Acoustic and articulatory vowel variation as quality shift and increased variance in anticipatory and carryover vowel-to-vowel coarticulation

Andrea Deme[1,2], Márton Bartók[1,2], Tamás Gábor Csapó[2,3], Tekla Etelka Gráczi[2,4], Alexandra Markó[1,2]

[1]*Eötvös Loránd University, Budapest, Hungary*
[2]*MTA-ELTE "Lendület" Lingual Articulation Research Group, Budapest, Hungary*
[3]*Budapest University of Technology and Economics, Budapest, Hungary*
[4]*Research Institute for Linguistics, Budapest, Hungary*

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

Abstract

In this paper we studied if we find increased coarticulatory resistance and aggression in V-to-V coarticulation in pitch-accented syllables, and we also tested if anticipatory effects are exceeded by carryover effects in these contexts. We analyzed acoustic and articulatory (EMA) data from 9 Hungarian female speakers, and gauged coarticulatory effects by two different means. First, in line with previous studies, we calculated differences of coarticulated and neutrally positioned vowels (quality shift), which captures nature and magnitude of centralization. Second, we determined across context dispersion of vowels using relative standard deviation of tokens, which reflects uniformity of vocalic targets and which measure is yet understudied in this topic. Our two measures revealed different trends of coarticulatory variation and its modulating factors, and we also find divergent tendencies in acoustics and articulation. We propose that extension of our analysis to further typologically different languages is needed.

Keywords: vowel to vowel coarticulation, pitch-accent, direction of coarticulation

1. Introduction

Vowel-to-vowel (V-to-V) coarticulation refers to the phenomenon that realization of vowels is affected by a transconsonantal vowel in a sound sequence. This was first demonstrated by Öhman (1966), who claimed that vowels in vowel-consonant-vowel sequences are produced with one single underlying diphthongal gesture to which the consonant's target is superimposed.

It is hypothesized that V-to-V coarticulation induced contextual variation of vowels is dependent on several factors, for instance, prosodic position of the target and context (or trigger) vowels (e.g., Fowler 1984, Cho 2004; Deme *et al.* 2019), and the direction of coarticulation (e.g., Cho 2004; Mok 2011; 2012).

Prosodically strong locations, that is, lexical stress, pitch-accent, or the edge of the prosodic domain, are expected to condition articulatory "strengthening" which means an increased spatio-temporal magnitude of gestures. This, in turn is also expected to increase coarticulatory resistance and aggression, that is, less contextual variability, and more coarticulatory effect on the transconsonantal vowels coming from the accented vowel, respectively (Cho 2004).

Previous studies in non-words showed evidence for increased coarticulatory resistance with respect to V-to-V coarticulation in American English. They revealed that the *quality shift* of vowels (i.e., acoustic or articulatory differences of vowels in different contexts, e.g., in coarticulating and non-coarticulating, or in other words, neutral context) was smaller in lexically stressed syllables than in unstressed syllables in acoustics (Fowler 1981), just as it was smaller in pitch-accented syllables than in unaccented syllables in articulation (Cho 2004).

Recent studies in Hungarian real-words, however, revealed that distances of coarticulated and non-coarticulated (neutrally positioned) vowels showed vowel quality (/i/ vs. /u/) and production domain (articulatory vs. acoustics) dependent trends, as we found accent to reduce quality shift (i.e., reduced centralization) only in /i/ in acoustics, and only in /u/ in articulation (Deme *et al.* 2019). Furthermore, in the cited study we also introduced a second measure of coarticulatory resistance, *across context dispersion* (to which we return below in more detail), and this measure was not shown to be conditioned by pitch-accent.

Several studies demonstrated that coarticulatory aggression and resistance are the "two sides of the same coin", at least in consonant-vowel coarticulation (see e.g., Recasens & Rodríguez 2016). In vowel-to-vowel coarticulation, however, no evidence of increased coarticulatory aggression was found for vowels in pitch-accented syllables, as determined on the basis of lingual displacement (Cho 2004).

As for the direction of coarticulation, effects of carryover coarticulation were found to be stronger than anticipatory in V-to-V coarticulation both in articulation and acoustics, for several vowels: it was demonstrated in /i/ and /a/ in articulation in American English (Cho 2004), and in open /æ/ (Mok 2011) and /i u a/ (Mok 2012) in acoustics in Thai, Cantonese, and Mandarin.

