Saarland University

Department of Language Science and Technology

Bachelor's Thesis

Are /ə/ and /ɨ/ Merged in Aromanian?

Examining Their Acoustic Quality Across Indexical, Syntactic and Phonetic Parameters

Supervisors:

Prof. Dr. Bernd Möbius
Prof. Dr. Bistra Andreeva

Michael Vrazitulis (2570166)
michaelv@lst.uni-saarland.de

Date of Submission: December 29, 2020

Declaration

I hereby confirm that the thesis presented here is my own work, with all assistance acknowledged. I assure that the electronic version is identical in content to the printed version of the Bachelor's thesis.

Saarbrücken, December 29, 2020

Signature:

M. Vrazitulis

Abstract

The dialect spectrum of the Aromanian language consists of two broader dialect regions. It is commonly assumed that in the Southern dialect region, a vowel contrast between /ə/ and /ɨ/ is maintained, whereas in the Northern one, the two vowels have merged. On the basis of a speech corpus featuring native speakers from both dialect regions (South: Turia, Kutsufliani; North: Andon Poçi, Gjirokastër, Stjar), this study provides a quantitative account of the Aromanian vowel system. The results indicate that Northern speakers, indeed, tend not to display such a vowel contrast. Instead, they produce a single vowel phoneme with an allophonic range from [ɛ] to [ə], with stressed realizations being exclusively distributed around an open-mid front quality. Several Southern speakers, on the other hand, can be shown to maintain the expected contrast, but only between distributions of stressed /ə/ and /ɨ/ tokens. After accounting for other context effects, however, the observed contrast appears to be less robust. It can more accurately be described as resulting from an occasional allophonic variation of, again, a single underlying vowel phoneme.

Contents

1 Introduction

Studying the phonetic properties of a minority language can raise some special challenges. In particular when it comes to languages that are endangered, sparsely researched and do not have an agreed-upon standard variety or written form, it might be difficult to acquire a reliable data basis upon which to conduct a quantitative analysis. Nevertheless, exploring a language which meets the criteria that were just mentioned, can also lead to meaningful insights with regard to the robustness of phonological categories. If a language exists only in spoken form for the majority of its speakers, one can assume that their underlying mental representations of phonological categories are not influenced, distorted or held in place in any way by a bias resulting from the exposure to and proficiency in the usage of, for example, a standardized alphabet motivated by phonemic distinctions.

Aromanian is a Romance minority language spoken in the Southern Balkans. Its speaker community is mainly spread among the countries Greece (the regions Epirus, Macedonia and Thessaly), Albania (the South), North Macedonia and Bulgaria. There is some debate on whether Aromanian should be considered a language of its own or rather a dialect of Romanian. The latter view considers Aromanian to be one of the four major dialects of Romanian, alongside Dacoromanian, the variety which forms the basis of the official standard adopted in Romania and Moldova, as well as the two even more marginalized varieties Istroromanian and Meglenoromanian. On the other hand, the former view insists that those four varieties represent four separate languages within the Eastern-Romance subbranch of the Romance language family. It is not within the aim of this study to take a definitive stance on the dialect-or-language question regarding Aromanian. But considering that both Aromanian and Dacoromanian each consist of several subvarieties (dialects) of their own, for the sake of simplicity, it is convenient to speak of the *Aromanian language* and the *(Daco-)Romanian language* and their respective *dialects* or *dialect groups*. That is also the terminology which is going to be used in this text. According to the Endangered Languages Project (see Austin, 2009), it is estimated that today there are roughly 350,000 native speakers of Aromanian left. Aromanian is assigned a Language Endangerment Index level of 2 (Threatened); the level scale ranges from 0 (Safe) up to 5 (Critically Endangered).

1.1 Research Question

The goal of this study is to examine the Aromanian vowel system through an acoustic-phonetic analysis, with a particular focus on the pair of vowel phonemes /ə/ and /ɨ/. They are of special interest as it is actually quite unclear whether they should indeed be considered as two separate phonemes, or rather as merged together, thus forming a single vowel phoneme, which could be denoted as $/ə/_{\text{single}}$

for clarity. With regard to that question, opinions in existing literature tend to vary, as will be revealed in more detail in Section 2. Common descriptions of the phonetic/phonological properties of Aromanian differentiate between two major dialect groups: The Southern varieties, spoken mainly in Greece, are typically considered—just like Dacoromanian—to maintain an /ə/-versus-/ɨ/ distinction, whereas the Northern (or Farsherot) varieties, spoken mainly in Albania and North Macedonia, are assumed to employ a merged /ə/$_{\text{single}}$ variant (see Schlösser, 1985).

However, it seems that there have been no attempts to verify those assumptions by quantifying the acoustic quality of Aromanian /ə/ and /ɨ/ realizations. If it could be shown that their distributions vastly overlap in the speech of some speaker, it would be reasonable to conclude that also his or her underlying phonological representations do not differentiate between /ə/ and /ɨ/. On the other hand, if one were to find the distributions to be distinct from each other, one could interpret that as indicating the existence of two different representations for /ə/ and /ɨ/. However, in the latter case, it might still be necessary to check for potentially confounding factors like stress, coarticulation and duration, among others. Factors like these are known to have an influence on the acoustic quality of phonetic segments.

Then again, within the framework of phonological theory, speaking of two different vowel phonemes would also require finding minimal pairs which only differ with respect to the two vowels in question. And indeed, Beis and Dasoulas (2017) have proposed that such minimal pairs for the contrast /ə/ versus /ɨ/ do exist in the Southern-Aromanian varieties. A short discussion of their validity will also follow in Section 2.

Formant frequencies of vowel realizations represent a reliable quantitative description of their acoustic quality. On the basis of an audio corpus featuring fourteen different native Aromanian speakers (seven from each dialect region; the recordings were conducted between 1999 and 2015), a dataset containing formant frequencies of 11,756 vowel realizations was obtained. 1,603 of them can be attributed to the *vowels of interest* /ə/ and /ɨ/. The dataset also contains information on speaker's sex, coarticulatory context and vowel duration. Additional information on word stress as well as part-of-speech tags for the words containing the examined vowel tokens were added to the subdataset only encompassing such vowels of interest. By applying methods of multivariate statistical analysis to the data, it will be possible to make an informed judgement on whether and under which circumstances /ə/ and /ɨ/ should be considered as merged or as distinct. Those circumstances are further determined by the various types of indexical (speaker's sex and dialect region), syntactic (part-of-speech tags) and phonetic (duration, stress, coarticulatory context) parameters that are taken into account. Furthermore, an exploration of the complete vowel dataset will reveal some general properties of Aromanian's vowel space.

The details of the whole data preparation and analysis process are going to be laid out in Section 3. Results will be presented in Section 4, and finally, a conclusive discussion of the observed effects will be carried out in Section 5.

2 Literature Review

An overview of the scientific literature consulted as theoretical background for this study is going to be provided in this section. This includes publications with a primary or secondary focus on the phonetic/phonological properties of Aromanian as well as sources providing general guidance on acoustic-phonetic methodology.

2.1 Aromanian Phonetics and Phonology

The first scientific account of Aromanian phonology dates back to as early as 1888. In part one of his dissertation, Gustav Weigand, a German linguist, balkanologist and ethnographer, pioneered in describing the Aromanian sound system in much detail (see Weigand, 1888). His description is based on materials and observations he gathered during field studies in the then-Ottoman Empire, in the region around Mount Olympus more specifically. Notably, that region lies within the geographical boundaries of the Southern-Aromanian varieties. For the assumed vowels of interest /ə/ and /ɨ/, Weigand uses the notations ⟨ă⟩ and ⟨â⟩ which are inspired by Romanian orthography, most likely due to the fact that, in his time, phonetic theory and notation were not yet as fleshed-out and standardized as they are known today. What is very interesting to note here is that Weigand claims to be able to define the Aromanian /ɨ/ sound, in rough articulatory terms, as an 'a' sound produced with the 'larynx significantly pressed down'. Weigand goes on to contrast Aromanian /ɨ/ both from its Dacoromanian counterpart (which is also commonly described as a close central unrounded vowel, i.e. [ɨ]) as well as the close back unrounded vowel found in Turkish, i.e. [ɯ]. In analogy to, and in the same terms as for his articulatory outline of Aromanian /ɨ/ production, Weigand states that Dacoromanian /ɨ/ is derived from a 'u' sound, again by pressing down the larynx, whereas Turkish [ɯ] is allegedly derived from an 'i' sound in the same way.

Even though it remains a bit unclear how exactly those somewhat imprecise descriptions of the respective auditory qualities of the three vowels from Aromanian, Dacoromanian and Turkish would translate into modern-day phonetic/phonological terms, it is still striking to see Weigand pointing to a difference in quality between Aromanian /ɨ/ and Dacoromanian /ɨ/ in particular. Weigand was an expert and quite proficient in both of those languages, which arguably did enable him to make a qualified auditory judgement like that. This could be valuable diachronic insight with respect to the research question, suggesting that the common notion of an exact equivalence between the Southern-Aromanian and the Dacoromanian /ə/-versus-/ɨ/ contrast might not have been very accurate even for speakers who lived back by the end of the 19th century. On the other hand, Weigand still maintains that some contrast between the vowel phonemes /ə/ and /ɨ/ did exist for the Aromanian variety that he examined. A highly speculative interpretation of those observations could be to view the speaker population explored by Weigand in his early dissertation study in the state of a merger-in-progress: Assuming that it turned out to be the case that more recent distributions of the two Southern-Aromanian vowels of interest tended to overlap one another (if that might or might not the case shall be examined by this study), this could be seen as the result of a vowel merger affecting two originally distinct sounds /ə/ and /ɨ/. In that case, Weigand's auditory observation of Southern-Aromanian /ɨ/ not sounding exactly the same as its counterpart found in the related language Dacoromanian, might indicate that, at some point before 1888, the Aromanian distribution of /ɨ/ realizations had started to move away from its original mean and to gradually approach the /ə/ distribution. If the corresponding Dacoromanian distributions, by contrast, had stayed more or less in place during the same period of time, this would then explain why Weigand noticed a qualitative difference between what could have been a 'genuine' close back unrounded vowel in Dacoromanian and a Southern-Aromanian counterpart slightly shifted towards a more central average realization due to the effects of that merger-in-progress. But again, this interpretation would rely on a lot of speculative assumptions and should thus, in any case, only be considered with great caution.

In addition to that, Weigand also claims to have noticed that /ə/ typically had a shorter duration than /ɨ/ in the speech of the individuals he interacted with. One other important aspect brought up by him with regard to the vowels of interest, is the following historical sound-change pattern: (Vulgar) Latin words that used to contain a stressed [a] vowel followed by a nasal have mostly been transformed in their development towards their Balkan-Romance forms in such a way that the [a] vowel has been reduced to either [ə] or [ɨ]. These are some of the examples from Aromanian that Weigand uses to highlight this phenomenon:

- Lat. *pānis* (bread) > Ar. [ˈpɨne] (bread)
- Lat. *quanti* (how many) > Ar. [ˈkɨt͡s] (how many)
- Lat. *nesciō quantae* (I don't know how many) > Ar. [nəsˈkənte] (several)

It is quite remarkable that no further explanation is given on whether and how one could distinguish cases where the Latin prenasal [a] reduces to [ɨ] from ones where it reduces to [ə], indicating that there might not even be such a systematic relationship. This perspective is further developed by Schlösser (1985), almost a hundred years later: Schlösser deals with the Aromanian sound system and its historical evolution.

His informants were native Aromanian speakers from the Greek town of Metsovo, so again from the Southern dialect region. Among other aspects, Schlösser also comes back to the phenomenon that was just mentioned, and he extends the list of examples by some more; three of them are:

- Lat. *campum* (field, *acc.*) > Ar. [ˈkəmbu] (field)
- Lat. *fontānam* (source, *acc.*) > Ar. [funˈdənə] (source)
- Lat. *grandinem* (hail, *acc.*) > Ar. [ˈgrɨndine] (hail)

Schlösser goes on to state that whether any of the Latin words affected by the particular sound-change pattern would transform its prenasal [a] to [ə] or to [ɨ] is consistent in Dacoromanian (always [ɨ]), but rather arbitrary in Aromanian. According to him, this gives reason to believe that—at least for the specific Southern-Aromanian variety spoken in Metsovo—[ə] and [ɨ] are not phonologically relevant, but are better described as allophones of one and the same underlying phoneme. This is a crucial statement in that Schlösser's publication seems to be one of very few that directly question or reject the existence of two distinct phonemes /ə/ and /ɨ/ even in some Southern-Aromanian dialectal variety. Schlösser contrasts his conclusion with the classification criteria of Aromanian dialects presented in an earlier work by Caragiu-Marioțeanu (1975; as cited in Schlösser, 1985), who defines two ideal types of Aromanian dialects, 'type A' and 'type F', corresponding to the Southern and the Northern dialect group respectively. According to her, there are certain distinct characteristics associated with either dialect type that can be used to determine which dialect group any Aromanian variety belongs to. One of the listed characteristics is the existence of two distinct phonemes /ə/ and /ɨ/ in type A, but the absence of a phoneme /ɨ/ in type F.

The issue raised by Schlösser points back to what was established earlier in the introduction, namely that claiming that any two sounds in a given language constitute a phoneme requires not only to show that they are separate phones, i.e. that their respective acoustic qualities substantially differ from each other on average. It also requires demonstrating that their acoustic contrast is semantically relevant for at least one word pair, ideally more. Beis and Dasoulas (2017) propose to have found such a contrast for the following two Aromanian word pairs:

1. /aˈrɨu/ (river) versus /aˈrəu/ (bad)
2. /aˈrɨd͡zɨ/ (you laugh) versus /aˈrəd͡zɨ/ (rows)

Now, this definitely seems to be a compelling argument. It should be taken into account, though, that both of the proposed minimal pairs are heterogeneous with respect to word class. That is, the noun /aˈrɨu/ is contrasted against the adjective /aˈrəu/, and the finite verb form /aˈrɨd͡zɨ/ against the noun /aˈrəd͡zɨ/ (plural form of

/aˈraðə/). Neither of the word pairs is actually in great danger of creating an ambiguity within a full linguistic context. Adjectives only rarely take on the syntactic role that nouns usually have as heads of nominal phrases. And it is nearly impossible, within a given sentence, to replace a finite verb form by a noun without violating some grammatical requirement. So not only by the semantic context, but already by the syntactic one, it seems to be possible to tell the elements of both alleged Southern-Aromanian minimal pairs apart quite reliably. The four underlying words could thus as easily be classified as constituting two pairs of homophones:

1. /aˈrəu/$_{\text{single}}$ (river / bad)
2. /aˈrəd͜zə/$_{\text{single}}$ (you laugh / rows)

There are few publications dealing with the Aromanian sound system from a phonetic rather than phonological perspective. While Turculeț (2010) provides a quantitative examination of interrogative intonation patterns in the speech of native Aromanian speakers living in North Macedonia, Kharlamova (2017) investigates properties of affricates in Aromanian spontaneous speech under the aspects of stop-phase loss in dental affricates, the affrication of dental stops before front-row vowels and the speakers' bilingualism. In a more recent paper by the same author (see Kharlamova, 2019), the phonetic properties of an Aromanian variety spoken in the town of Selenica, located in Albania, are discussed. In that latter paper, Kharlamova briefly mentions the fact that Capidan (1932; as cited in Kharlamova, 2019) was historically the first author to propose a general dialectal distinction between Southern and Northern varieties of Aromanian. She also refers to a more recent classification of the distinctive features of each Aromanian dialect group, proposed by Namurov (2001; as cited in Kharlamova, 2019): That classification appears to be a further developed version of the distinction between features attributed to dialects of 'type A' and 'type F' by Caragiu-Marioțeanu, mentioned further above. Among other aspects, Namurov's classification, again, includes a statement that in Southern dialects, a contrast between /ə/ and /ɨ/ is present, whereas in Northern ones, such a contrast is absent. Notably, one of Kharlamova's main findings is that the specific subdialect spoken in Selenica showed both some phonetic characteristics classified as typical for the Northern dialect region as well as some classified as typical for the Southern one, effectively rendering it an 'intermediate idiom' between the two ideal dialect types.

In their grammatical description of the Aromanian language, based on the Southern dialectal varieties predominantly spoken in Greece, Katsanis and Ntinas (1990) provide an extensive overview of Aromanian phonology and morphology. Their analysis proposes the Aromanian vowel space to be constituted of six distinct phonemes /a/, /e/, /i/, /o/, /u/, /ə/ and /ɨ/. Similarly to Schlösser, they also cede a lot of relevance to explanations through a diachronic perspective. With regard to the vowel

/ə/, they state that it most likely evolved from a Vulgar Latin /a/ in unstressed positions, except for word-initial ones. On the other hand, the emergence of /ɨ/ can often be attributed to Vulgar Latin contexts of a prenasal /a/, according to them, which is matching the diachronic patterns observed by Weigand and Schlösser.

Kahl and Pascaru (2018) explore how similar the cultural and linguistic identities of Aromanian speakers from either broader region of the dialect spectrum are to each other. While only partially concerning phonology, their paper offers some valuable insights that could be interpreted from a sociophonetic perspective: For instance, they discover a tendency for older as well as female speakers of the Northern dialects to rather conserve the allophonic realization of the /r/ phoneme as [ʁ], which is typical for the Farsherot varieties, whereas younger, less traditionally living as well as male speakers are more inclined towards a realization as [r]. Kahl (2005) provides a comprehensive overview of the state of research, with regard to Aromanian, and summarizes the most relevant open questions. Those concern the sparseness of localized texts, insufficient consideration of localisms, documentation of everyday language use, syntactic descriptions, effects of language contact, sociolinguistic aspects, bilingualism, onomastics as well as the emergence of a written form of the Aromanian language. With respect to phonology and the commonly made distinction between a Southern and a Northern dialect group, this latter article makes the interesting remark that—somewhat in line with the later findings by Kharlamova (2019)—feature-based classifications (e.g. between a type A and a type F; see above) are not as airtight as they might seem, given that in both dialect regions a high degree of local variation can be observed which runs counter to the general trend postulated by such classifications.

