ‘distances’ in [4], and to the acoustic differences of these tokens established in the present paper). To quantify **vowel variability**, we calculated the standard deviation of the horizontal displacement of the tongue dorsum for the 3 repetitions of the same token by the same speaker, again similarly to the quantification of acoustic variation.

2.3. Statistical analysis of acoustic and articulatory data

Data were analyzed with linear mixed effects models in R [14], by using the lmerTest package and obtaining p-values by Satterthwaite-approximation [15]. Random slopes and intercepts were added to the models for speakers if they improved the performance of the model (assessed on the basis of AIC). Graphs display mean and corrected confidence intervals.

3. Results

3.1. Acoustic data

3.1.1. Degree of coarticulation

Locus equations showed steep slopes for /i/ and slopes of approx. 0 for /u/ in both conditions, reflecting the fact that /i/s were produced more stationary in time than /u/s, irrespective of the presence of accent (Figure 1).

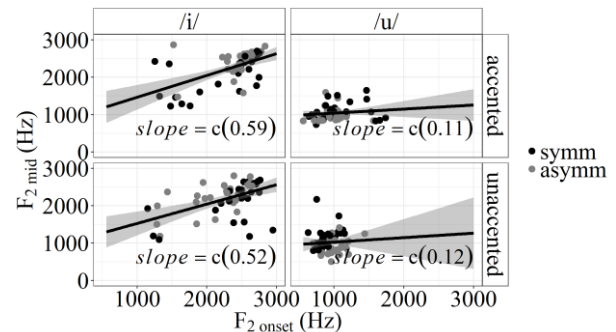


Figure 1: “Locus equations” for the target vowel in coarticulating (asymm) and non-coarticulating (symm) contexts, as a function of prominence.

F_{2onset} differences showed the effect of accent condition as a function of vowel quality (Figure 2) (vowel quality × condition interaction: [$F(1, 30) = 16.04, p < 0.01$]). Regarding /u/, F_{2onset} differences reflected that the onset of the vowels was not modified by coarticulation in the accented condition (recall that values around 0 mean no difference between /i(##)hu/ and /u(##)hu/ realizations), but in the unaccented condition, the F₂ of /u/s was “pulled up” by the preceding front /i/ in /i##hu/.

Regarding /i/, data surprisingly showed that in the accented position, tokens were realized as more “front”, i.e., more peripheral in the coarticulating /u##hi/ context than in the non-coarticulating /i##hi/ context (as positive values reflect that coarticulated tokens have a higher F₂ than non-coarticulated tokens). In unaccented position, however, F₂ of /i/s was lower in coarticulating positions (revealed by negative values in the graph), that is, coarticulated /i/s were more “back” acoustically

than non-coarticulated /i/s. In the unaccented condition, the effect of coarticulation was comparable in the two vowels.

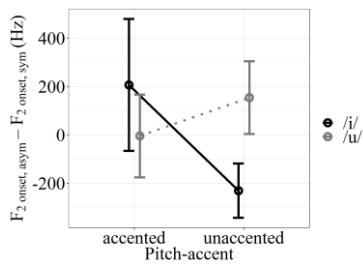


Figure 2: Differences of F_2 s in coarticulated (asymm) and non-coarticulated (symm) vowel onsets.

3.1.2. Vowel variability

The analysis of the SD of midpoint F_2 data revealed that /i/ tokens were more variable than /u/ tokens in the accented condition (especially in the non-coarticulating context, where /i/s showed the highest SD: 279 Hz on average), but in the unaccented, non-coarticulating context /u/s varied more than /i/s, and in the unaccented coarticulating context, /u/s varied less acoustically (and showed the lowest values of variability: 72 Hz on average) (Figure 3) [vowel \times context \times condition interaction effect: $F(1, 176) = 19.48, p < 0.01$]. As far as /i/ tokens are concerned, their variability was comparable in all contexts and conditions but the accented symmetrical condition.

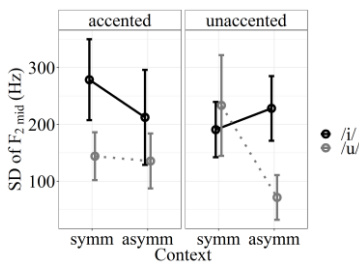


Figure 3: SD of F_2 measured in the vowel midpoint.