Note that most of the above studies captured V-to-V induced coarticulatory resistance or "contextual variation", in quality differences of coarticulated and non-coarticulated tokens. Contextual variability is, however, very often interpreted and visualized by dispersion of vowel tokens in the acoustic and articulatory spaces (dispersion ellipses), especially in studies focusing on within-category vowel dispersion specific to a given language (see e.g., Manuel 1990; Mok 2012: 194), or in other words, "phoneme size" (Mok 2012: 194). Characteristic vowel dispersion patterns can be calculated using vowel realizations from different contexts. And this *across context dispersion* is not yet well explored under the conditioning effect of the above factors, while, as mentioned, our previous study showed that it may exhibit fundamentally different trends from those observable in *quality shift* data (Deme *et al.* 2019).

Furthermore, limited amount of available results (see Deme et al. 2019) warrants for further exploration of the question if coarticulatory resistance and aggression of the same tokens in the same utterances is detectable both in acoustics and in articulation in V-to-V coarticulation.

It is important to note here, that according to a well-known (but in some respect, understudied) hypothesis we already briefly mentioned above, variability or dispersion of V realizations (the above mentioned "phoneme size") is also affected by the density of the vowel space (Manuel 1990). Therefore, it is safe to assume that none of the effects claimed to influence V-to-V induced variability in English, generalize automatically across languages.

The three questions of our present study are the following.

1. Does V-to-V induced vocalic variation depend on the direction of coarticulation?
2. Are V-to-V effects influenced by prosodic position of the target vowel?
3. Does prosodic strengthening of the trigger vowel have an effect on variation in the target vowel?

And we investigate these questions

- using the highest level of prosodic prominence, pitch-accent,
- both in dispersion (that is, across-context variance), and quality shift (that is, distances) of vowels,
- in both domains of production, namely, articulation and acoustics, and
- in Hungarian, an obligatory syntactic focus marking non-Germanic (but Finno-Ugric) language with fixed word stress.

We hypothesize that vowels under the effect of V-to-V coarticulation show increased coarticulatory aggression and resistance in pitch-accented syllables, both in acoustics and in articulation. We also expect that the effect of carryover coarticulation exceeds that of anticipatory in these contexts.

2. Methods

For the purposes of the present study, we analyzed synchronously collected acoustic and EMA data from 9 adult female speakers of Hungarian (aged 25.2±5.9 years).

We analyzed the /i u/ point vowels both as target and trigger vowels. Vowels were placed in the context of the minimally constrained labial stop /p/ in nonsense sequences (**Table 1**) similarly to Cho (2004) and Mok (2011; 2012). We recorded minimally 6 repetitions of each token per speaker. We varied the position of pitch accent (and had target and trigger vowels in pitch-accented and unaccented syllables), and created coarticulating (e.g., /p<u>u</u>p**i**pipi/) and non-coarticulating (e.g., /p<u>i</u>p**i**pipi/) contexts (bold: target V; underline: trigger V). Since Hungarian has a fixed first syllable stress, however, not every combination of factors was possible.

In articulatory data processing, head movement and bite plane corrections were done by the Carstens software, while further post-processing (3D-2D conversion, and production of Emu-compatible ssff tracks) was carried out by 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 (Reichel 2012) and MAUS (Schiel 1999). For gestural labelling we used Emu (Winkelmann *et al.* 2018).

As we analyzed the /i u/ point vowels, in accordance with previous studies (e.g., Fowler 1981; Cho 2004), we obtained second formant values at the onset, offset, and temporal midpoint of target vowels, and measured and averaged the horizontal position of the backmost two tongue body sensors as "dorsum" data. Position data were normalized to the maximum and minimum *x*-axis displacement of the given sensor for each speaker using the sensor positions of the backmost /u/ variant (0%), and the most fronted /i/ variant (100%) in each case (**Figure 1**).