2.2 Acoustic-Phonetic Methods in General

Harrington (2010) presents a general introduction into the domain of acoustic phonetics and aspects of the established theory behind it. This includes the motivation for modelling acoustic vowel quality using formant-frequency measurements. A review of the current best practices in measuring potential vowel mergers is laid out by Nycz and Hall-Lew (2013), which includes a comparison of the merits and disadvantages different metrics for measuring vowel overlap, e.g. the *Pillai score*, display. Three rather practical online tutorials on automatic formant extraction (see Stanley, 2017), formant analysis (see Harrington, 2018) and measuring vowel overlap (see Stanley, 2019) were also consulted.

3 Methods

This section will cover an extensive overview of the methodological approach chosen for conducting the present acoustic-phonetic corpus study.

3.1 Aromanian Speech Corpus

The speech corpus that was examined in the course of this project is based on a loose collection of Aromanian spontaneous-speech recordings featuring individuals from Turia (Greece, in 2000/2001), Kutsufliani (Greece, in 1999), Andon Poçi (Albania, in 2004/2005), Gjirokastër (Albania, in 2003) and Stjar (Albania, in 2015). Those recordings originate from field studies conducted by Thede Kahl and others. Reference works by Dietrich, Kahl and Sarros (2001), Bara, Kahl and Sobolev (2005) as well as Rădulescu and Kahl (2006) provide a more detailed overview of the circumstances around and motivation behind the earlier field studies. Two of the recordings are also available as videos on YouTube, published by the Commission Vanishing Languages and Cultural Heritage (VLACH); see the references Kahl et al. (2019) as well as Kahl, Poci and Fichera (2020). Several Aromanian native speakers were interviewed and asked to recite local fairy tales or to talk about their traditional customs, their everyday life or some historical event they had witnessed.

1. Southern dialect region (Greece)
   - Turia (2000/2001)
     - WT_1, female speaker, 04:34 min, 2,469 segments *(noisy audio)*
     - WT_2, female speaker, 03:12 min, 1,481 segments
     - MT_1, male speaker, 05:11 min, 2,801 segments
     - MT_2, male speaker, 01:58 min, 906 segments *(noisy audio)*
   - Kutsufliani (1999)
     - WK_1, female speaker, 01:23 min, 595 segments
     - WK_2, female speaker, 00:34 min, 315 segments
     - WK_3, female speaker, 01:04 min, 582 segments
2. Northern dialect region (Albania)
   - Andon Poçi (2004/2005)
     - WA_1, female speaker, 04:36 min, 2,287 segments
     - WA_2, female speaker, 04:13 min, 2,034 segments
     - WA_3, female speaker, 00:26 min, 242 segments

  - MA_1, male speaker, 02:08 min, 1,256 segments *(bird chirping)*
  - MA_2, male speaker, 01:30 min, 702 segments *(noisy audio)*
- Gjirokastër (2003)
  - MG_1, male speaker, 14:46 min, 6,513 segments
- Stjar (2015)
  - MS_1, male speaker, 04:35 min, 2,282 segments

As it can be noticed, the first character within the anonymizing label assigned to each speaker reflects their sex (*M* for male, *W* for female). The second character is always identical with the initial letter of a speaker's place of origin, among the given possibilities Turia, Kutsufliani (Southern dialects) as well as Andon Poçi, Gjirokastër and Stjar (Northern dialects). Overall speech time sums up to 32:10 min. Given like this, though, that interval does not exclude the minor portions of time that are taken up by the interviewers' own speech in some recordings, which are, of course, not the object of study here. A more comparable figure is the number of analysable segments for each speaker that is given right next to the corresponding recording duration. Those numbers only include actual phonetic segments, i.e. vowels or consonants, not speech pauses or other disturbances. A somewhat challenging property of the speech corpus is that the numbers of segments are not at all evenly distributed among the fourteen speakers. The mean length of analysable speech per speaker lies at 1,748 segments; the median is 1,369. With only 242 segments, Southern speaker WK_2 is featured with the shortest amount, whereas Northern speaker MG_1 holds a lone record for the largest quantity of provided speech, with 6,513 segments in total. Obviously, this does not lead to a very ideal situation with regard to between-speaker comparability of results, but it is a situation that has to be dealt with either way. It can be expected that some more granular speaker-specific features that are going to be analysed, will only be able to show meaningful results for the speakers featured with larger speech quantities, while for other, more general aspects, even shorter speech recordings might turn out to be sufficient as a data basis. When it comes to comparing groups of speakers to each other, e.g. clustered together based on sex and/or dialect region, it is necessary to assign equal weight to each speaker, not each examined segment, as well as for any subsequent interpretation to be qualified with respect to the sizes of the compared groups (e.g. there are five Southern female speakers, but only two Southern male ones). All recordings were made between 1999 and 2005, with the only exception of speaker MS_1 whose speech was recorded a bit more recently, in 2015. No information is given, unfortunately, on the age of each individual speaker during the time of the recording. Judging just by auditory means, they all seem to be somewhere between 45 and 75 years old.

Another thing to keep in mind here is that these recordings were not primarily created for the purpose of phonetic analysis, but rather to serve as an archive for the study of Aromanian language and culture in other domains like dialectology, ethnology or Romance studies. That is why a few of the recordings were not made in very ideal conditions with regard to audio quality. In the recordings of three of the speakers (marked with *noisy audio* in the overview above), a higher-than-average steady noise was noticeable just by a quick auditory impression. Such noise can probably be attributed either to the specific equipment used for the recording session, or to the environment in which it was conducted. That issue was trivial enough to repair by applying a noise-reduction effect to the affected audio files, which was done using the open-source audio editing software Audacity (see Mazzoni and Dannenberg, 2000). The parameters chosen for this process of noise reduction were *Noise reduction (dB): 12; Sensitivity: 6.00; Frequency smoothing (bands): 3.* Speaker MA_1's speech recording was affected in a different and very particular manner: It contained sounds of bird chirping, most prominently present roughly within the frequency bands 3,500 and 5,500 Hz. This could partially be tackled by applying a high-pass filter effect, again in Audacity, with the parameters *Frequency (Hz): 3,400; Roll-off (dB per octave): 48.* The bird chirping still remains audible after the application of the filter, but its presence is way more subtle. Speech remains comprehensible and does not sound distorted. A lower frequency parameter for the high-pass filter would have removed the unwanted background noises more effectively, but would have also given rise to new difficulties as a reliable extraction of the third formant frequency (which can range roughly up to 3,500 Hz) would not be guaranteed anymore.

3.2 Transcriptions

The Aromanian speech recordings were kindly provided together with transcriptions by the authors of the field studies they stem from. Even though they do not come directly in a standard phonetic script like IPA, most transcriptions use a convention very close to ideal phonemic orthography. Some, but not all of the transcriptions also denote relevant phonetic information like word stress explicitly. An orthographic distinction between /ə/ and /ɨ/ was not maintained, however, in the vast majority of transcriptions. Typically, tokens of either of the two vowel types were simply denoted as ⟨ă⟩ / ⟨ā⟩. Only for a single one of the recordings, featuring speaker WT_2, an /ə/-versus-/ɨ/ contrast was reflected orthographically as ⟨ă⟩ versus ⟨î⟩. For the Northern-Aromanian recordings, the provided transcriptions often used the symbols ⟨ϵ⟩ / ⟨ε⟩ alongside ⟨ă⟩ / ⟨ā⟩ for expressing instances of what would typically be assumed to be the unified phoneme $/ə/_{\text{single}}$. The motivation for this last aspect seems to be a consideration of a distinct feature that can be found in the sound system of the Northern dialectal varieties, which is that realizations of $/ə/_{\text{single}}$ are distributed across a noticeably vast space of possible vowel quality, ranging roughly from an open-mid front position around [ɛ] to a more central one around [ə]. So in those transcriptions, ⟨ϵ⟩ / ⟨ɛ⟩ stands for rather fronted realizations of the phoneme $/\text{ə}/_{\text{single}}$. For illustration, here is the unedited transcription of a 00:32-min excerpt of Northern speaker *WA₂*'s recording:

> Tundea oili dolj frats shi si-acātsϵ cu frati-su, el, shi-acātsarā aco iu tundea oili sh-frat-su lungiunghe sh-muri Ioryou. Sh-iel dzϵse: A, dzϵse, ϵnj yine s-plϵng! dzese frati-su, cara vātāmā frati-su s-fétsi pishmani. Sh-dzϵse: Ia s-mi fϵtsiam un pulju ta s-lu kϵftam Ioryo njedzϵnoptsϵ, dzϵse. Tu doaupradzli sihatea di noptsϵ, dzϵse. Sh-tora nóptea, kϵndu nédzi sihatea doauspratsi lu caftϵ Ioryul. Dzetse: Ioryooo... pulju. Lu-ai avdzϵtϵ tini Spiro? E.

To complement this, the unedited transcription of Southern speaker *WK₂*'s complete recording, which is 00:34 min long, can be found right below:

> Siára di Áyiu Vasíli, pirín Áyiu Vasíli, nă adunámu tu căsi viʦínată. Múlʦă. Fiáti, fičóri. Ș-ave̯ám foc tu fucurínă. (Ș-aliʥe̯á) adunámu grắnu di tu păhnía-a právdiloru, ʦi măngá bólʲi pálʲi. Adunámu bubólʲi di grăn di-aclóʦi ș-băgámu unu bubólʲi unu fičór, alántu e̯ará fiată, ala 'ʦe̯á la cári vre̯á unu ș-alántu. Áma si-arde̯á apóia, di si-anvărte̯á do̯áuli grăni 'ʦe̯ámu si aɣăpiséscu, va s-lʲe̯á, áma vʥe̯á unu cătră au̯á, alántu cắtră acló, 'ʦe̯ámu nu s-vrúră. Aʦe̯á.

In order to acquire an automatic segmentation of the speech recordings, the web interface of the HMM-based Munich Automatic Segmentation System was used (a.k.a. WebMaus, see Kisler, Reichel and Schiel, 2017). The WebMAUS General application requires raw audio files to be uploaded together with corresponding phonetic transcriptions encoded e.g. in the SAMPA script. So, as an initial step the speech recordings had to be re-transcribed into SAMPA manually, leaving out all notation regarding stress and phrase boundaries. The existing transcriptions proved to be a great help here since they already provided the valuable information on word boundaries and phone classes, which would be quite hard to reliably extract from the audio alone without an advanced level of proficiency in the Aromanian language.

This is how the raw SAMPA transcription of the short recording excerpt from Northern speaker *WA₂*'s speech, mentioned above, looked like:

```
tundea oili doLi fRats S si ak@ts@ ku fRati su el Si
ak@tsaR@ ako iu tundea oili S fRat su ludZuNge S muRi jorGu
S iel dzse a dz@se J jine s pl@N dz@se fRati su kaR v@t@m@
fRat s s fets piSmani S dz@se ia s mi f@tsiam un puL ta s
```

```
lu k@ftam jorGo Jadz@nopts dzs tu daulispRadz di nopts@ dzs
S noptea toRa Jadzenoptea k@n nedzi sihatea dauspRatsi lu
kaft@ jorGul dz@tse jorGo puLu l ai avdz@t tin spiro e
```

Again, for comparison, this is how it turned out for speaker *WK₂*:

```
siara di aju vasili pirin aju vasili n adunam tu k@si
vitsinat@ mults@ fiati ftSori s aveam fok tu fukurin@ S
alidzea adunam gr@n di tu p@xnia a pravdilor tsi m@Nga boLi
paLi @l adunam buboLi di gr@n di aklotsi S b@gamu un buboLi
un ftSor alant eara fiat@ ala tsea kari vrea unu S alantu
ama si ardea apoia di s anv@rtea doauli gr@ni tseam si
aG@psesku va s Lea ama vdzea un k@tr ua alant k@tr aklo nu
s vrur@ atsea
```

Note that all instances of vowels of interest (i.e. realizations of any assumed /ə/$_{\text{single}}$, /ə/ or /ɨ/) have been represented with SAMPA's schwa symbol ⟨`@`⟩ here. This entailed grouping some Northern vowel realizations auditorily closer to [ɛ] together with some Southern ones that clearly sounded like somewhere on the [ə]∼[ɨ] spectrum into the same underlying category of representation. By contrast, instances of any other phone have been denoted in SAMPA with sensitivity to their particular phonetic appearance. As an example, the Aromanian consonant phoneme /r/ has been transcribed as ⟨`r`⟩ $\hat{=}$ [r]∼[ɾ] (e.g. above for WK_2), ⟨`R`⟩ $\hat{=}$ [ʀ]∼[ʁ] (e.g. above for *WA₂*) or sometimes even ⟨`x`⟩ $\hat{=}$ [x]∼[χ], depending on the specific type of its realization. Using a phonetically more granular encoding for sounds other than the vowels of interest was motivated by methodological considerations: In examining the potential effect of coarticulatory context on vowel quality, it is expectable for, say, a preceding consonantal context as in /rə/ to exert a different kind of influence whether realized as an alveolar trill or a uvular fricative, due to the dissimilar articulatory properties between either type of realization.

Moreover, the SAMPA transcriptions already account for some pronunciation reductions that were not captured by the original transcriptions, for instance:

- ⟨frati-su s-fétsi⟩ → ⟨`fRat s s fets`⟩
- ⟨unu cătră au̯á⟩ → ⟨`un k@tr ua`⟩

3.3 Segmentation Process

For further analysis, the speech recordings had to be segmented and annotated first. As already hinted at, this was done using an HMM-based automatic-alignment tool (see Kisler, Reichel and Schiel, 2017), but also required manually correcting the automatically determined segment boundaries afterwards.

3.3.1 Automatic Alignment

Best results for the automatic segmentation process were achieved by setting the WebMAUS' language option to *German (DE)*. Even if that might not be very plausible intuitively, after some experimentation it turned out that, with that option enabled, the resulting automatic segmentations for the examined Aromanian speech excerpts were substantially more precise, compared to the performance of other options like *Language indep. (sampa)*. Using *Romanian (ro)* was also considered, but had to be discarded as that would require the SAMPA-transcribed input texts to only contain sounds present in the Dacoromanian phone inventory, which is a problem for several consonants specific to Aromanian, like [ʎ] or [x]. Of course, the German phone inventory is not a superset of the Aromanian one either, but at least as it is defined in WebMAUS, the only Aromanian sounds not recognized for the *German (DE)* option are the affricate ⟨`dz`⟩ ≙ [d͜z], which would be analysed as two separate sounds ⟨`dz`⟩ ≙ [d.z], and the consonants ⟨`J`⟩ ≙ [ɲ], ⟨`G`⟩ ≙ [ɣ] and ⟨`c`⟩ ≙ [c]. This was dealt with by temporarily replacing all ⟨`J`⟩ by ⟨`nj`⟩, ⟨`G`⟩ by ⟨`R`⟩ and ⟨`c`⟩ by ⟨`k`⟩ in the input text files. The effects of those edits were then manually reverted in the output *TextGrid* files; it was also ensured for instances of ⟨`dz`⟩ to be assigned only one instead of two separate segment intervals. On a sidenote, in Aromanian the sound [c] is an allophonic variant of /k/ that is only present in the speech of some individuals and that typically occurs when followed by an /e/ or /i/ vowel. All remaining tweakable parameters provided by WebMAUS General were left on their default settings.

3.3.2 Manual Adjustment

The resulting automatic segmentations were quite decent for large portions of the analysed speech. However, there was also a considerable number of cases where the proposed alignment was completely mismatching the true position in the recording. In order to ensure a reliable subsequent analysis, each individual segment boundary was checked and adjusted by hand whenever necessary. While doing that, it was also made sure to exclude sections of fillers and similar semi-linguistic utterances, of sporadic background noise and of (overlapping) speech not uttered by the main speaker, but by the interviewers or some unidentified bystanders. Some speakers displayed a tendency to code-switch between Aromanian and Greek, in which case the Greek speech sections were also excluded. Wherever effects of coarticulation were so heavy that entire phonetic segments could not be meaningfully mapped to the speech signal, they were deleted as well.

3.4 Formant Tracking

By running a script through the open-source software tool Praat (see Boersma, 2001), formant frequencies for F1, F2 and F3 were extracted for all vowel segments, i.e. all segments labelled as ⟨`a`⟩, ⟨`e`⟩, ⟨`i`⟩, ⟨`o`⟩, ⟨`u`⟩ or ⟨`@`⟩. The exact approach chosen here was to calculate the mean formant values of the middle 60 percent of each relevant interval, thus discarding 20 percent from the beginning and another 20 percent from the end of any segment. As heavy coarticulation effects are most present on the edges of phonetic segments, they might distort formant values if those were to be obtained by averaging over entire segment intervals. However, even under the chosen conditions, it will still be necessary to control for the effects of coarticulatory context, as it has been shown that such effects can have an influence reaching beyond just the edges of adjacent phonetic segments (see Grosvald, 2010).

The parameters *Number of formants*, *Windows length*, *Dynamic range* and *Dot size* were set to 5, 0.025 s, 30 dB and 1.0 mm respectively. According to general recommendations (see Mayer, 2017), when performing formant extraction it is best practice to set the *Maximum formant* parameter to a value of 5,500 Hz for an average female voice or to 5,000 Hz for an average male voice. For this study, it was set to the average of those two values, 5,250 Hz, for all of the featured speakers. Choosing a universal ceiling value like this was not only done for reasons of simplicity: Considering that speaker sex is going to be included as a parameter in parts of the subsequent analysis, any risk of creating a bias by choosing different parameters for the extraction of male-voice formants and female-voice formants should better be avoided. This parameter setting was also chosen by Heeringa, Johnson and Gooskens (2009) in their study on Norwegian dialect distances.

3.5 Annotating the Vowel Dataset

Vowel duration was trivially measured as the difference between the start and the end point of any segment. For each vowel, information on F1, F2, F3 (in Hertz), duration (in milliseconds) and the word token inside which it appeared (as a string) were collected and stored in a preliminary dataset. That dataset was further enriched by information on coarticulatory context (more details right below in Section 3.5.1) which could be retrieved automatically.