3.2. Articulatory data

3.2.1. Degree of coarticulation

Locus equations showed that x -axis dorsum positions in the onset predict very well the x -axis dorsum positions measured in the vowel midpoint, irrespective of vowel quality, the presence of pitch accent, or the V-to-V coarticulatory effect exerted by the preceding transconsonantal vowel (Figure 4).

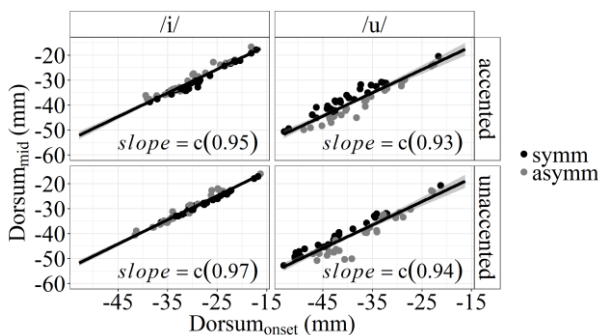


Figure 4: "Locus equations" for the target vowel in coarticulating (asymm) and non-coarticulating (symm) contexts, as a function of prominence.

Steep slopes of the linear models tell us that, as captured in the articulatory domain, vowels were realized much less dynamically than the acoustic data reflected, and that these stationary articulatory patterns were comparable in all vowels. It is also observable, however, that /u/ tokens in asymmetrical context are more likely to fall below the line of best fit (while /u/s in symmetrical context fall above), that is, the dorsum in the onset of the coarticulated /u/s was positioned more front than that of the symmetrical contexts, where /u/s were generally more stable through time. We observed no similar tendencies in /i/ realizations.

As expected, values of the differences of dorsum positions in coarticulated vs. non-coarticulated targets (measured at the onset of the target vowel) were generally below 0 for /i/, and above 0 for /u/. That is, due to coarticulation, the tongue dorsum in /i/ was, in general, more back, while in /u/ it was more front if vowels occurred in contexts that promoted V-to-V coarticulation (Figure 5). (Note that in these data negative values represent forward movement of the sensor, in other words, with respect to asymm to symm differences, centralization of coarticulated tokens is reflected by positive values for back, and negative values for front vowels.)

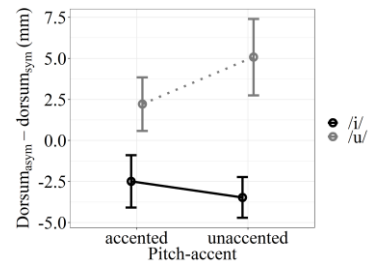


Figure 5: Differences of dorsum positions in coarticulated (asymm) and non-coarticulated (symm) vowel onsets.

Just as we have seen in the F_2 -differences, dorsum distances of coarticulated and non-coarticulated vowels showed a significant two-way interaction effect of condition and vowel quality [$F(1, 36) = 7.35, p < 0.05$]. In these articulatory data, the distances established in the accented condition were comparable in both vowels (2.22 and -2.49 mm on average), and increased in the unaccented condition (reflected by more negative values for front, and more positive values for back vowels). However, in the unaccented position, /u/ clearly showed a greater carryover V-to-V coarticulation effect than /i/, that is the distance between /ihu/ and /uhu/ ($M = 5.08$ mm) was greater than that of /uhi/ and /ihi/ ($M = -3.47$ mm).

3.2.2. Vowel variability

Lastly, as opposed to the acoustic data, SD of dorsum positions (Figure 6) showed that (in the articulatory domain) /u/ was in general more variable than /i/, especially in the accented syllables /u#'hu/ and /i#'hu/ [condition \times vowel interaction: $F(1, 207) = 7.06, p < 0.01$]. Moreover, variation in the articulatory target of /u/ was the highest, when /u/s were uttered in non-coarticulating accented /u#'hu/ syllables, which again, goes against the observation we made for acoustics. (Recall that in acoustics, we found a moderate amount of variation in /u/ in non-coarticulating accented /u#'hu/ syllables, and the highest amount of variation in non-coarticulating unaccented /uhu/ syllables.) Additionally, SD of dorsum data was the lowest in the unaccented non-coarticulating context (/V1hV1/: 1.95 mm), the highest in the accented non-coarticulating context (/V1#'hV1/: 2.83 mm) in general, and took an in between value

in both coarticulating contexts ($/V_1\#hV_2/$: 2.09 mm; $/V_1hV_2/$: 2.22 mm) [condition \times context interaction: $F(1, 207) = 7.38, p < 0.05$].