Table 1: *Stimuli and factors of the study (antic. = anticipatory; carryo. = carryover; targ. = target; acc = accented; unacc = unaccented).*

		Context /i/		Context /u/	
		acc	*unacc*	*acc*	*unacc*
targ. /i/ + *unacc*	*antic.*	–	ˈpip**i**p<u>i</u>pi	–	ˈpip**i**p<u>u</u>pu
	carryo.	ˈp<u>i</u>p**i**pipi	ˈpip<u>i</u>p**i**pi	ˈp<u>u</u>p**i**pipi	ˈpup<u>u</u>p**i**pi
targ. /i/ + *acc*	*antic.*	–	ˈp**i**p<u>i</u>pipi	–	ˈp**i**p<u>u</u>pupu
	carryo.	–	–	–	–
targ. /u/ + *unacc*	*antic.*	–	ˈpup**u**p<u>i</u>pi	–	ˈpup**u**p<u>u</u>pu
	carryo.	ˈp<u>i</u>p**u**pupu	ˈpip<u>i</u>p**u**pu	ˈp<u>u</u>p**u**pupu	ˈpup<u>u</u>p**u**pu
targ. /u/ + *acc*	*antic.*	–	ˈp**u**p<u>i</u>pipi	–	ˈp**u**p<u>u</u>pupu
	carryo.	–	–	–	–

To quantify variability, first we calculated distances of coarticulated and non-coarticulated tokens measured at the vowel edge which was located closer to the trigger vowel, in accented and unaccented syllables respectively (e.g., F_2 p**u**p<u>i</u>pupu – F_2 p**u**p<u>u</u>pupu vs. F_2 pup**u**p<u>i</u>pu – F_2 pup**u**p<u>u</u>pu, and Dorsum p**u**p<u>i</u>pupu – Dorsum p**u**p<u>u</u>pupu vs. Dorsum pup**u**p<u>i</u>pu – Dorsum pup**u**p<u>u</u>pu). This is what we refer to as *distances* data, where the greater the value, the greater the difference is between target vowel qualities. Second, we also calculated relative standard deviation for vowel midpoint data across contexts (e.g., pooled RSD of p**u**p<u>u</u>pupu and p**u**p<u>i</u>pupu), which is referred to as *across context dispersion*. Here, the greater the value the greater the variability is in realization of a given vocalic target.

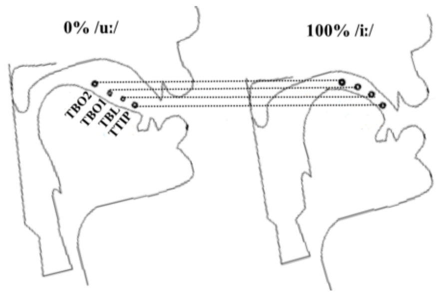

Figure 1: *References used for sensor position normalization (/u/ = 0%), /i/ = 100%), and the scale defined by these extreme positions on the x axis (based on Cho 2004).*

Data were analyzed with linear mixed effects models in R (R Core Team 2018), by using the lmerTest package and obtaining *p*-values by Satterthwaite-approximation (Kuznetsova *et al.* 2017). 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). Post hoc analyses (Tukey tests) were carried out by the lsmeans package (Lenth 2016). Graphs display mean and corrected confidence intervals.

3. Results

Results showed divergent tendencies for the two measures we obtained, and in some cases, trends also differed as a function of the production domain (i.e., in articulation, and in acoustics).

As mentioned previously, not all combinations of factors were present in our dataset. Therefore, we tested our three hypotheses in three different subsets of data. Subsets and their respective position on each figure are the following:

- direction hypothesis: unaccented trigger & unaccented target vowels (upper right and lower right panels, right hand side);
- resistance hypothesis: accented, and unaccented target vowels and unaccented trigger vowels (upper right panel);
- aggression hypothesis: unaccented target vowels, accented and unaccented trigger vowels (bottom panels, right hand side).

3.1. Distances of coarticulated and non-coarticulated vowel tokens

We start with the *distances* data (**Figure 2**). Note that here centralization is represented by negative values for /i/, and positive values for /u/ both in acoustics and in articulation.

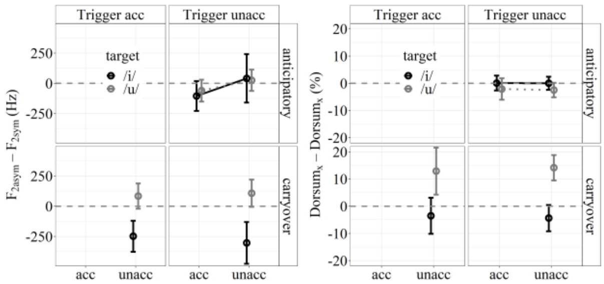

Figure 2: *Acoustic (left) and articulatory (right)* distances *of coarticulated and non-coarticulated vowel tokens as determined on the basis of vowel midpoint data.*

In acoustics, we generally found greater centralization in /i/ than in /u/ (i.e., greater absolute values in distances in /i/ than in /u/). We found a significant interaction effect of direction and vowel quality ($F(1, 37) = 9.89$, $p < .05$), as effects of carryover coarticulation exceeded that of anticipatory, but only in /i/. Data also revealed increased resistance in accented syllables (accent main effect; $F(1, 28) = 4.22$, $p < .05$), as /i/ was more centralized if accented, while /u/ was more peripheral. Increased aggression of accented targets was, however, not evidenced.