While all of the above was done for all vowel types, the subdataset restricted to only the vowels of interest, marked as ⟨`@`⟩, was extended by manual annotation to also encompass values for the additional variables word stress (see Section 3.5.3), part of speech (see section 3.5.4) and a mapping to either /ə/ and /ɨ/ based on ideal phonological assumptions (see Section 3.5.2).

3.5.1 Context of Coarticulation

As complete phonetic segmentations for all of the speech recordings were available by now, it was possible to extract the class of each vowel's preceding and succeeding segment. That class could either be a type of phone or a speech pause. The overall unique set of possible segment classes contained 40 elements. In order to simplify a bit further, several consonant phone types with similar articulatory-acoustic properties were grouped together into unified categories. Note that a slightly different kind of grouping was used for preceding and for succeeding contexts, as given here:

1. Common for both preceding and succeeding contexts:
   - *C1:* Speech pause
   - *C2–C7:* [a], [e], [i], [o], [u], [ɛ]-[ə]-[ɨ] *(vowel of interest)*
   - *C8–C15:* [m], [n], [ŋ], [ɲ], [l], [ʎ], [r], [h]
   - *C16–C19:* [b]-[p], [v]-[f], [ð]-[θ], [ʁ]-[ɣ]-[x]
2. Only for preceding contexts:
   - *P1–P3:* [ɟ]-[ç]-[c], [g]-[k], [d]-[t]
   - *P4–P5:* [z]-[d͜z]-[s]-[t͜s], [ʒ]-[d͜ʒ]-[ʃ]-[t͜ʃ]
3. Only for succeeding contexts:
   - *S1–S2:* [g]-[k]-[c], [ɟ]-[ç]
   - *S3–S5:* [z]-[s], [ʒ]-[ʃ], [d]-[d͜z]-[d͜ʒ]-[t]-[t͜s]-[t͜ʃ]

Grouped like this, there are now 24 different categories for preceding contexts (*C1 . . . C19* and *P1 . . . P5*) as well as 24 for succeeding contexts (*C1 . . . C19* and *S1 . . . S5*). Note that the difference between the category subsets *P1 . . . P5* and *S1 . . . S5* is motivated by asymmetric phonetic properties of the affricates [d͜z], [t͜s], [d͜ʒ] and [t͜ʃ] and the palatal stop [c]. As far as the affricates are concerned, that asymmetry is pretty straightforward: An affricate begins as a stop and is released as a fricative, thus it is expectable for it to exert an anticipatory coarticulatory influence similar to a stop, but a perseveratory coarticulatory influence similar to a fricative. The situation is a bit less clear as far as the palatal stop [c] is concerned. If pronounced ideally, it should not be treated as 'asymmetric' within a categorization like the one above. However, in its specific Aromanian appearance as an occasional allophone of /k/, its acoustic manifestation often retains a clearly audible velar onset, making it plausible to alternatively describe that sound as a *palatalized* velar plosive, i.e. [k^j] or even [kç]. Described like this, it becomes clear why that sound is grouped together with [ɟ] and [ç] in *P1*, but with [g] and [k] in *S1*. Effectively treating [c] as equivalent to [kç] also avoids running into categorization inconsistencies with respect to [c]'s voiced counterpart, [ɟ], which only speaker MG_1 robustly employed as an allophonic variant of /g/ in contexts before /e/ or /i/. It was denoted using the analogous approximation ⟨gj⟩ ≙ [gɟ].

Defining the coarticulatory context of a vowel realization merely by the category of the immediately preceding and the immediately succeeding phonetic segment shall be sufficient for the purposes of this study. It makes up for two nominal variables, each for either side, that can be added to every vowel in the dataset. An initial consideration to include some encoding of syllable structure as another predictor variable for acoustic vowel quality was dismissed, mainly in order not to overcomplicate the analysis. Information on syllable structure is likely to be somewhat associated with the factors context of coarticulation and duration anyway: For example, suppose there is a schwa realization with preceding context *C14*, [r], and succeeding context *C1*, speech pause. Then it is quite safe to assume that it forms the nucleus of a syllable with a structure of either CV, e.g. [rə], or CCV, e.g. [drə]. On the other hand, if the succeeding context in this example was replaced by *C9*, [n], there would be some ambiguity left as to whether this is an open or a closed syllable, i.e. if that succeeding nasal segment belongs to the coda of the syllable with the schwa nucleus, or if it is part of the following syllable's onset. Here, information on vowel duration might be indicative for resolving the ambiguity as it can be expected to observe longer vowel durations for the nucleus of a syllable that is open, compared to one that is closed.

3.5.2 Vowel Type

If a vowel token was associated with one of the five types /a/, /e/, /i/, /o/ and /u/, the transcriptions already provided the information which type it was. For the supposed vowel-of-interest types /ə/ and /ɨ/, however, such distinctive information was initially not available (except for WT_2), and thus had to be annotated manually for each individual vowel token marked as ⟨@⟩. The goal was to retrieve a reliable mapping of any vowel-of-interest occurrence to either /ə/ or /ɨ/, based on what might be called *ideal Southern-Aromanian phonology*: As laid out earlier, it can be noted that in theoretical literature it is the Southern varieties that are assumed to maintain a (phonemic) distinction between such two non-central open vowel types. This provides the motivation to apply that theoretical mapping to actual speech data and find out if it can be confirmed for such a mapping to be, indeed, relevant in a quantitative examination of Southern-Aromanian speech, but rather irrelevant in one of Northern-Aromanian speech, as literature would often suggest.

That is why, in this context, /ə/ and /ɨ/ are not referred to as *vowel phonemes*, but instead, more vaguely, as *vowel types*, given that their phonemic status is actually unclear and something that is going to be tested for. So, for example, whenever a wording like 'the /ɨ/ realizations in the Northern varieties' is used hereinafter, it will neither entail that a distinct phoneme /ɨ/ exists for the respective varieties, nor that the vowel realizations referred to are acoustically resembling a closed front unrounded one, i.e. [ɨ]. Instead, such a wording will only suggest which vowel type those vowel realizations *would* correspond to if classified along the lines of an ideally assumed phonology for the *Southern* varieties. In order to acquire such a mapping for the speech corpus examined here, several scientific as well as literary sources were consulted. Weigand (1888), Schlösser (1985) and Katsanis and Ntinas (1990) provided helpful explanations of general morphophonological patterns in those regards, for instance, about the phenomenon that verb forms with a monosyllabic stem and a nucleus /a/, e.g. /ˈbagə/ (he/she puts, *present tense*), consistently reduce this /a/ to /ə/ in other inflected forms where the stem is not stressed, e.g. /bəˈgə/ (he/she put, *past tense*), or that word-initial vowels of interest can only ever correspond to /ɨ/, unless some phonetic reduction happened. While morphophonological patterns like these already account for a reliable mapping of a substantial number of the concerned vowel tokens to either /ə/ or /ɨ/, this was—expectably—still not sufficient overall. For the remaining tokens, the mapping had to be retrieved on a lexical level. In addition to the three theoretical works just mentioned, a sparse collection of glossaries and Aromanian literary texts that employed a graphemic convention reflecting a distinction between /ə/ and /ɨ/ was thoroughly searched for entries and occurrences of word forms associated with the same lexical lemma as any word token from the speech corpus which was containing an as yet unspecified vowel of interest. Sources included in this process were a comprehensive glossary appended to an anthology of Aromanian folk tales and songs that was compiled by Papahagi (1922), a somewhat smaller glossary provided with a collection of traditional Aromanian song texts from the town of Metsovo that was put together by Papadiotis (1988), as well as a booklet accompanying a rather recently published album of musicalized Aromanian poetry by Nitsiakos, Liakos and Seviloglou (2019).

It turned out that one would often find ambiguous or contradicting information within a single source or when comparing several sources to each other. Just to give a few examples: The interrogative adverb /k(ə|ɨ)ˈt͜se/ (why?) is transcribed with a grapheme corresponding to /ɨ/ by Papahagi (1922), but with one corresponding to /ə/ in the remaining sources. Katsanis and Ntinas (1990) explicitly mention the existence of variants with either vowel type for the prepositions /ˈf(ə|ɨ)rə/ (without) and /ˈp(ə|ɨ)nə/ (until). Papahagi (1922) lists the noun /aˈr(ə|ɨ)u/ (river) with a grapheme for /ɨ/, in line with what was discussed in Section 2.1 about the minimal pairs proposed by Beis and Dasoulas (2017). However, Papadiotis (1988) transcribed the same noun with a grapheme corresponding to /ə/. In all those unclear cases, an annotation decision had to be made eventually, sometimes motivated by which variant was used in the majority of sources (e.g. motivating /kəˈt͜se/ and

/aˈrɨu/), sometimes by analogies to equivalent or similar words in the related language Dacoromanian (e.g. motivating /ˈfərə/ and /ˈpɨnə/) or a contact language of Aromanian. A prior consideration to mark such ambiguous cases separately in the annotation, as some sort of third, in-between category, was discarded, as that would have given rise to unnecessary methodological complications during the subsequent statistical analysis. Even after careful consideration of all available sources, some of the collected vowels of interest could not be confidently associated with either of the two types /ə/ and /ɨ/, often because they appeared in words that were quite rare or presumably regionalisms. This was the case for 65 of the total 1,603 vowel-of-interest observations, and they were excluded from the resulting dataset.

3.5.3 Word Stress

For each realization of a vowel of interest, it was determined by hand whether it appeared in a stressed (a.k.a. *tonic*) or unstressed position within the word token containing it. Among unstressed positions, a further differentiation was made between syllables that were *pretonic* (immediately preceding the stressed syllable), *posttonic* (immediately succeeding the stressed syllable) or *atonic* (neither). For illustration, these are two word tokens produced by speaker MT_1:

1. /bə.ˈgə/ (he/she put)
2. /və.tə.ˈma.rə/ (they killed)

In example 1, the first schwa is in *pretonic* position, and the second schwa is in *tonic* position. Example 2 features three schwas which are in *atonic*, *pretonic* and *posttonic* position respectively. Some words in Aromanian are clitics. These include short forms of direct and indirect pronouns and monosyllabic prepositions or conjunctions like /tə/ or /kə/. Here are three examples of clitics extracted from the speech of speaker WT_1:

3. /kə/ (that) in /kə.a.ˈvet͜s/ (that you *pl.* have)
4. /ɨɲ/ (to me) in /ɨɲ.ˈd͜zɨ.si/ (he/she said to me)
5. /ʎu/ (to it) in /ˈdə.ʎu/ (give to it!)

Examples 3 and 4 show proclitics, whereas example 5 features an enclitic word. The idea is to define word stress for any clitic as if it were belonging to the adjacent word token which it depends on phonologically. In consequence, it can be said that the position of /ə/ in example 3 is *atonic*, /ɨ/ in example 4 is *pretonic*, and /u/ in example 5 is *posttonic* (although note that the last case is not directly relevant here as it does not concern a vowel of interest).

3.5.4 Part-of-Speech Tags

In the domains of natural language processing and corpus linguistics, part-of-speech (POS) tagging describes the process of assigning each word in a text a label that represents its lexical category. Any such label is an element of a set of POS tags that is specific to the language and the task in question. Usually, the goal is to automatize that process by initially training statistical or deep-learning models on a manually annotated gold-standard corpus. The resulting models are called POS taggers.

It is not within the scope of this study to apply or implement any POS tagger. However, the categorical framework that a POS tag set provides can be used to encode some information on the syntactic function of individual word tokens in a quite straightforward way. Of course, the information contained in a POS tag is not syntactic in a stricter sense. For instance, even if some word token W is assigned a tag CN that supposedly represents countable nouns, this still does not reveal anything about whether the noun phrase whose head W is appears as a subject, an object or a complement of some other phrase type within the surrounding sentence. POS tags are not designed to capture dependency or constituency information. Nevertheless, knowing the lexical category of a word token does provide an approximation of its syntactic role. Tokens that share the same lexical category are likely to fulfil equivalent syntactic roles on average.

Simionescu (2011) developed a POS tagger for Dacoromanian which combines statistical with more traditional rule-based approaches. That system, sometimes called the UAIC-Ro POS tagger, seems to be the most established one of its kind for that particular language. There have already been discussions on whether it might be applicable to corpus studies on related varieties like Istroromanian, Meglenoromanian or Aromanian (see Bobicev, Mărănduc and Perez, 2017). Simionescu's POS tag set is very granular as it encodes detailed morphological properties of a word form in addition to its general lexical category. As an example, the complete POS tag for the Dacoromanian word /ˈfrat͡sii/ (the brothers) would be *Ncmpry*, which stands for:

Tag	*N*	*c*	*m*	*p*	*r*	*y*
Meaning	noun	common noun	male	plural	direct case	with article

Table 1: Example of an UAIC-Ro POS tag. The first character (here, *N*) represents lexical category (noun); each following character encodes a certain morphological feature.

The Aromanian equivalent would be /ˈfrat͡sʎi/ (the brothers) and could be classified along the same lines. Although morphologically detailed tags like *Ncmpry* are very informative, they would not be suited as variables in a limited corpus study like this one, due to their sheer number. There are 376 such granular UAIC-Ro POS tag types in total. By considering only the first character of such a tag, a label representing a broader lexical category can be obtained, e.g. *N* for nouns, *V* for verbs, *Q* for particles etc.

For this study, each of the collected vowel-of-interest realizations has been further annotated with one of the 14 broader-category tags proposed by Simionescu, based on the lexical category of the word that it appeared in. Two of those tags, *X* for foreign words and symbols as well as *Y* for abbreviations, were never used as they rather apply to written text anyway. The remaining 12 tags, together with a few examples from the vowel-of-interest dataset, are given in Table 2. Note that some word forms can be mapped to more than one lexical category due to the fact that they fulfil different grammatical roles in different contexts. For instance, /aˈestə/ and /ˈvərə/ can function either as pronouns or as determiners. In the same way, /ˈunə/ might be interpreted as an indefinite article in some contexts, but as a numeral in others. The word form /ˈpɨnə/, in analogy to its English translation, can be either a preposition or a subordinating conjunction. Finally, a past-participle form like /ˈgritə/ or /arvuɲiˈsitə/ can be used both for constructing a perfect-tense verbal phrase and as an attribute of a nominal phrase.

As far as the motivation for representing part-of-speech information in the dataset for /ə/ and /ɨ/ realizations is concerned: There is no obvious intuition for why the acoustic quality of vowels would vary depending on an approximative encoding of the syntactic role of the words that contain them. Even so, if it turns out that there is indeed almost no association between POS tags and vowel quality, including them will at least have served as a sanity check. In the opposite case, i.e. if there can be discovered such an association, this would lead to further interesting questions along the lines of whether it is really the lexical category per se causing the observed effect, or rather a specific correlated property of certain lexical categories, e.g. female nouns in Aromanian often ending in an unstressed /ə/. It had briefly been considered to keep track not only of individual POS tags, but also of tag sequences of a given length that precede a target word token, i.e. POS tag n-grams. In case of opting for 4-grams, meant as including the tag for the target word, the POS label for a vowel realization /ə/ in the noun /ˈkasə/ (house) would then be *P–V–S–N* instead of just *N* in the context /lu arˈka pr ˈkasə/ (they scattered it in front of the house). A label like *P–V–S–N* is surely syntactically more informative than just *N*, but it runs into the same problem discussed above for complete UAIC-Ro tags like *Ncmpry*: Too much granularity makes the part-of-speech information unsuited for a statistical analysis of within-speaker variation in a small speech corpus like the one at hand.

POS	Category	Examples	
N	noun	/ˈkɨɲ/	(dogs)
		/ˈɡrɨnlu/	(the wheat)
		/ˈxo̯arəʎei/	(of the village)
V	verb	/mərˈtə/	(he/she married)
		/t͜sɨˈnet͜s/	(you *pl.* hold)
		/ˈɡritə/	(shouted, *perfect tense*)
A	adjective	/vɨrˈtos/	(strong)
		/ˈmərli/	(the big ones)
		/arvuɲiˈsitə/	(engaged, *attributive*)
P	pronoun	/nə/	(us)
		/aˈestə/	(this one)
		/ˈvərə/	(someone)
D	determiner	/ˈmult͜sɨ/	(many . . .)
		/aˈestə/	(this . . .)
		/ˈvərə/	(some . . .)
R	adverb	/naˈfo̯arə/	(outside)
		/aˈʃɨt͜si/	(like this)
		/kəˈt͜se/	(why?)
S	adposition	/tə/	(for)
		/ˈkətrə/	(towards)
		/ˈpɨnə/	(until, *preposition*)
C	conjunction	/kə/	(that)
		/ˈkɨndu/	(when)
		/ˈpɨnə/	(until, *subord. conj.*)
M	numeral	/ˈunə/	(one)
		/ˈdo̯auə/	(two)
		/patruˈd͜zət͜s/	(forty)
T	article	/ˈunə/	(an/a . . .)
I	interjection	/ˈləi/	(hey . . .)
Q	particle	/sə/	(to)

Table 2: General UAIC-Ro POS tags associated with their represented lexical category as well as corresponding Aromanian example words that appear in the examined speech corpus and contain an /ə/ or /ɨ/ vowel.

3.6 Outlier Rejection

Summing up what was discussed in the previous sections, the two resulting vowel datasets look as presented in Tables 3 and 4 below. As indicated in the captions, the vowel datasets will hereafter be referred to as VOI and AV respectively. For all following steps of analysis, the values for F1, F2 and F3 have been converted from Hertz to Bark scale.