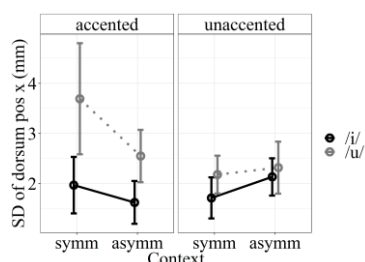


Figure 6: SD of the horizontal dorsum displacement.

4. Discussion

In the present paper, we presented parallel acoustic and articulatory data on V-to-V carryover coarticulatory effects of high front /i/ and high back /u/ in real words, and minimally constrained consonantal context, in 9 speakers of Hungarian. Our main aims were to observe the conditioning effect of pitch-accent, and vowel quality in carryover V-to-V coarticulation, that is, we aimed to reveal if sentence-level prominence provokes coarticulatory resistance, and decreases vowel variability, and if vowels show a difference in their resistance to coarticulation, as claimed before, but not supported abundantly by empirical evidence. An important novelty of our approach was to analyze the effect of coarticulation parallel in the acoustic and articulatory domains (obtaining flesh-point information via EMA). We attempted to capture both **the degree of coarticulation** and **the variability of vowels** resulting from this coarticulatory effect in the affected syllable by comparable measures obtained in both domains.

Most importantly, our results demonstrated the well-known fact that acoustic and articulatory data are not in a linear relationship, which also highlights the fact that articulatory and/or acoustic data alone may lead to fairly divergent results with respect to the effect of coarticulation and vowel variation. Nevertheless, our data in both domains unanimously showed that pitch-accent exerts its effect on the tested vowels. However, the effect itself was quite different as a function of vowel quality.

Acoustic data showed that, in general, /u/ was realized in a more dynamic fashion through time than /i/, but this difference did not persist so clearly in the articulatory domain (in which we saw more stationary realizations for both vowels). The distances of coarticulated and non-coarticulated vowel instances further revealed that the behavior of /u/ and /i/ also differed as a function of pitch-accent. While in the accented syllable /u/ targets were similarly “fronted/backed” irrespective of coarticulation, /i/ targets appeared hyperarticulated under coarticulation, as coarticulated /i/ instances were more “front” than non-coarticulated ones. However, this difference revealed itself only in the acoustics; in the articulatory domain, the distance of coarticulated and non-coarticulated vowels was highly similar (and showed some degree of centralization due to coarticulation) in both qualities. As expected, distances of coarticulated and non-coarticulated instances increased in the unaccented syllables in both vowels. However, as the acoustics showed a similar degree of centralization (approx. 200 Hz) for /i/ and /u/, the analysis of the articulatory data revealed that to yield the same amount of acoustic modification, the tongue was displaced more in /u/ than in /i/.

As for vowel variability resulting from the effect of V-to-V coarticulation is concerned, we found that acoustic and

articulatory data showed opposite tendencies: while the acoustics showed that /i/ was in general more variable than /u/, the analysis of dorsum positions revealed that the variability of the tongue position was in general higher in /u/ than in /i/, especially in the accented syllables. These data again reflect that the acoustic stability of /u/ targets results from articulator displacement of greater magnitude in the case of /u/ than in /i/.

On the basis of degree of coarticulation and vowel variation measures together, we can say that sentence-level accent provoked acoustic hyperarticulation in /i/ (achieved by the same displacement of the tongue between coarticulated and non-coarticulated instances than that in the case of /u/), and similar acoustic qualities in /u/ under coarticulation and no coarticulation. These tendencies were accompanied by greater acoustic variation in /i/ than in /u/ achieved by less variation in the articulator displacement in /i/ than in /u/. In unaccented syllables, we found that /u/ showed less variability in the coarticulating context, which points to the possibility of /u/ showing greater acoustic adaptation by reaching the modified (coarticulated) quality with more precision across repetitions than /i/.