In articulation, in general, /u/ showed greater centralization than /i/. Here, anticipatory effects were again exceeded by carryover effects, but only in /u/ ($F(1, 25)$ 42.32 = , $p < .01$). Lastly, in articulatory distances, no increased resistance and aggression were found.

3.2. Across context dispersion of vowels

Let us now turn to the across context dispersion of vowels (**Figure 3**).

In this measure, in general, /u/ showed greater acoustic and articulatory variability. Interestingly, here in acoustics (**Figure 3**, left) we found neither a significant direction effect, nor increased coarticulatory resistance in accented syllables. However, pitch-accent on the trigger vowels turned out to be inducing greater across-context variability in unaccented /u/ ($F(1, 439) = 21.99$, $p < .01$), which reflects increased coarticulatory aggression in /i/ under the effect of pitch accent.

Lastly, across context articulatory dispersion of vowels (**Figure 3**, right) showed evidence only for the resistance hypothesis, as /u/ targets were more resistant if they were produced in pitch-accented syllables than in unaccented syllables ($F(1, 390) = 8.12$, $p < .01$).

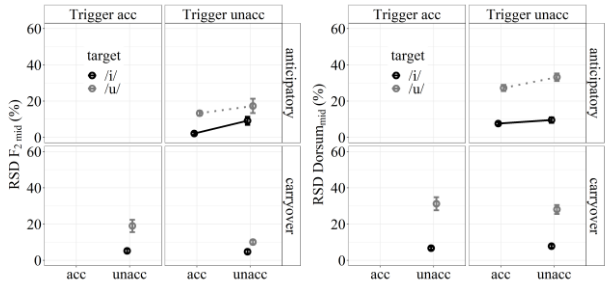

Figure 3: *Acoustic (left) and articulatory (right)* across context dispersion *of vowel tokens as determined on the basis of vowel midpoint data.*

4. Discussion and conclusion

In this study we analyzed if vowels under the effect of V-to-V coarticulation show increased coarticulatory aggression (AH) and resistance (RH) in pitch-accented syllables, both in acoustics and in articulation (**Table 2**). We further tested if the effect of carryover coarticulation exceeds that of anticipatory (DH) in V-to-V coarticulation. We gauged the effect of coarticulation by two separate metrics: quality shift or distances (distances of coarticulated and neutrally positioned tokens) and across context target variability (dispersion). For ease of comprehension, we summarized our results with respect to the above three hypotheses and two measures in **Table 2**.

Table 2: *Summary of hypothesis testing with respect to the two obtained metrics and the two domains of production (DH = direction hypothesis; RH = resistance hypothesis; AH = aggression hypothesis; ✓ = corroborated; ✗ = not corroborated; →← = trends opposing to expectations).*

		DH	RH	AH
Distances	*Acoustics*	✓/i/	→← /i/ ✓/u/	✗
	Articulation	✓/u/	✗	✗
Dispersion	*Acoustics*	✗	✗	✓/i/
	Articulation	✗	✓/u/	✗

To sum up, we found divergent tendencies in the two measures we obtained, and the two domains of production, but also as a function of vowel quality. On the basis of our findings, we can conclude that we replicated results not showing increased coarticulatory aggression in quality shift (Cho 2004) (while we found increased aggression in /i/ in dispersion), and we found less clear effect of accent on coarticulatory resistance than suggested previously for American English (see Fowler 1981, Cho 2004).

This latter discrepancy of results may possibly be explained by several factors. Probably one of the most important of these is the fact that, as opposed to English, Hungarian is an obligatory syntactic focus marking language, for which reason, prosodic means are suggested to play a limited role in expressing prominence of a constituent (Mády & Kleber 2010).

Furthermore, it may as well be the case that /i/ and /u/ are simply not exerting the same type of effects or same magnitude of effects on each other as /i/ and /a/ (see in Cho 2004), and on this basis, we did not find the claimed effects showing up in our data.