/ə/ and /ɨ/: 1,538 observations							
Speaker	WT_1	WT_1	WT_1	...	MS_1	MS_1	MS_1
Region	*South*	*South*	*South*	...	*North*	*North*	*North*
Sex	*Female*	*Female*	*Female*	...	*Male*	*Male*	*Male*
Vowel in Word	`d(@)vani`	`ndreag(@)`	`d(@)`	...	`r(@)m@J`	`r@m(@)J`	`ai(@)n`
POS Tag	*N*	*D*	*V*	...	*N*	*N*	*P*
Stress Position	*pretonic*	*posttonic*	*tonic*	...	*pretonic*	*tonic*	*tonic*
Vowel Type	/ə/	/ə/	/ə/	...	/ə/	/ɨ/	/ɨ/
Prec. Context	*P3*, `d`	*P2*, `g`	*P3*, `d`	...	*C14*, `r`	*C8*, `m`	*C4*, `i`
Succ. Context	*C17*, `v`	*C9*, `n`	*C13*, `L`	...	*C8*, `m`	*C11*, `J`	*C9*, `n`
Duration (ms)	81	40	71	...	75	107	95
F1 (Hz)	623	713	631	...	535	651	751
F2 (Hz)	1,599	2,187	2,015	...	1,720	1,973	2,048
F3 (Hz)	3,020	2,874	2,879	...	2,445	2,382	2,752

Table 3: Head and tail of the dataset VOI (vowels of interest).

/a/, /e/, /i/, /o/, /u/ and $/ə/_{single}$: 11,756 observations							
Speaker	WT_1	WT_1	WT_1	...	MS_1	MS_1	MS_1
Region	*South*	*South*	*South*	...	*North*	*North*	*North*
Sex	*Female*	*Female*	*Female*	...	*Male*	*Male*	*Male*
Vowel in Word	`ts(i)`	`(i)ara`	`i(a)ra`	...	`St(i)u`	`Sti(u)`	`m(i)n`
Vowel Type	/i/	/i/	/a/	...	/i/	/u/	/i/
Prec. Context	*P4*, `ts`	*C4*, `i`	*C4*, `i`	...	*P3*, `t`	*C4*, `i`	*C8*, `m`
Succ. Context	*C4*, `i`	*C2*, `a`	*C14*, `r`	...	*C6*, `u`	*C8*, `m`	*C9*, `n`
Duration (ms)	93	90	81	...	39	45	66
F1 (Hz)	430	757	716	...	341	289	523
F2 (Hz)	1,833	2,237	2,037	...	1,717	1,623	2,315
F3 (Hz)	2,595	3,403	2,342	...	2,596	2,486	3,097

Table 4: Head and tail of the dataset AV (all vowels).

Given that automatic formant extraction is prone to some error which can lead to outlier observations, it is advisable to detect and remove such observations in order to acquire a distribution that is more normal and a more accurate estimate of any particular vowel category's acoustic space. Vowel quality is multidimensional in that it is described by the values of at least two formants; in this case the three formants F1, F2 and F3. In consequence, a multivariate measure of outlyingness needs to be employed. Here, this was done by computing the squared Mahalanobis distance (D^2) for each vowel realization with respect to the distribution of all realizations of the same vowel type produced by the same speaker. As an example, Figure 1 shows the distribution of D^2 for realizations of /ə/ produced by the speakers MT_1

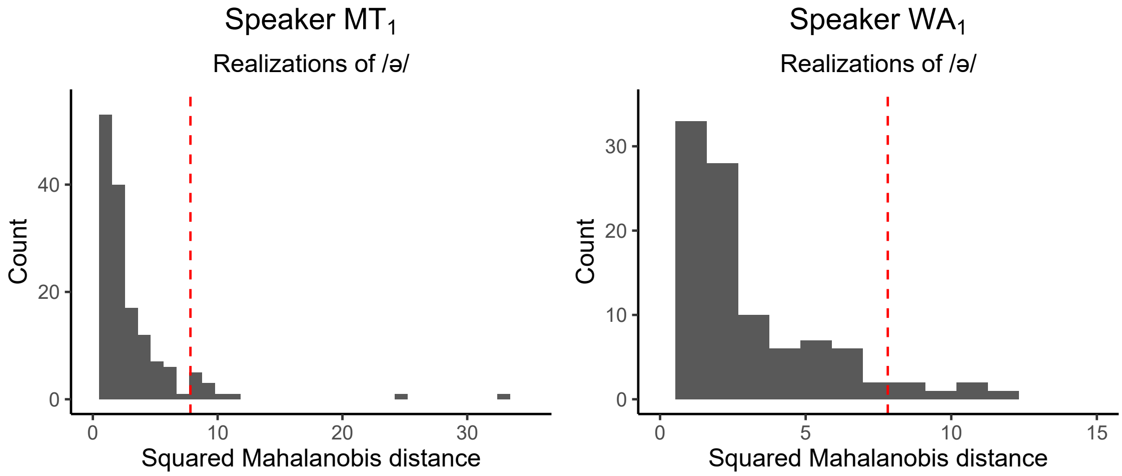

Figure 1: Histograms of D^2 (squared Mahalanobis distance) distributions for realizations of /ə/ uttered by speakers MT_1 and WA_2. The dashed, red vertical line at 7.82 indicates the threshold above which observations are going to be classified as outliers.

Region	Speaker	Pre-Excluded	Outliers	Remaining
South	WT_1	2	7	150
	WT_2	7	6	82
	WK_1	2	0	32
	WK_2	1	0	16
	WK_3	0	2	36
	MT_1	8	16	176
	MT_2	6	3	50
North	WA_1	8	8	152
	WA_2	5	11	116
	WA_3	0	0	15
	MA_1	1	3	67
	MA_2	2	2	41
	MG_1	19	24	369
	MS_1	4	9	145
Total		**65**	**91**	**1,447**

Table 5: VOI dataset: Number of observations per speaker grouped by whether they were pre-excluded because their vowel type could not be retrieved (see Section 3.5.2), rejected as outliers, or neither, i.e. remaining.

and WA_1. The dashed, red vertical line indicates a threshold at 7.82, above which observations will be treated as outliers. That specific threshold is motivated by the fact that D^2 typically approximates a χ^2 distribution, with the number of variables under investigation (in this case, three) corresponding to the degrees-of-freedom value set for χ^2. Thus, the critical χ^2 value for $p = .05$ and $df = 3$, which is 7.82, can be used as an outlier-rejection threshold for D^2 distributions that will exclude

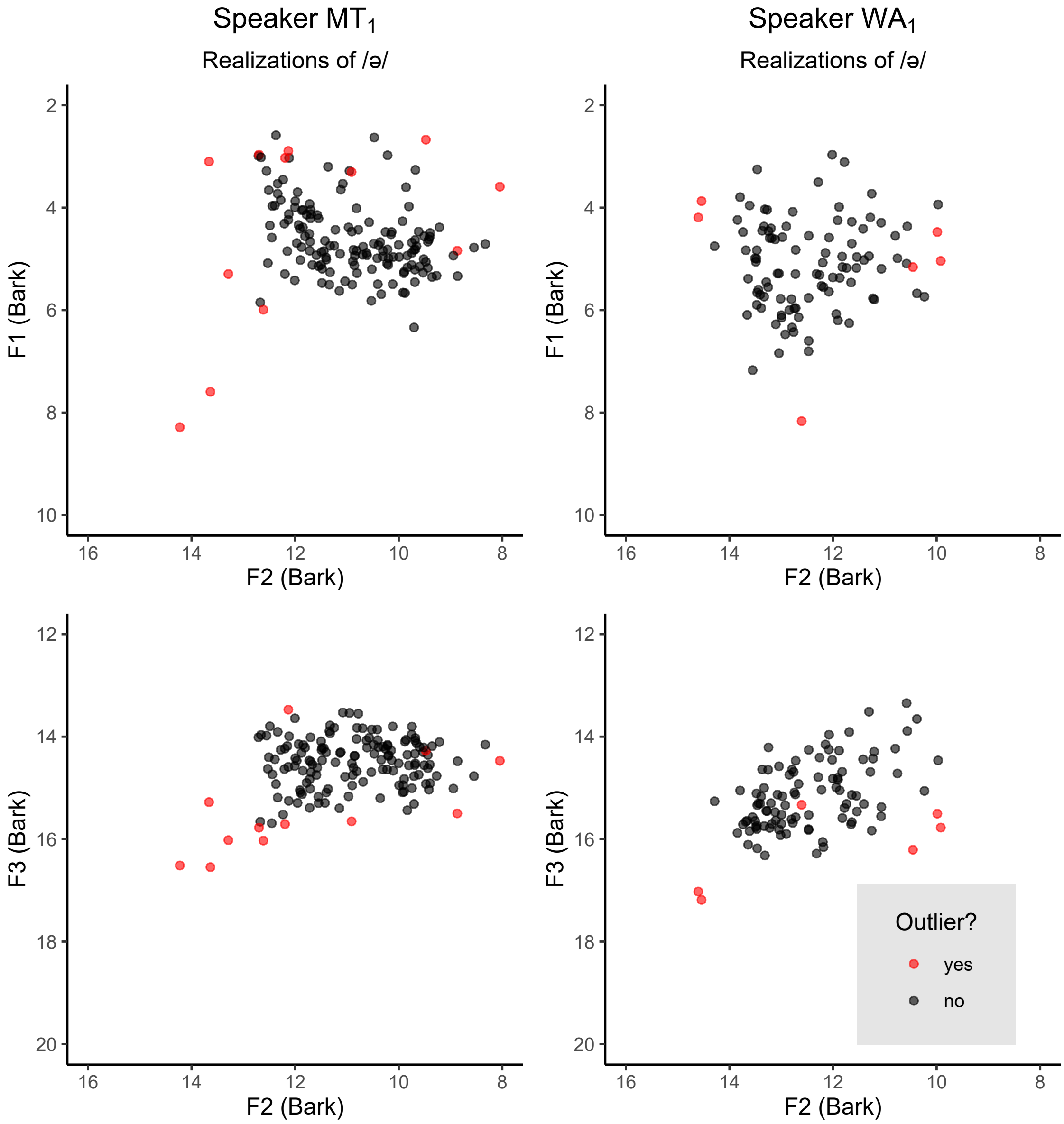

Figure 2: Scatter plots of the distributions of /ə/ realizations uttered by speakers MT_1 and WA_1 on F1×F2 and F2×F3 planes. Observations displaying a D^2 value above 7.82 are treated as outliers and marked red.

extreme observations which are roughly less-than-five-percent likely to have occurred by chance.

The histograms in Figure 1 indicate that there are a handful of moderate outliers for both speakers, with squared Mahalanobis distance values ranging somewhere between 8 and 12. For speaker MT_1, two more extreme outliers, with D^2 values of 24.2 and 33.0, can also be found. Figure 2 shows plots of F1×F2 and F2×F3 planes featuring those same two speaker-specific distributions of the /ə/ vowel. Observations that were classified as outliers based on the criteria established above have been marked red.

In the particular example, one can find twelve outlying /ə/ realizations for speaker

Region	Speaker	Outliers	Remaining
South	WT_1	75	1,163
	WT_2	39	657
	WK_1	18	267
	WK_2	11	143
	WK_3	14	272
	MT_1	91	1,273
	MT_2	24	417
North	WA_1	62	1,040
	WA_2	72	928
	WA_3	6	109
	MA_1	42	540
	MA_2	17	320
	MG_1	206	2,822
	MS_1	62	1,066
Total		**739**	**11,017**

Table 6: AV dataset: Number of observations per speaker grouped by whether they were rejected as outliers or not.

MT_1 as well as six such realizations for speaker WA_2. The process described here was repeated for each vowel type (/ə/ and /ɨ/ in the VOI dataset, /a/, /e/, /i/, /o/, /u/ and /ə/$_{\text{single}}$ in the AV dataset) and each of the fourteen speakers. Table 5 and Table 6 inform about how many outlying observations had to be rejected for each speaker and dataset.

3.7 Pillai Scores

The *Pillai-Barlett trace*, sometimes referred to as *Pillai's trace* or—in particular in its application to acoustic phonetics—*Pillai score*, is a test statistic for multivariate analysis of variance (MANOVA) and as such also used as an established metric for measuring the overlap between vowel distributions. Pillai scores are especially relevant in examinations of vowel mergers, as laid out by Nycz and Hall-Lew (2013). In order to determine how distinct speaker-specific distributions of /ə/ and /ɨ/ appear in the given Aromanian speech corpus, Pillai scores are going to serve as the primary measure to interpret in this study.

4 Results

To what extent the collected data is indicative of an overlap between speaker-specific distributions of /ə/ and /ɨ/, is going to be discussed right below (see Section 4.1). An impression of how those vowels tend to relate to the remaining Aromanian vowel space will follow (see Section 4.2), as well as an examination of effects attributable to specific syntactic and phonetic context (see Section 4.3). Finally, factors contributing to acoustic variation within the distribution of the vowels of interests are going to be investigated, with a particular focus on speakers from the Northern-Aromanian dialect region (see Section 4.4). Visualizations of the vowel data's most relevant aspects were created using the *ggplot2* package (see Wickham, 2016) developed for the R programming language (see R Core Team, 2020).

4.1 Degree of Overlap Between /ə/ and /ɨ/

On the basis of the remaining VOI dataset, Pillai scores for the contrasts between speaker-specific distributions of /ə/ and /ɨ/ were computed. They are represented by the orange bars in Figure 3 and are also given numerically in Table 7. It was verified that for each speaker and each vowel type, there were at least four observations. This is relevant to MANOVA's assumption of adequate sample size, according to which reliable effect-size measures can only be expected if each group contains more observations than the number of outcome variables (i.e. in this case, more than three, as there are three outcome variables F1, F2 and F3). If one were to control for duration by adding it as a covariate to the model from which the Pillai scores were extracted, one would get slightly higher results (see the grey bars in Figure 3). The differences, however, appear to be marginal, so, for the sake of simplicity, interpretation will be limited to the scores not controlled for duration.

It is interesting to observe that Northern female speaker WA_3, among all speakers, displayed by far the least overlap between /ə/ and /ɨ/, reflected by the considerably high Pillai score of .730. At the same time, none of the Southern speakers scored above .258, and the averages scores for Northern speakers of either sex (.353 for females, .213 for males) are also higher than those for their Southern counterparts (.205 for females, .135 for males). Caution is advised, though, before prematurely interpreting this as meaning that—in contradiction to the most common theoretical classifications—a phonemic contrast /ə/ and /ɨ/ is actually more relevant in Northern rather than Southern varieties.

One aspect immediately raises some suspicion regarding the interpretability of the observed differences between speaker-specific Pillai scores: The highest Pillai score was displayed by WA_3 who also happens to be the speaker with the fewest analysable observations (only 15, as given in Table 5). Complementary to that, the lowest Pillai score of just .059 can be attributed to MG_1 who is the speaker with the—by far—largest number of analysable observations in the VOI dataset (369, see Table 5), hinting at a possible correlation. And indeed, computing the Pearson correlation coefficient between all speaker-specific numbers of analysable observations and all speaker-specific Pillai scores gives –0.470, which suggests a medium-strength

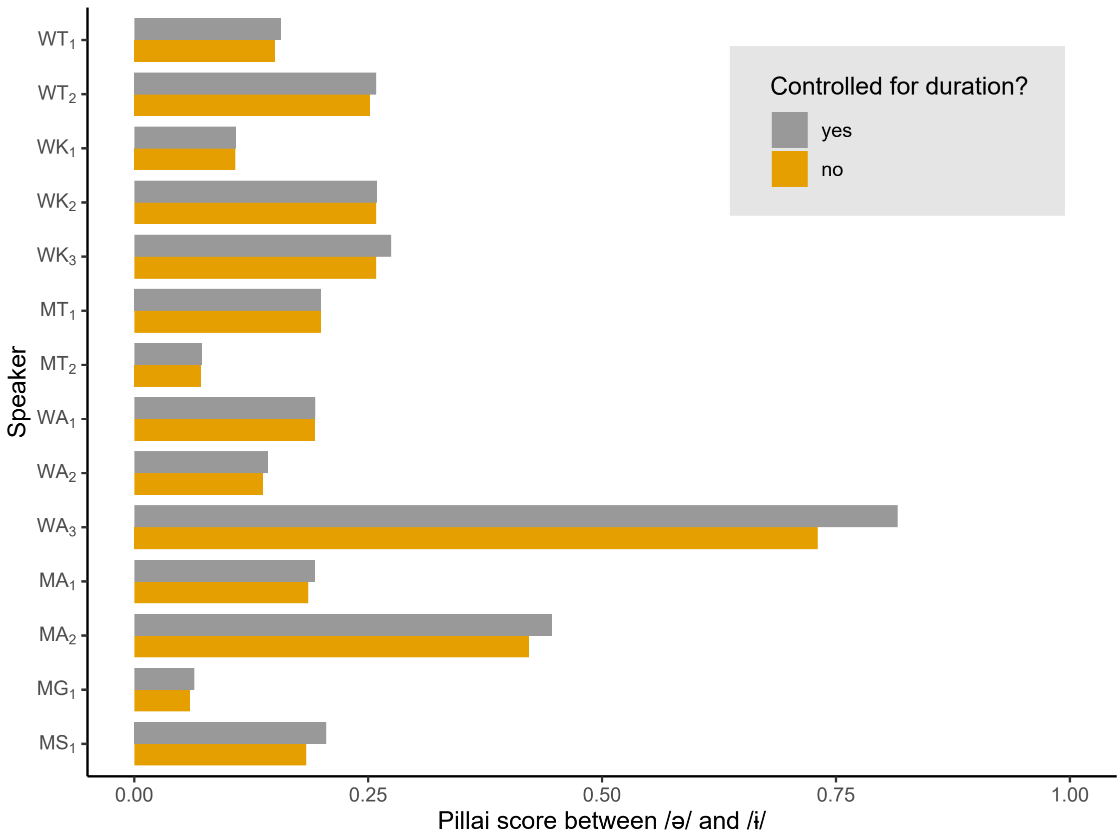

Figure 3: Bar chart of speaker-specific Pillai scores, quantifying the distinctness of /ə/ and /ɨ/ distributions. Orange bars are the default; the values indicated by the grey bars are marginally higher as they have been controlled for effects of duration.

negative association between those two measures. In order to make sure that the observed differences between Pillai scores were not just an artefact of the underlying sample sizes, the Pillai scores for the five vowel contrasts /a/ vs. /e/, /e/ vs. /i/, /i/ vs. /u/, /u/ vs. /o/ and /o/ vs. /a/ were also computed, for each speaker, on the basis of the AV dataset. They are given in Table 7 as well. By dividing a Pillai score for the contrast /ə/ vs. /ɨ/ by the mean score for the other five contrasts, one can obtain a measure that indicates how unusually relevant or irrelevant that contrast is when compared to the rather established phonemic contrasts between other vowel types within the speech of the same speaker. This is what the *Rel. %* column in Table 7 is about. Given like this, as 149.7 % and 15.8 %, the aforementioned extreme difference between the /ə/-vs.-/ɨ/ scores for speakers WA_3 and MG_1 does appear to be slightly less marked as it is now put into relation with the respective mean contrast scores, which are .488 for WA_3 and .375 for MG_1.