Due to parallel data acquisition, our results are to some extent, difficult to compare with previous findings. Essentially, on the basis of the acoustic data we may claim that we corroborated previous proposals for the effect of pitch-accent on coarticulatory resistance [3, 4, 8], as we have seen that in certain respects, both vowels behaved more resistant in accented syllables. However, the phonetic implementation of this resistance was highly different as a function of vowel qualities and even domains. Not independently of the above, our data showing higher acoustic variability of /i/ than /u/ (especially in accented syllables) contradicted previous findings [5-7]. As a result, we believe that the interpretation of the present data as evidence of the conditioning effect of prosody and vowel quality on V-to-V induced CR may be misleading, and it rather appears to be motivated by finding phonetic correlates of CR, and not by answering the question if accent did really increase CR (as a function of vowel quality). Therefore, instead of proposing further claims on the effect of V-to-V coarticulation and coarticulatory resistance, we prefer to raise two questions as concluding remarks that, we believe, could lead to new insights on the topic.

1. How should we define coarticulatory resistance? Should it be conceptualized as decreased acoustic/articulatory variability of targets, and should we therefore claim the /i/ was found to be less resistant than /u/ in the present study? Or is it more like the increased capacity of dissimilation under coarticulatory effects, while decreased CR is the reduction of this capacity, that is, more adaptation and less variability in the adapted target, just as we have seen in the case of /u/ in unaccented conditions in the acoustic domain?

2. What is the domain of CR? Is it to be measured in the acoustic or the articulatory domains? Is it necessarily the acoustic output that constraints phonetic variability as suggested by some authors, or it is rather the motor domain, which allow less or more precision (and thus variation) in different regions of the articulatory space?

5. Acknowledgements

Help of Zsófia Weidl, Zsófia Puzder, Valéria Krepsz, Doris Mücke, Anne Hermes, and Theodor Klinker is gratefully acknowledged. The first author was supported by the Hungarian Ministry of Human Capacities (EMMI), Hungarian Talent Program (NTP).

6. References

- [1] S. E. G. Öhmanm “Coarticulation in VCV utterances: spectrographic measurements,” *Journal of the Acoustical Society of America*, vol. 39, no. 1, pp. 151–168, 1966.
- [2] H.-Y. Bang, “The acoustic counterpart to coarticulation resistance and aggressiveness in locus equation metrics and vowel dispersion,” *The Journal of the Acoustical Society of America*, vol. 141. no. 4, pp. EL345–EL350, 2017.
- [3] C. A. Fowler, “Production and perception of coarticulation among stressed and unstressed vowels,” *Journal of Speech and Hearing Research* vol. 24, no. 1, pp. 127–139, 1981.
- [4] T. Cho, “Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English,” *Journal of Phonetics*, vol. 32, no. 2, pp. 141–176, 2004.
- [5] E. Farnetani, K. Vaggés and E. Magno-Caldognetto, “Coarticulation in Italian /VtV/ sequences: a palatographic Study,” *Phonetica* vol. 42, no. 2-3, pp. 78–99, 1985.
- [6] A. Butcher and E. Weiher, “An electropalatographic investigation of coarticulation in VCV sequences,” *Journal of Phonetics* 4, pp. 59–74, 1976.
- [7] P. K. Mok, “Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation”, *Language and Speech*, vol. 54, no. 4, pp. 527–544, 2011.
- [8] D. Recasens, “The effect of stress and speech rate on vowel coarticulation in Catalan vowel–consonant–vowel sequences,” *Journal of Speech, Language, and Hearing Research* vol. 58, no. 5, pp. 1407–1424, 2015.
- [9] A. Deme, R. Greisbach, A. Markó, M. Meier, M. Bartók, J. Jankovics, Zs. Weidl, “Tongue and jaw movements in high-pitched soprano singing: A case study” *Beszédkutató* vol. 24, pp. 121–138, 2016.
- [10] B. Paul and D. Weenink, *Praat: doing phonetics by computer [Computer program]*. Version 5.4. <http://www.praat.org/>, 2014.
- [11] Reichel, U. D. 2012. PermA and Balloon: Tools for string alignment and text processing. Proc. Interspeech, paper: 346.
- [12] Schiel, F. 1999. Automatic phonetic transcription of nonprompted speech. Proc. Int. Cong. Phon. Sci 607–610.
- [13] Winkelmann, R., Klaus, J., Cassidy, S., Harrington, J. 2018. emuR: Main Package of the EMU Speech Database Management SystemR, package version 1.1.1.
- [14] R Core Team *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2018.
- [15] A. Kuznetsova, P. B. Brockhoff and R. H. Christensen, lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, vol. 82, pp. 1–26, 2017.