However, it is also important to note that due to the prosodic characteristics of the Hungarian language (i.e., fixed first syllable stress), we tested all three hypotheses in a very controlled fashion: i) resistance hypothesis was tested only in cases were trigger vowels were always unaccented, ii) aggression hypothesis was tested only in contexts, where target vowels were always unaccented, and iii) direction effects were tested using unaccented target and trigger vowels exclusively. This design goes against most previous studies where either one or the other of the target and trigger vowels were accented, but in this way, and in some sense, it also helps to disentangle prosodic effects on the target and the trigger vowels more, than previous attempts.

In our design, increased resistance in accented syllables was tested in cases where trigger vowels were "weaker" (according to the coarticulatory aggression hypothesis). Therefore, the role accent might play in modulating coarticulatory resistance was tested without the confounding effect of the accent position of the trigger being also manipulated, but we could only observe the behavior of an allegedly "weaker" trigger. Similarly, as accent effects on the trigger were tested only with unaccented, i.e., "weaker" targets (according to the coarticulatory resistance hypothesis), accent effects on the trigger vowel were completely disentangled from accent position of the target, but in a context where maximum effects can be expected. And finally, we need to keep in mind that direction effects were also tested here using "weaker" segments (in both relevant positions) exclusively.

An interesting finding of the present study is that the two measures we obtained provided complementary pictures to each other. *Across context dispersion* (that reflects the uniformity of vocalic targets) seemed to show accent effects, as it revealed increased coarticulatory resistance and aggression tendencies not seen in distances data, but we found no difference as a function of direction of coarticulation. *Distances data,* however, (that gauges the magnitude and character of coarticulation induced quality hift) revealed direction effects not seen in dispersion data, and some of the expected resistance effects in acoustics. On the basis of dispersion data, we can conclude that accent affects by all means should be taken into consideration in studies where phonemic size or vowel space density are investigated. While distances data indicates that studies concerned with V-to-V coarticulation induced centralization should take direction effects into consideration and acknowledge (or probably even exploit) the very likely fact, that in V-to-V coarticulation induced centralization tendencies carryover effects exceed those of anticipatory.

We suggest that the dispersion measure should be tested in future cross-linguistic studies involving prosodic focus marking languages (for instance, English and German), to help us further clarify what resistance is, and how we should measure it. Is it contextual invariance, or rather it is quality shift (i.e., some measure of centralization)?

5. Acknowledgements

The authors acknowledge the kind help of Zsófia Weidl, Zsófia Puzder, Valéria Krepsz, Ákos Buza, Anne Hermes, Doris Mücke, Theo Klinker. The research was funded by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences, the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund, the Thematic Excellence Program of the Ministry for Innovation and Technology, and the TKA-DAAD.

6. References

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics,* 32(2), 141-176.

Deme, A., Bartók, M., Gráczi, T. E., Csapó, T. & Markó, A. (2019). V-to-V coarticulation induced acoustic and articulatory variability of vowels: The effect of pitch-accent. In: *Interspeech 2019*: 3317-3321.

Fowler, C. S. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research,* 24(1), 127-139.

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

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software,* 69, 1-33.

Mády, K., & Kleber, F. (2010). Variation of pitch accent patterns in Hungarian. *Proceedings of Speech Prosody 2010*. http://speechprosody2010.illinois.edu/papers/100924.pdf

Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America,* 88(3), 1286-1298.

Mok, P. K. (2011). Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation. *Language and Speech,* 54(4), 527-544.

Mok P. K. (2012). Does Vowel Inventory Density Affect Vowel-to-Vowel Coarticulation? *Language and Speech,* 56(2), 191-209.

Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America,* 39(1), 151-168.

R Core Team R (2018). *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Recasens, D. & Rodríguez, C. (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics,* 59(C), 58-75.

Reichel, U. (2012). PermA and Balloon: Tools for string alignment and text processing. In *Interspeech, 13th Annual Conference of the International Speech Communication Association,* 9.-13. September 2012, Portland, USA. Paper no. 346.

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In *Proceedings of International Congress of Phonetic Sciences,* 607–610.

Winkelmann, R., Klaus, J., Cassidy, S., & Harrington, J. (2018). *emuR: Main Package of the EMU Speech Database Management SystemR,* package version 1.1.1.