Nonetheless, the fact that WA_3 is the only one among all fourteen speakers whose /ə/-vs.-/ɨ/ contrast is associated with a Pillai score that is higher than the mean score for the five other contrasts, still makes it seem likely that there might be something amiss at least with that particular speaker's measurement. A more detailed

Region	Speaker	/a/–/e/	/e/–/i/	/i/–/u/	/u/–/o/	/o/–/a/	**Mean**	/ə/–/ɨ/	**Rel. (%)**
South	WT_1	.370	.358	.561	.099	.315	**.341**	.150	**44.1**
	WT_2	.538	.121	.791	.373	.736	**.512**	.252	**49.2**
	WK_1	.329	.234	.771	.151	.414	**.380**	.108	**28.5**
	WK_2	.556	.527	.621	.216	.670	**.518**	.258	**49.9**
	WK_3	.355	.206	.473	.316	.469	**.364**	.258	**71.1**
	MT_1	.281	.116	.396	.199	.350	**.268**	.199	**74.2**
	MT_2	.459	.251	.654	.035	.334	**.347**	.071	**20.5**
Avg. South Female		.430	.289	.643	.231	.521	**.423**	.205	**48.5**
Avg. South Male		.370	.184	.525	.117	.342	**.307**	.135	**47.4**
North	WA_1	.720	.137	.678	.281	.548	**.473**	.193	**40.8**
	WA_2	.524	.169	.568	.135	.775	**.434**	.137	**31.6**
	WA_3	.455	.287	.748	.275	.673	**.488**	.730	**149.7**
	MA_1	.489	.146	.739	.145	.298	**.363**	.186	**51.1**
	MA_2	.717	.077	.490	.376	.639	**.460**	.422	**91.7**
	MG_1	.564	.105	.617	.161	.428	**.375**	.059	**15.8**
	MS_1	.516	.207	.744	.193	.564	**.445**	.184	**41.3**
Avg. North Female		.566	.197	.664	.230	.666	**.465**	.353	**74.0**
Avg. North Male		.572	.134	.647	.219	.482	**.411**	.213	**50.0**

Table 7: Summary of Pillai scores computed for several vowel contrasts in the F1×F2×F3 space. The *Mean* column contains averages over the five preceding columns. The *Rel. (%)* column represents any Pillai score for the /ə/-vs.-/ɨ/ contrast in terms of the corresponding *Mean* value.

look into the underlying distribution of WA_3's 15 analysed vowel-of-interest realizations does, in fact, reveal a striking association between vowel type and stress: Although it is true that the vowels classified as /ə/ and /ɨ/ form two very obviously separate distributions on the F1×F2 plane, as it can be seen on the left-hand side of Figure 4, the plot on the right-hand side of Figure 4 reveals that the /ɨ/ tokens are exactly matching the ones that appeared in a stressed (or tonic) syllable. The reasonable suspicion that the effect observed here is rather one that can be attributed to stress than to vowel type is not only supported by the fact that Northern-Aromanian dialects are commonly assumed *not* to distinguish between /ə/ and /ɨ/, but also by the effect's direction: For a prototypical phonetic contrast between [ə] and [ɨ], one would expect higher F1 values to be attributable to tokens of [ə] and lower F1 values to tokens of [ɨ]. In the speech of WA_3, however, one can remark the opposite; furthermore, the distributions also seem to be distinguished by the F2 dimension, with higher F2 values corresponding to tokens of /ɨ/ rather than /ə/. From a diachronic perspective, it would be a bit hard to explain how an original /ɨ/ distribution shifted its mean from a historically original position around [ɨ] all the way to an open-mid front one around [ɛ] while leaving the other categories in the vowel space, and in particular /ə/, unaffected.

Some of the alleged /ɨ/ (or $/ə/_{\text{single}}$) tokens, e.g. the one in /ˈkɨne/ (or $/ˈkəne/_{\text{single}}$), fall into an F2 range around 14 Bark (i.e. about 2,300 Hz), which would be extremely

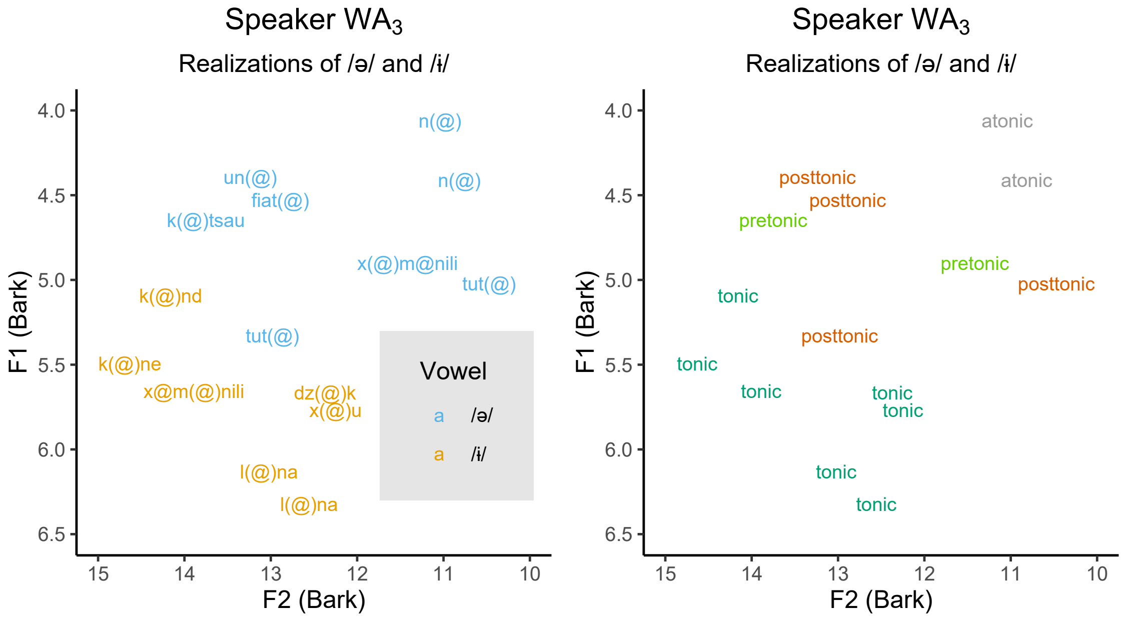

Figure 4: Locations of /ə/ and /ɨ/ realizations, produced by speaker WA_3, on the F1×F2 plane. The plot on the left displays the distinction between the two supposed vowel types as well as the surrounding word tokens transcribed in the SAMPA convention used for this study. The plot on the right informs about the same tokens' word-stress positions.

unusual for any type of central vowel. This is pointing at what was discussed in Section 3.2, namely that it is typical for speakers from the Northern dialect region to frequently produce an open-mid front allophone of the /ə/$_{\text{single}}$ vowel phoneme. Judging only by the somewhat sparse distribution plots for speaker WA_3, it seems plausible that this allophonic variation can be mainly attributed to vowel realizations in tonic word-stress positions. If and to what extent such an explanation through word stress would be sufficient and generalizable to other, less sparsely recorded Northern speakers, is going to be assessed and discussed further below.

	Atonic	Pretonic	Posttonic	Tonic	**Total**
/ə/	101	302	434	203	**1,040**
/ɨ/	16	40	17	334	**407**

Table 8: Contingency table of vowel type and word-stress position. The two cross-tabulated variables are not independent, as suggested by a chi-square test, with $p < .001$.

The key take-away from the inspection of the extreme example of Northern female speaker WA_3 is that word stress can both be heavily associated with the theoretical distinction between /ə/ and /ɨ/ and potentially responsible for large variation in the acoustic quality of vowel-of-interest observations. In order to get an impression of how vastly the discussed association is affecting not only WA_3, but the overall speech corpus, the contingency table of the variables *Vowel Type* and *Stress Position*, as retrieved from the entire VOI dataset (excluding outliers), is given as Table 8.

For atonic, pretonic and posttonic word-stress positions, the frequencies of /ə/ are 6 to 25 times larger than the ones of /ɨ/. Tonic positions are the only category for which the occurrences of /ɨ/ form the majority. A chi-square test confirms that vowel type and word-stress position are heavily interdependent, with $p < .001$. Thus, the Pillai scores for the /ə/-vs.-/ɨ/ contrast, reported earlier, are not only quantifying the overlap between distributions of the two vowel types, but to some degree also the overlap between stressed and unstressed vowels of interest. That could explain why the average scores paradoxically turned out to be higher for Northern rather than for Southern speakers: If Northern speakers are indeed generally inclined to produce $/ə/_{\text{single}}$ as [ɛ] in stressed syllables, but as [ə] in most other cases, as was observed for WA_3, it makes sense for this effect to manifest itself in the corresponding Pillai scores too.

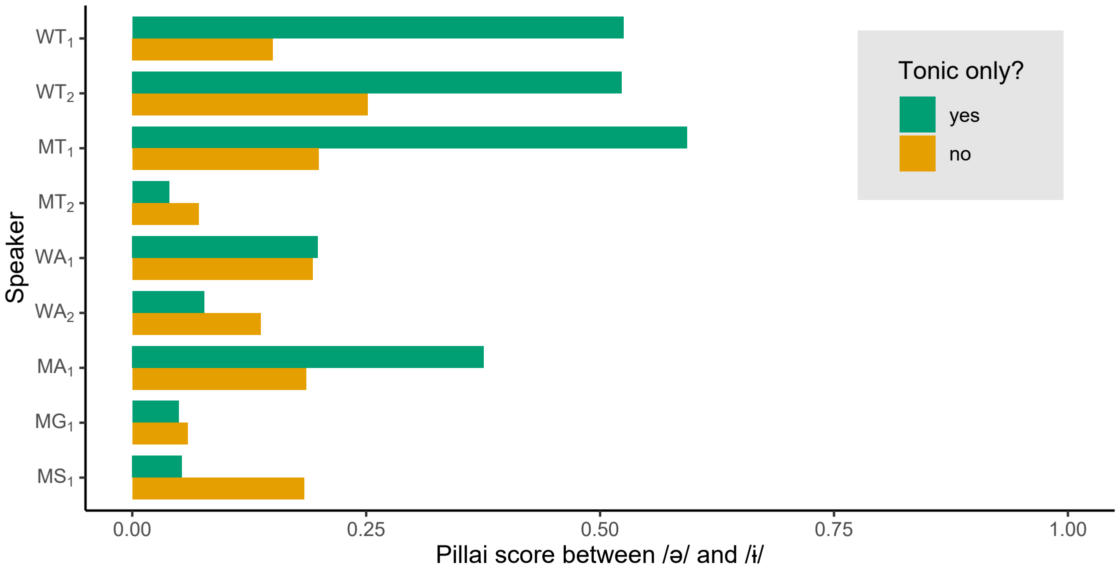

Figure 5: Bar chart of speaker-specific Pillai scores, quantifying the distinctness of /ə/ and /ɨ/ distributions. The values indicated by the green bars only consider vowel tokens in tonic word-stress positions.

Restricting the analysis only to vowel tokens in tonic positions would give a clearer insight into any potential contrast between observations classified as /ə/ and /ɨ/: For one thing, they are the only category for which the ratio beween the two vowel types is not extremely unbalanced. For another, unstressed vowels in spontaneous speech are likely to be reduced towards a schwa position anyway. Of the fourteen speakers featured in the VOI dataset, only nine fulfil the condition of providing at least four tonic observations of each vowel type. The five speakers who cannot be considered, in consequence, are WK_1, WK_2, WK_3, WA_3 and MA_2. Computing the Pillai scores between tonic /ə/ and /ɨ/ distributions, for the remaining nine speakers, yields the results represented by the green bars in Figure 5 and given numerically in Table 10's $F1 \times F2 \times F3$ column.

Here, the results look very different. Three of the four Southern speakers, WT_1, WT_2 and MT_1, who were displaying moderate-to-low Pillai scores of .150, .252 and .199 in the prior analysis, are now associated with scores of .525, .523 and .593; only for Southern speaker MT_2 the score stayed more or less in place, at a very low value, even decreasing a bit from .071 to .039. Among Northern speakers, the score noticeably increased only for MA_1, from prior .186 to .376 under the tonic-only condition. It stayed in place for speakers WA_1 (.193 to .198) and MG_1 (.059 to .050), but it showed a decrease for speakers WA_2 (.137 to .077) and MS_1 (.184 to .053). On average, the relationship observed earlier has been reversed: When only considering tonic vowel tokens, it is not the Northern, but the Southern speakers who display higher Pillai scores, and thus less overlap between their /ə/ and /ɨ/ distributions. Could this now be related to the common assumption of two separate mid-to-close central vowel phonemes, /ə/ and /ɨ/, in Southern-Aromanian dialects, but only a merged single one, $/ə/_{\text{single}}$, for the Northern-Aromanian ones?

Maybe, but before drawing generalizing conclusions, it should be kept in mind that for each dialect group, there was one speaker not falling in line with that trend, MT_1 in the South and MA_1 in the North. Examining the direction of the observed effects will also be necessary: For a *genuine* /ə/-vs.-/ɨ/ contrast, again, one would expect F1 to be the most relevant dimension. Table 9 contains three further columns with Pillai scores, apart from the rightmost one labelled $F1 \times F2 \times F3$, whose values are quantifying the overlap/distinctness of distributions within the three-dimensional space spanned by F1, F2 and F3. The values presented in the columns $F1 \times Dur.$, $F2 \times Dur.$ and $F3 \times Dur.$ are Pillai scores computed for determining the overlap/distinctness of distributions on the plane spanned by one of the formants and the duration variable (note that, here, duration is treated as another outcome variable instead of a covariate). On a quick sidenote on duration, only for two of all fourteen speakers, WA_2 and MG_2, a Welch's t-test pointed to a significant difference ($p < .05$) between the duration means of /ə/ and /ɨ/, with /ɨ/ being associated with longer durations on average.

By comparing, for example, a value for $F1 \times Dur.$ to one for $F2 \times Dur.$, it can be estimated whether F1 or F2 is a more relevant dimension for the overall distinction between the distributions of /ə/ and /ɨ/ in the vowel space. In each row of Table 9, the maximum value among those three formant-specific measures has been marked by underlining it. In congruence with theoretical expectations, a quick glance at the table reveals that among Southern speakers, the F1 is clearly the most distinctive formant for the /ə/-vs.-/ɨ/ contrast, whereas among Northern speakers—if such a contrast were to be assumed at all—it would rather be the third formant.

Region	Speaker	F1×Dur.	F2×Dur.	F3×Dur.	**F1×F2×F3**
South	WT_1	.507	.127	.054	**.525**
	WT_2	.408	.082	.016	**.523**
	MT_1	.580	.175	.071	**.593**
	MT_2	.156	.134	.153	**.039**
Avg. South Female		.457	.105	.035	**.524**
Avg. South Male		.368	.155	.112	**.316**
North	WA_1	.195	.169	.215	**.198**
	WA_2	.041	.034	.081	**.077**
	MA_1	.086	.045	.140	**.376**
	MG_1	.025	.031	.026	**.050**
	MS_1	.056	.086	.042	**.053**
Avg. North Female		.118	.101	.148	**.138**
Avg. North Male		.056	.054	.069	**.159**

Table 9: Summary of Pillai scores for the /ə/-vs.-/ɨ/ vowel contrast, among observations in tonic word-stress positions. The column names indicate which continuous variables were included as outcome variables in the linear model from which the Pillai scores were retrieved. In each row, the maximum value among *F1×Dur.*, *F2×Dur.* and *F3×Dur.* is underlined.

The plots in Figures 6, 7 and 8 aim at illustrating the overlap between distributions of tonic /ə/ and /ɨ/ realizations in the speech of the nine speakers listed in Table 9. For each vowel type, confidence ellipses plotted on top of the surrounding-word text labels indicate an area on the F1×F2 plane within which approximately 95 % of all observations would fall, assuming a normal distribution. Each pair of such ellipses provides some visual intuition for the degree of overlap between the corresponding distributions, but note that this intuition is less precise than the quantification through the Pillai score: Confidence ellipses are quite sensitive to off-bounds data points like the first realization of /ə/ in ⟨m@Ng@⟩ ≙ /ˈməŋgə/ that can be found in the upper right of the plot for Southern speaker WT_2 in Figure 6. That observation causes the /ə/ ellipse to get inflated to an extent that it covers a majority of the area enclosed by the competing /ɨ/ ellipse, even though only very few /ə/ observations can actually be found there.

For the most part, though, the plots do confirm what was already suggested by the Pillai scores collected in Table 9: Figure 6 shows that Southern speakers WT_1, WT_2 and MT_1 produce two quite distinct distributions of the vowels of interest, with the distribution of /ɨ/ being centred at a lower F1 value than the one of /ə/. On the other hand, speaker MT_2, who also was associated with the exceptionally low

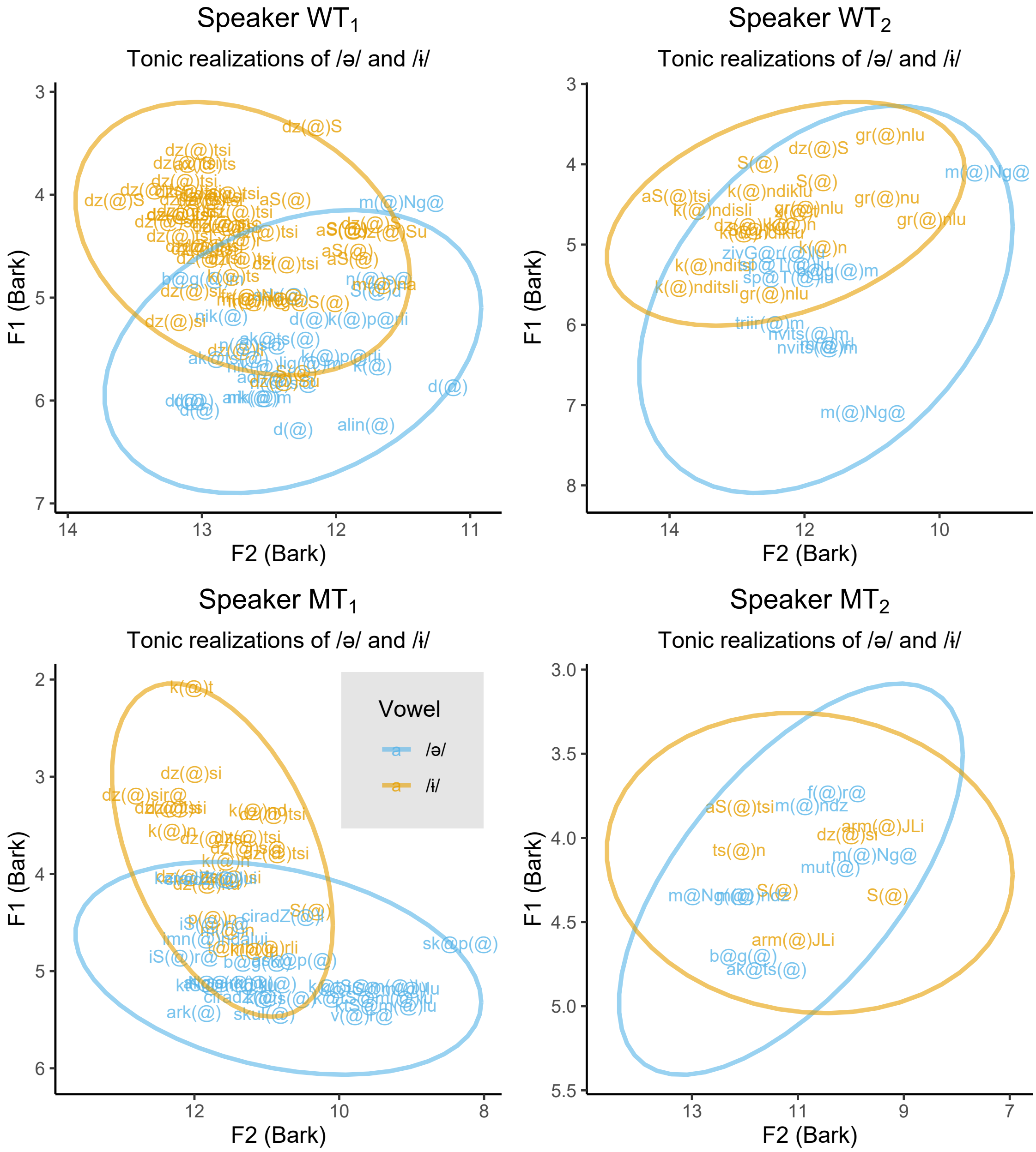

Figure 6: Four Southern speakers: Locations of tonic /ə/ and /ɨ/ realizations on the F1×F2 plane, marked by transcriptions of the words containing them. For each vowel type, confidence ellipses with a probability level of .95 have been added to the plots.

Pillai score of .039, displays two very overlapping distributions. It might be worth considering, though, that among the four Southern speakers featured here, MT_2 was also the speaker with the sparsest contribution to the vowel dataset. Perhaps further observations of his speech would have yielded a different result. As expected, high overlap can be identified in the plots for the five Northern speakers in Figures 7 and 8. This includes the plot for speaker MA_1 despite the rather high Pillai score of .376 that was computed for him. But as it has been established by the formant-specific measurements in Table 9, this should give no reason to wonder since the most distinctive dimension for his distributions was F3, which is not visualized

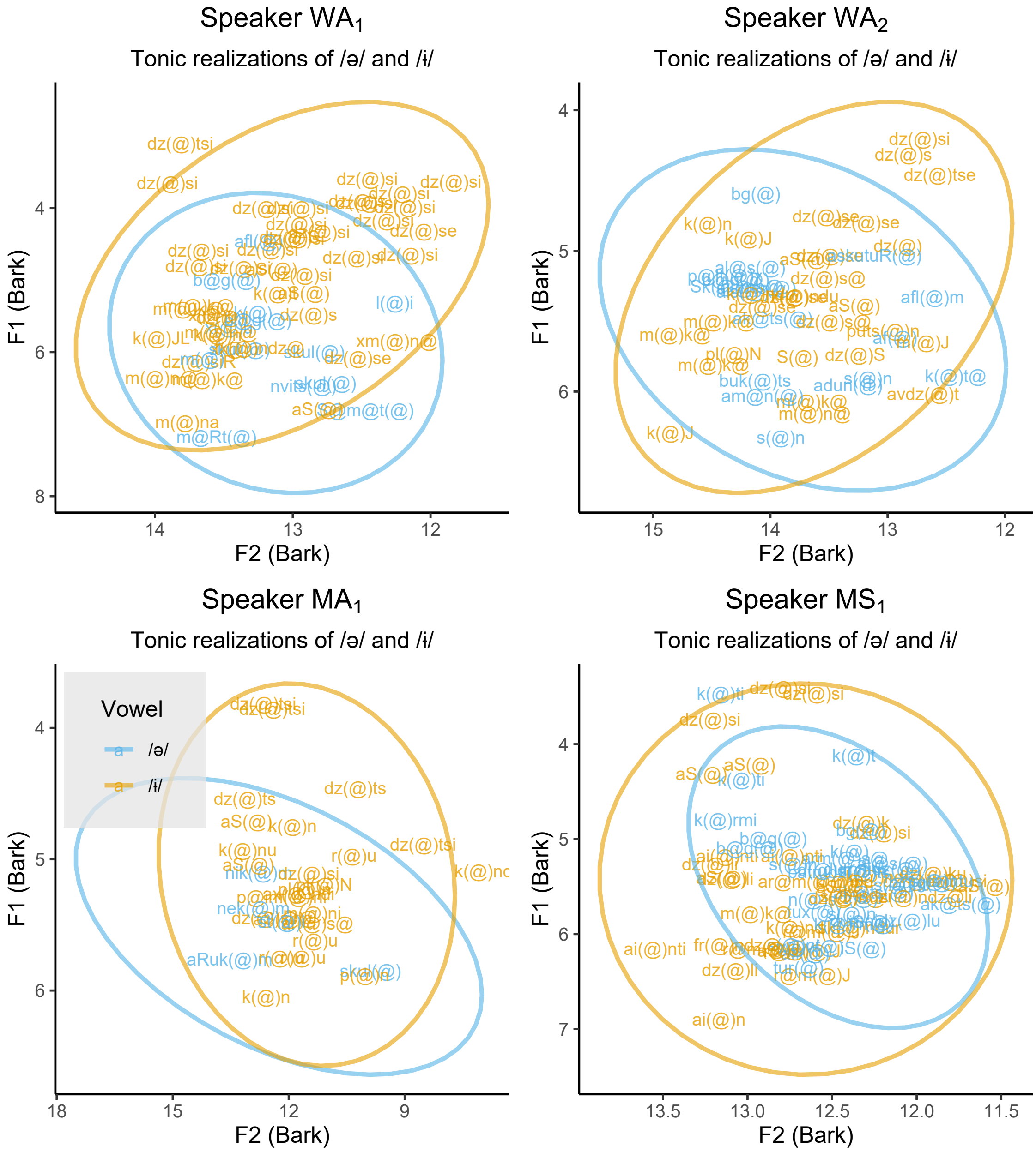

Figure 7: Four Northern speakers: Locations of tonic /ə/ and /ɨ/ realizations on the F1×F2 plane, marked by transcriptions of the words containing them. For each vowel type, confidence ellipses with a probability level of .95 have been added to the plots.

here. Another important observation, concerning almost all speakers from both dialect regions, is the noticeably frequent presence of inflected forms of the verb /ˈd͜zɨku/ (I say) in the lower F1 regions, i.e. visually upper regions in the plots, which includes word tokens like ⟨`dz@tsi`⟩, ⟨`dz@S`⟩, ⟨`dz@sir@`⟩ etc. This observation could hint at an effect of coarticulatory context and/or lexical category that has not yet been accounted for. The particular case of the extent of influence exerted by those verb forms is going to be investigated in more detail in Section 4.3.

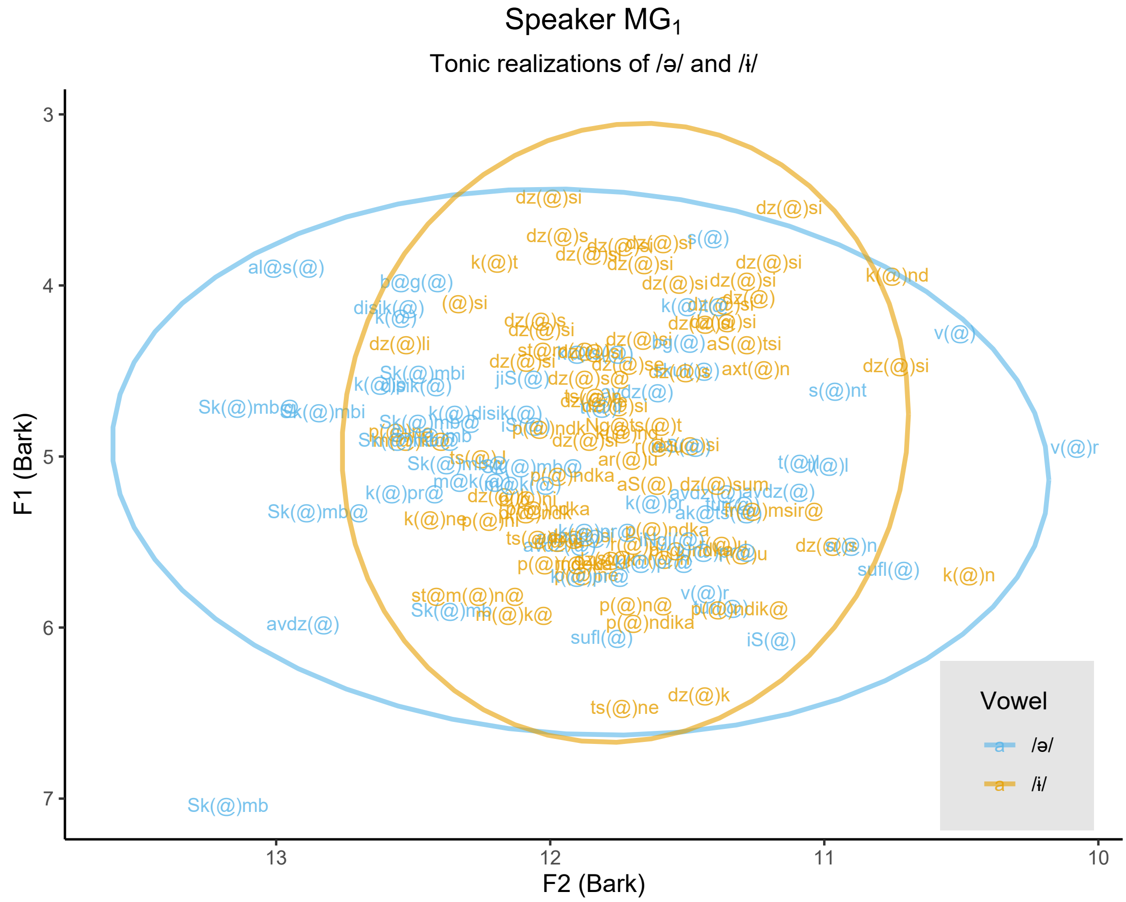

Figure 8: Northern speaker MG_1: Locations of tonic /ə/ and /ɨ/ realizations on the F1×F2 plane, marked by transcriptions of the words containing them. For each vowel type, confidence ellipses with a probability level of .95 have been added to the plot.

4.2 Estimated Vowel Space

Up until now, the focus has been lying on determining the degree of overlap between the distributions of the vowels of interest. It would also be insightful to get an impression of where their means are located within the broader Aromanian vowel space. On the basis of the datasets VOI and AV, the mean formant frequencies of any individual speaker and vowel type can be calculated. Further averaging over mean values that are associated with speakers who share the same indexical properties, i.e. identical dialect affiliation and sex, results in more general, speaker-independent estimates of vowel space for each such indexical group. The dispersion of those general vowel-type means on the F1×F2 plane has been plotted in Figure 9.

One observable difference across the parameter of speaker sex is the position of the /a/ vowel on the F2 scale. It is consistently higher for females than for males from both dialect regions, around 12 Bark (i.e. about 1,700 Hz) for females, but only around 11.5 Bark (i.e. about 1,600 Hz) for males. That difference is likely attributable to physiological differences across sex. There is a clear pattern that

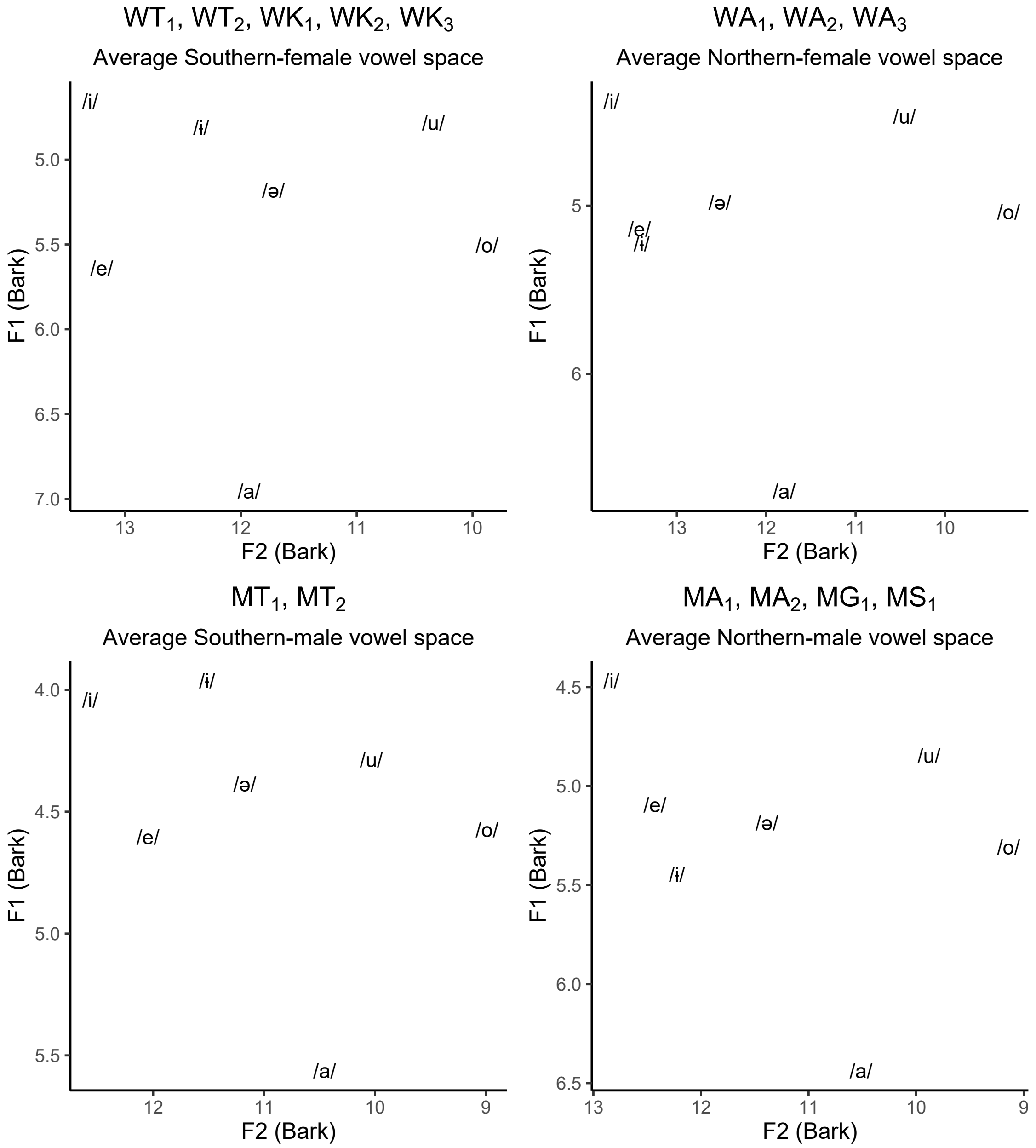

Figure 9: Plots displaying the mean points of each vowel type on the F1×F2 plane, grouped by the indexical speaker properties dialect affiliation and sex. The speaker labels above each plot indicate which speakers were considered. Averages have been calculated over the means of speaker-specific vowel distributions, so each speaker is given equal weight within any particular plot.

differentiates the Southern from the Northern vowel spaces: For Southern speakers, the mean of the vowel tokens classified as /ɨ/ is visually located somewhat above (i.e. lower F1) the one of /ə/, with a noticeable leftwards shift (i.e. higher F2) among females. In contrast to that, for Northern speakers, the vowel tokens labelled /ɨ/ appear to be distributed a bit below and to the left of (i.e. slightly higher F1 and higher F2) their /ə/ counterparts. This seems to be related to the observations made for speaker WA_3 based on the plots in Figure 4, Section 4.1: Tokens of /ɨ/ are substantially more likely to appear in stressed positions; such positions might in turn be associated with higher F1 and F2 values, judging at least from the distributions resulting from WA_3's speech production. Among Northern speakers, the means of both /ə/ and /ɨ/, which are likely not phonemically distinct in Northern dialects anyway, are located within close proximity to the one of /e/; for Northern females, the means of /ɨ/ and /e/ are even pretty much identical with F1 values of 5.22 and 5.14 as well as of F2 values of 13.39 and 13.41 respectively. That seems logical given that Northern-Aromanian realizations of the vowels of interest are know to often approximate the sound [ɛ]. It could even justify doubts about the acoustic and phonemic distinctness between $/ə/_{\text{single}}$ and /e/ for some speakers of the Farsherot varieties. Regarding the contrasts between the remaining vowel types, /a/ appears to be the most distinct one for all four speaker groups, when compared to any another sound in the vowel space. This is consistent with the high Pillai scores recorded for the /a/-vs./e/ and /a/-vs.-/o/ contrasts in Table 7 back in Section 4.1, ranging up to averages of .566 and .666 for Northern female speakers. The distinction between /e/ and /i/ is mainly constituted by a difference in F1 and slight difference in F2, with /e/ being associated with higher F1 and slightly lower F2 values. Interestingly, for the contrast between /o/ and /u/, the F2 dimension has a quite substantial relevance for some speaker groups: Tokens of /u/ are pronounced less back, on average, than tokens of /o/, which is most clearly the case for the group of male speakers from the Southern dialect region. Of course, they are also distinguished by F1, albeit a bit less than it is the case for the aforementioned /e/-vs./i/ contrast; the mean of /u/ tends to be located at somewhat higher (i.e. visually lower) values of F1 than that of /i/. In any case, it should not surprise that the contrasts /e/ vs. /i/ and /o/ vs. /u/ turn out to be less marked than other contrasts like /a/ vs. /e/ or /a/ vs./o/. That is because, in a handful of phonological contexts, the vowels in those two former contrast pairs are highly interchangeable, especially as the less prominent part of rising diphthongs like /e̯a/ and /o̯a/ that are mostly equivalent to /i̯a/ and /u̯a/. Interchangeability of /e/ and /i/ can also be observed whenever they appear as the coda of an unstressed last syllable of a word, e.g. in /ˈfrate/—/ˈfrati/ (brother), more rarely even in a few stressed contexts like /aˈestu/—/aˈistu/ (this / this one). It has been argued (see Caragiu-Marioțeanu, 1975; as cited in Schlösser, 1985) that, in cases like this, there is a tendency for /e/ variants to be a feature of the Northern, Farsherot varieties, while /i/ variants are more typical for the Southern varieties. However, as Schlösser (1985) points out, this tendency cannot be generalized to all subvarieties, and the transcriptions provided for the speech recordings this study is examining also suggest a high within-region variability in those regards, sometimes even within the speech of a single individual.

Figure 10 goes on to display equivalent plots as Figure 9, only that this time just the nine speakers listed in Table 9 are considered, i.e. the ones for whom the vowel-of-interest distributions under the tonic-only condition were not too sparse for either

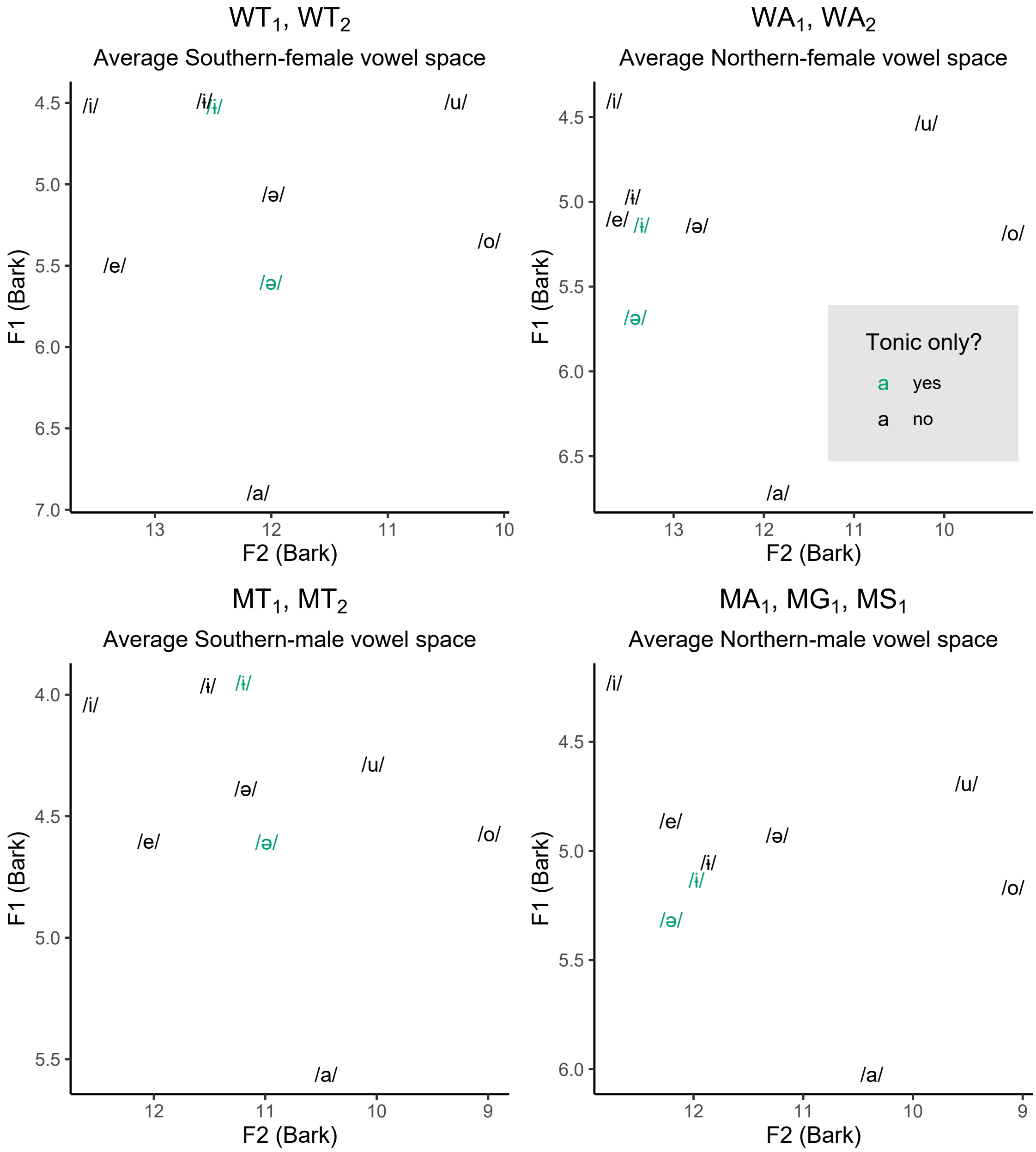

Figure 10: Plots displaying the mean points of each vowel type on the F1×F2 plane, grouped by the indexical speaker properties dialect affiliation and sex. The speaker labels above each plot indicate which speakers were considered. Averages have been calculated over the means of speaker-specific vowel distributions, so each speaker is given equal weight within any particular plot. For the vowels /ə/ and /ɨ/, the mean points of the distributions only containing tokens in tonic word-stress positions have also been plotted.

vowel type. Additionally, the locations of each vowel of interest, /ə/ and /ɨ/ have been plotted twice; once as determined by taking into account all available observations, once by only averaging over observations marked as tonic. The result is another confirmation of the preliminary conclusions drawn in Section 4.1: General and tonic variants of the /ɨ/ mean share roughly the same location among all four indexical speaker groups, which makes sense given that most realizations of /ɨ/ appear in tonic positions anyway. As opposed to that, one can find the tonic variant of the mean of /ə/ at an about 0.54 Bark higher (visually lower) point on the F1 scale for Southern female speakers, and about 0.21 Bark higher for Southern males. This is matching the observation that lower degrees of distribution overlap, indicated by higher Pillai scores, were recorded between tonic /ə/ and /ɨ/ among Southern speakers. On the Northern side, notably, the tonic /ə/ mean displays a substantial shift towards higher F1 and F2 values, compared to its general counterpart. Being located at F1 and F2 values of 4.61 and 10.99 for males as well as 5.60 and 12.00 for females, it reaches an even more open and fronted position than the corresponding mean point for /ɨ/. A possible interpretation of this phenomenon would be that especially stressed realizations of the Northern-Aromanian vowel phoneme $/ə/_{\text{single}}$ tend to approximate the acoustic quality of [ɛ], and even more so whenever they appear in contexts that would be attributed to the phoneme /ə/ rather than /ɨ/ in the Southern dialect group.

4.3 Impact of Syntactic and Phonetic Context

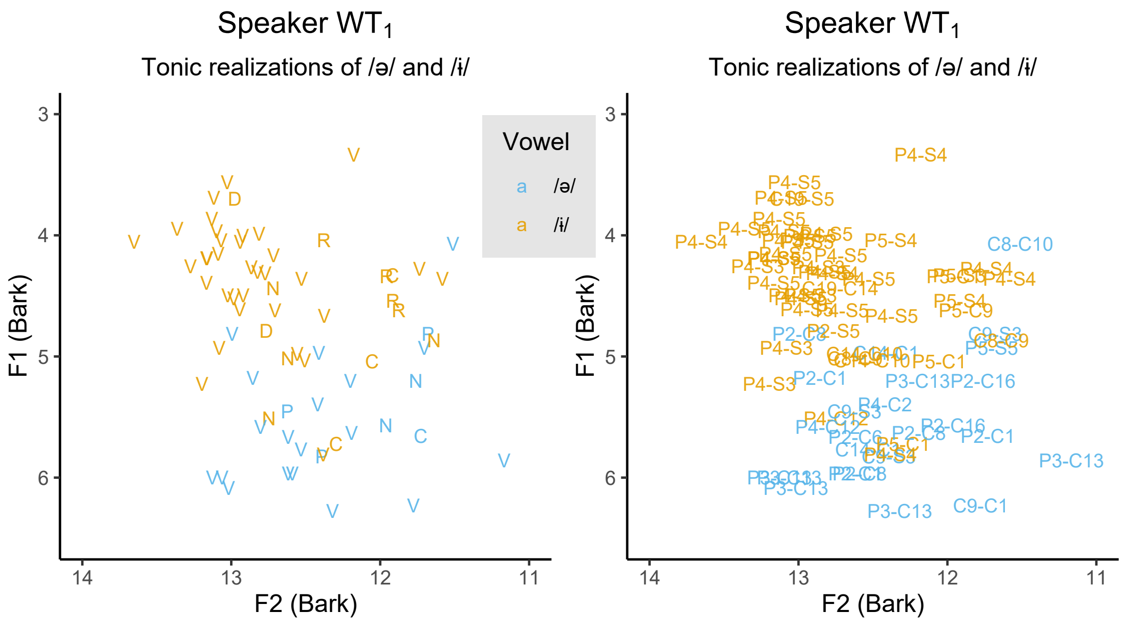

Figure 11: Southern speaker WT_1: Locations of tonic /ə/ and /ɨ/ realizations on the F1×F2 plane. They are represented by text labels with POS tags of the words containing them in the plot on the left side, and with tags representing their preceding and succeeding contexts of coarticulation in the plot on the right side.

Coming back to the issue regarding the frequent presence of /ˈd͡zɨku/ verb forms in low-F1 regions, observed in Figures 6, 7 and 8, it could be interesting to see how much a specific combination of syntactic and phonetic parameters can affect the acoustic quality of a vowel of interest. The two plots in Figure 11 are based on the same data as the plot drawn for speaker WT_1 in Figure 6; here, however, the text labels are informing about the POS tag and the preceding as well as succeeding context of coarticulation associated with each vowel observation. Coarticulatory contexts are denoted as defined in Section 3.5.1. For instance, a label like *P4-S4* indicates that the corresponding vowel observation was preceded by a [z], [d͜z], [s] or [t͜s] segment and succeeded by a [ʒ] or [ʃ] segment. Vowel tokens appearing in the stem of inflected forms of the verb /ˈd͜zɨku/ all share the same preceding context, *P4*, and POS tag, *V*. They are responsible for most of the *P4* contexts that can be discerned in the upper-left edges of the plot on the right side of Figure 11.

	N	*V*	*A*	*P*	*D*	*R*	*S*	*C*	*M*	*T*	*I*	*Q*	**Total**
Tonic /ə/	49	117	4	12	3	0	6	7	1	0	4	0	**203**
Tonic /ɨ/	93	162	1	3	7	30	3	35	0	0	0	0	**334**

Table 10: Contingency table of vowel type and POS tag, only considering vowels in tonic word-stress positions. The two cross-tabulated variables are not independent, as suggested by a Monte-Carlo-simulated Fisher's exact test, with $p < .001$.

	C1 Sp. pause	*C2* [a]	*C3* [e]	*C4* [i]	*C5* [o]	*C6* [u]	*C7* [ɛ]-[ə]-[ɨ]	*C8* [m]	*C9* [n]
Tonic /ə/	0	0	0	0	0	0	0	19	9
Tonic /ɨ/	0	0	0	5	0	0	0	35	0

	C10 [ŋ]	*C11* [ɲ]	*C12* [l]	*C13* [ʎ]	*C14* [r]	*C15* [h]	*C16* [b]-[p]	*C17* [v]-[f]	*C18* [ð]-[θ]
Tonic /ə/	0	0	15	0	7	0	2	8	2
Tonic /ɨ/	0	0	5	0	28	0	18	0	0

	C19 [ʁ]-[ɣ]-[x]	*P1* [j]-[ç]-[c]	*P2* [g]-[k]	*P3* [d]-[t]	*P4* [z]-[d͜z]-[s]-[t͜s]	*P5* [ʒ]-[d͜ʒ]-[ʃ]-[t͜ʃ]	**Total**
Tonic /ə/	3	1	71	16	37	13	**203**
Tonic /ɨ/	8	0	41	1	156	37	**334**

Table 11: Contingency table of vowel type and preceding context of coarticulation, only considering vowels in tonic word-stress positions. The two cross-tabulated variables are not independent, as suggested by a Monte-Carlo-simulated Fisher's exact test, with $p < .001$.

At this point, before further analysing this particular case, some general impression of the dependence or independence between the variables vowel type and POS tag as well as vowel type and preceding or succeeding context of coarticulation would be helpful. With consideration still restricted to the tonic subset of the VOI dataset, contingency tables for those variable pairs are given as Tables 10, 11 and 12. It can be remarked that the data in all three tables, especially the ones regarding coarticulation, is quite sparse, with some categories being associated only with very few or even no occurrences. For that reason, it is not possible to apply a chi-square

	C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	*C9*
	Sp. pause	[a]	[e]	[i]	[o]	[u]	[ɛ]-[ə]-[ɨ]	[m]	[n]
Tonic /ə/	18	6	1	8	0	5	0	27	15
Tonic /ɨ/	13	4	1	1	0	15	1	1	99
	C10	*C11*	*C12*	*C13*	*C14*	*C15*	*C16*	*C17*	*C18*
	[ŋ]	[ɲ]	[l]	[ʎ]	[r]	[h]	[b]-[p]	[v]-[f]	[ð]-[θ]
Tonic /ə/	5	3	19	6	15	0	23	5	1
Tonic /ɨ/	4	15	9	0	2	0	0	3	0
	C19	*S1*	*S2*	*S3*	*S4*			*S5*	
	[ʁ]-[ɣ]-[x]	[g]-[k]-[c]	[ʝ]-[ç]	[z]-[s]	[ʒ]-[ʃ]			[d]-[d͜z]-[d͜ʒ]-[t]-[t͜s]-[t͜ʃ]	**Total**
Tonic /ə/	3	3	0	11	2			27	**203**
Tonic /ɨ/	0	19	0	85	12			50	**334**

Table 12: Contingency table of vowel type and succeeding context of coarticulation, only considering vowels in tonic word-stress positions. The two cross-tabulated variables are not independent, as suggested by a Monte-Carlo-simulated Fisher's exact test, with $p < .001$.

test here; a Fisher's exact test can be used instead. A Monte Carlo simulation of such a test reveals that for all three tables, the examined variables are interdependent, with $p < .001$ in each case. Table 10 shows that verbs are the most frequent one among the 12 lexical categories represented by POS tags, followed by nouns and conjunctions. The ratio between /ə/ and /ɨ/ is more balanced among verbs than within the entire tonic subdataset, while it is even more uneven among nouns. Only for less prominent categories like pronouns, adjectives and adpositions, the general trend of tonic /ɨ/ tokens outnumbering their /ə/ counterparts was reversed. As can be read off from Table 11, alveolar fricatives and affricates, labelled *P4*, appear to be the most frequent preceding context for tonic /ɨ/ realizations, accounting for almost half of all observations for that vowel type. For /ə/, however, it is velar plosives, labelled *P2*. Now, with respect to Table 12, it can be stated that tonic /ɨ/ tokens appear most frequently before an alveolar nasal or fricative, i.e. *C9* or *S3*, which already covers more than half of all such observations. By contrast, /ə/ tokens do not follow that trend at all as their most frequent succeeding contexts are bilabial nasals and plosives, *C8* and *C16*, although those only make up for roughly a quarter of all tonic /ə/ observations. Overall, the contingency tables given here make it clear that the two vowel-of-interest types are distributed across two very different ranges of coarticulatory contexts, on average. Their distributions across lexical categories, each associated with different syntactic roles, are also differing significantly, although that relationship is a bit harder to interpret as the two most prominent categories, nouns and verbs, already account for almost 80 percent of all tonic vowel-of-interest observations. For the assumption of two distinct vowel *phonemes* /ə/ and /ɨ/, this aspect could become a little challenging. If context already serves as a significant predictor of which of the two types any vowel-of-interest token can be associated with, there would not be much room left for a vast number of cases constituting minimal pairs like the ones mentioned in Section 2.1.

If one were to go on to cross-tabulate the variables POS tag, preceding context and succeeding context against each other pair-wise, one would find that among all possible two-way combinations, the group of tonic tokens associated with preceding *P4* context and the POS tag *V*, i.e. verb forms in which a stressed vowel of interest is preceded by an alveolar fricative or affricate, form the largest group, with 172 occurrences. Coincidentally, this is also the combination that has already been suspected, based on visual intuition from the plot in Figure 11, to be responsible for having a heavy influence on the results. Among those occurrences, vowel type /ɨ/ appears 143 times, type /ə/ only 33 times. In fact, 133 of the 172 occurrences are inflected forms of the aforementioned verb /ˈd͜zɨku/.

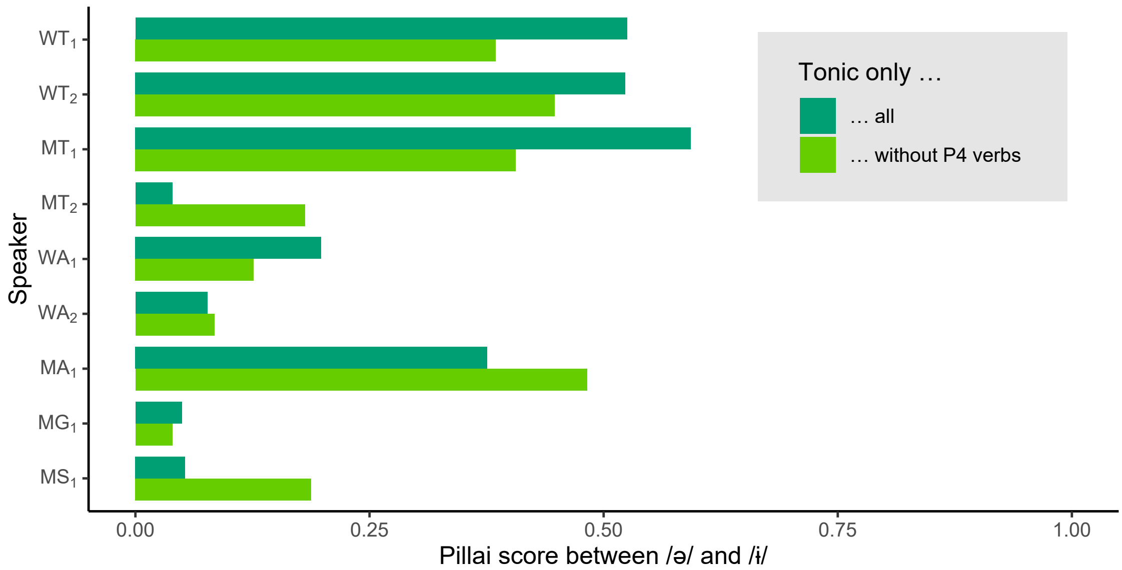

Figure 12: Bar chart of speaker-specific Pillai scores, quantifying the distinctness of tonic /ə/ and /ɨ/ distributions. The values indicated by the light-green bars only consider vowel tokens that did not appear both within a verb and right after a velar-fricative or velar-affricate segment.

How impactful is that largest two-way group then, consisting mostly of verb forms of a single lemma, on the results presented in Section 4.1? After reducing the tonic subdataset even further, so that it only encompasses observations not belonging to the *P4* & *V* group, the assumption of there being at least four observations for each vowel type among the nine speakers listed in Table 9, is still met. So, it is possible to compute Pillai scores again, and see how they are going to compare to the ones collected previously. The results are given in Figure 12. Effectively, the first three Southern speakers listed, WT_1, WT_2 and MT_1 now display more overlap between their two vowel-of-interest distributions, with Pillai scores ranging only from .385 to .448 instead of the earlier .523 to .593 that included *P4* & *V* tokens. Southern

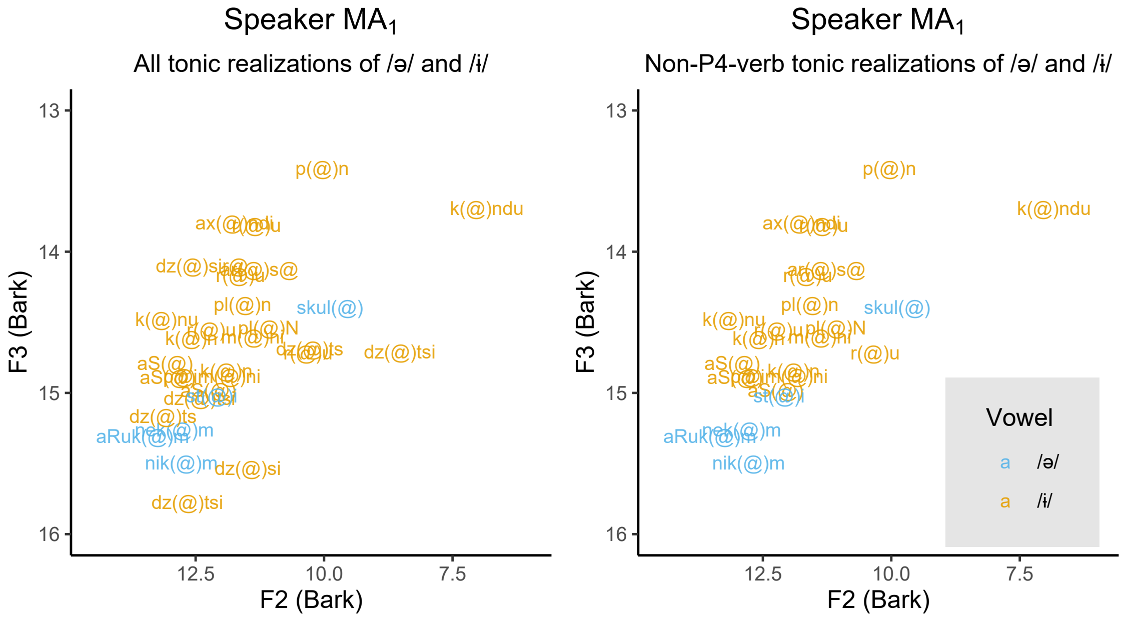

Figure 13: Northern speaker MA_1: Locations of tonic /ə/ and /ɨ/ realizations on the F3×F2 plane, marked by transcriptions of the words containing them. The plot on the right only considers vowel tokens that did not appear both within a verb and right after a velar-fricative or velar-affricate segment.

speaker MT_2 and Northern speaker MS_1 saw their Pillai scores rise from the quite low .039 and .053 up to .181 and .188, once that two-way group was excluded. While the scores for WA_2 and MG_1 stayed in place at relatively low values and showed a medium decrease for speaker WA_1, it is quite remarkable that the Pillai score for speaker MA_1, who—among the Northern dialect group—had already displayed the highest score under the all-tonics condition, saw his score go up even further here, from .376 to .482. That renders him the speaker associated with the most distinct /ə/ and /ɨ/ distributions under this particular condition. Only by looking at his F1×F2 plot in Figure 7, such an increase would not be very much expectable as most of the /ɨ/ tokens appearing in somewhat distinct positions are, in fact, verb forms of /ˈd͜zɨku/. But again, it has been established that the most distinctive formant for MA_1's tonic distributions is actually F3 (see Table 9), so it is not far-fetched to assume this will still be the case after excluding the *P4 & V* group. Figure 13 sheds some light onto this aspect as it shows distributions on the F3×F2 plane for tonic vowel-of-interest realizations uttered by MA_1. On the left side, all tokens are included, whereas the plot on the right side displays just those observations that were not classified as belonging to *P4 & V*. Comparing the two plots, it can easily be seen that the speaker in question tended to produce the /ɨ/ in /ˈd͜zɨku/ verb forms predominantly in high-F3 positions, somewhat close to the mean of the rather sparse distribution of tonic /ə/ tokens. Once those verb forms were removed, the F3-based distinction between /ə/ and /ɨ/ became even clearer, hence also resulting in a higher Pillai score.

All in all, the suspicion that *P4* verb forms, and inflected forms of /ˈd͡zɨku/ in particular, exert a substantial influence on the degree of overlap between distributions of tonic /ə/ and /ɨ/, was clearly justified. This specific example serves as a general reminder that context effects like this can manifest themselves as confounds in measures of vowel-distribution overlap. The fact that, for three Southern speakers, the distinctness between the two vowel types became less marked once tokens appearing in *P4* verb forms were excluded, could raise some doubts about the robustness of the observed results and more importantly of the proposed conclusion that /ə/ and /ɨ/ are, in fact, distinct in Southern Aromanian. Nonetheless, even under the condition excluding those verb forms, it was still the case that the average scores for Southern speakers were higher than the ones for Northern speakers, with .416 versus .105 for females and .294 versus .236 for males.

4.4 Variation of $/ə/_{\text{single}}$ in the Northern Varieties

This section is focusing on the factors that contribute to the acoustic variation of Northern-Aromanian vowels of interest. As previously stated, those are commonly assumed to form only one phoneme, denoted as $/ə/_{\text{single}}$ for the purposes of this text. Nevertheless, the results presented in Sections 4.1 and 4.2 also confirmed that the vowel space covered by that phoneme is rather large, reaching more open and fronted realizations around [ɛ] especially in stressed syllables.

The goal is to examine to what precise extent each of the five nominal variables stored in the VOI dataset, namely vowel type, POS tag, preceding and succeeding context as well as word-stress position, is responsible for the variance of the three continuous variables F1, F2 and F3, given in Bark scale. When trying to fit a linear model—containing the five nominal variables as predictors and the three continuous ones as outcome variables—to each of the fourteen speaker-specific subdatasets, it turns out that just the four speakers MG_1, MS_1, WA_2 and MT_1 provide data that is not too sparse to compute coefficients for each category of each nominal variable. Because of that, but also in order to keep the results overseeable, this part of the analysis is taking into account only the speech of those four speakers, among whom one can find three from the Northern dialect region and only one from the Southern one. As the object of examination is going to be the variation of acoustic quality for the Northern-Aromanian phoneme $/ə/_{\text{single}}$, this ratio is quite convenient. Southern speaker MT_1 is going to be included anyway, and can serve as a sort of control for verifying whether any common patterns observed for the three Northern speakers are exclusive to their dialect region or not.

To each of the four resulting, speaker-specific linear models, a MANOVA test was applied. As the data dealt with in this study is unbalanced, the type II sum-of-squares approach was chosen. Note that type II and III are equivalent here, as no higher-order factors are included in the models. While it was possible to compute meaningful-looking results like that, it should be pointed out that the assumption of adequate sample size is being violated here: Among the variables preceding context, succeeding context and POS tag, there were a lot of categorical groups only containing between zero and three entries, i.e. less than or equal to the number of outcome variables. In the case of zero entries, this is rather unproblematic as such categories are simply going to be treated as non-existent by the test. For values between one and three, this might have a distorting effect on the test results since the presence of such groups is known to make the significance estimation more conservative, meaning that some significant effects might either not be detected or at least judged as more moderate. Therefore, an additional test was run, as a sort of sanity check, under slightly different circumstances. For that additional test, only the first two formants were included as outcome variables in the linear models, thus effectively decreasing the threshold of what would be considered a too sparse categorical group from three to only two entries and below. Then, each of the four speaker-specific subdatasets underwent a sifting process with the aim to exclude observations belonging to some group with only one or two entries, until a state was reached where, for all nominal variables, all groups were either empty or containing at least three vowel observations. As a result, 10 observations were removed from speaker MG_1's data, 18 for speaker MS_1, 20 for speaker WA_2 and 7 for speaker MT_1. By running a MANOVA test now, on the models that were reduced as described, it has been assured that the assumption of adequate sample size is not being violated; as a trade-off, one dimension of acoustic quality, F3, as well as a handful of vowel observations had to be excluded. One could refer to this latter kind of test as test B, and to the original one, which—under the violation of the adequate-sample-size assumption—did not make the aforementioned reductions to dataset size and number of outcome variables, as test A. The further assumptions of multivariate normality and homogeneity of covariances were often, but not always met. For practical purposes, that should not matter too much, though, as MANOVA is considered to be a relatively robust procedure even in cases of moderate violations of those assumptions (see Weinfurt, 1995).

Tables 13, 14, 15 and 16 show the results of the MANOVA tests conducted for speakers MG_1, MS_1, WA_2 and MT_1 respectively. Pillai scores (a.k.a. Pillai's trace values) have been chosen as the test statistic, although they would show slightly different figures for vowel type than the ones collected in the earlier sections: That is because here they have been computed for models with multiple categorical predictor variables, which means the effects of those variables are being controlled for each other. On the basis of the resulting Pillai's trace values, partial η^2 measures were obtained, a metric for effect size that can be interpreted as how much variation in the multivariate outcome each predictor variable can account for.

***MG₁*, 369 Obs.**	Partial η^2		Approx. F		Effect df		Error df		$Pr(>F)$	
Test	A	B	A	B	A	B	A	B	A	B
Vowel Type	.034	.021	3.74	3.44	3	2	322	320	.0114	.0331
POS Tag	.066	.073	2.55	3.17	27	16	972	642	$<.0001$	$<.0001$
Prec. Context	.164	.149	4.22	5.12	45	22	972	642	$<.0001$	$<.0001$
Succ. Context	.185	.177	4.60	4.93	48	28	972	642	$<.0001$	$<.0001$
Stress Position	.077	.074	9.03	8.49	9	6	972	642	$<.0001$	$<.0001$

Table 13: Speaker MG_1: MANOVA table for tests A and B, as defined above.

***MS₁*, 145 Obs.**	Partial η^2		Approx. F		Effect df		Error df		$Pr(>F)$	
Test	A	B	A	B	A	B	A	B	A	B
Vowel Type	.056	.039	1.93	1.87	3	2	98	92	.1302	.1593
POS Tag	.124	.139	1.58	2.15	27	14	300	186	.0374	.0111
Prec. Context	.284	.187	3.05	2.14	39	20	300	186	$<.0001$	.0046
Succ. Context	.318	.226	2.59	2.27	54	24	300	186	$<.0001$	.0012
Stress Position	.142	.194	5.51	7.47	9	6	300	186	$<.0001$	$<.0001$

Table 14: Speaker MS_1: MANOVA table for tests A and B, as defined above.

***WA₂*, 116 Obs.**	Partial η^2		Approx. F		Effect df		Error df		$Pr(>F)$	
Test	A	B	A	B	A	B	A	B	A	B
Vowel Type	.074	.044	1.94	1.55	3	2	73	67	.1308	.2199
POS Tag	.100	.115	1.04	1.76	24	10	225	136	.4164	.0736
Prec. Context	.211	.194	2.00	2.33	30	14	225	136	.0024	.0065
Succ. Context	.317	.285	1.94	2.46	54	22	225	136	.0004	.0008
Stress Position	.130	.172	3.72	4.72	9	6	225	136	.0002	.0002

Table 15: Speaker WA_2: MANOVA table for tests A and B, as defined above.

***MT₁*, 176 Obs.**	Partial η^2		Approx. F		Effect df		Error df		$Pr(>F)$	
Test	A	B	A	B	A	B	A	B	A	B
Vowel Type	.102	.101	5.11	7.54	3	2	135	134	.0022	.0008
POS Tag	.081	.113	1.72	2.45	21	14	411	270	.0249	.0029
Prec. Context	.272	.328	4.27	7.33	36	18	411	270	$<.0001$	$<.0001$
Succ. Context	.219	.260	2.56	3.65	45	26	411	270	$<.0001$	$<.0001$
Stress Position	.074	.107	3.66	5.38	9	6	411	270	.0002	$<.0001$

Table 16: Speaker MT_1: MANOVA table for tests A and B, as defined above.

It can be observed that coarticulatory context is associated with the largest effect sizes, as indicated by partial η^2 values from test A ranging between .164 and .284 for preceding context and between .185 and .318 for succeeding context. The corresponding values from test B display the same overall trend. Only in the speech excerpt produced and provided by Southern speaker MT_1, it is preceding rather than succeeding context to have more weight in accounting for variation in vowel quality. Relative to the other variables, word-stress position can be associated with a medium effect size; its partial η^2 values are ranging between .077 and .142 for test A as well as .074 and .194 for test B. The binary vowel-type distinction based on ideal Southern-Aromanian phonology is associated with the smallest effect sizes on average, being assigned partial η^2 values from .034 to .102 and from .021 to .101 for test A and B respectively. It makes sense that the maximum value of .102/.101 is attributed to the only Southern speaker featured here, MT_1. When considering a conventional significance threshold like $p < .05$, test A and test B lead to the same judgement on the presence of significance, in all examined cases, even though sometimes the difference between them can be quite extreme, most notably with regard to the POS-tag variable in the tests conducted for speaker WA_2; $p = .4164$ for test A, but merely $p = .0736$ for test B. The measured effects for the phonetic parameters preceding and succeeding context as well as word stress are statistically significant for all four speaker-specific subdatasets and both types of tests. POS tags show less significant effects, with the comparatively small set of observations for speaker WA_2 displaying non-significant p-values, as was just mentioned. Effects of vowel-type distinction were the least significant ones, only passing below the $p < .05$ threshold for the two comparatively large subdatasets encompassing vowel observations from the speech of speakers MG_1 and MT_1 respectively.

The effects of vowel type have already been dealt with extensively in Sections 4.1–4.3. Although the variables for POS tag and coarticulation context have been associated with rather large effect sizes, further examination could not reveal any systematic relationship or pattern, generalizable beyond the speech of individual speakers, as to which specific contexts were exerting an effect, and in what direction within the vowel space. For word-stress position, on the other hand, the earlier suspicion that Northern-Aromanian /ə/$_{\text{single}}$ tokens are likely to be realized as more fronted, in stressed positions in particular, can clearly be confirmed by the impression of the plots in Figure 14: For stressed observations, all three included Northern speakers show a more concentrated and dense realization space, with a mean clearly shifted towards higher F2 values. For speakers MS_1 and WA_2, this trend goes hand in hand with somewhat higher F1 values as well, corresponding to more open vowel realizations. By contrast, for Southern speaker MT_1, the distributions appear to be highly overlapping. Figure 15 shows analogous plots, but this time for the distinction between pretonic and posttonic word-stress positions. Here, there does not seem

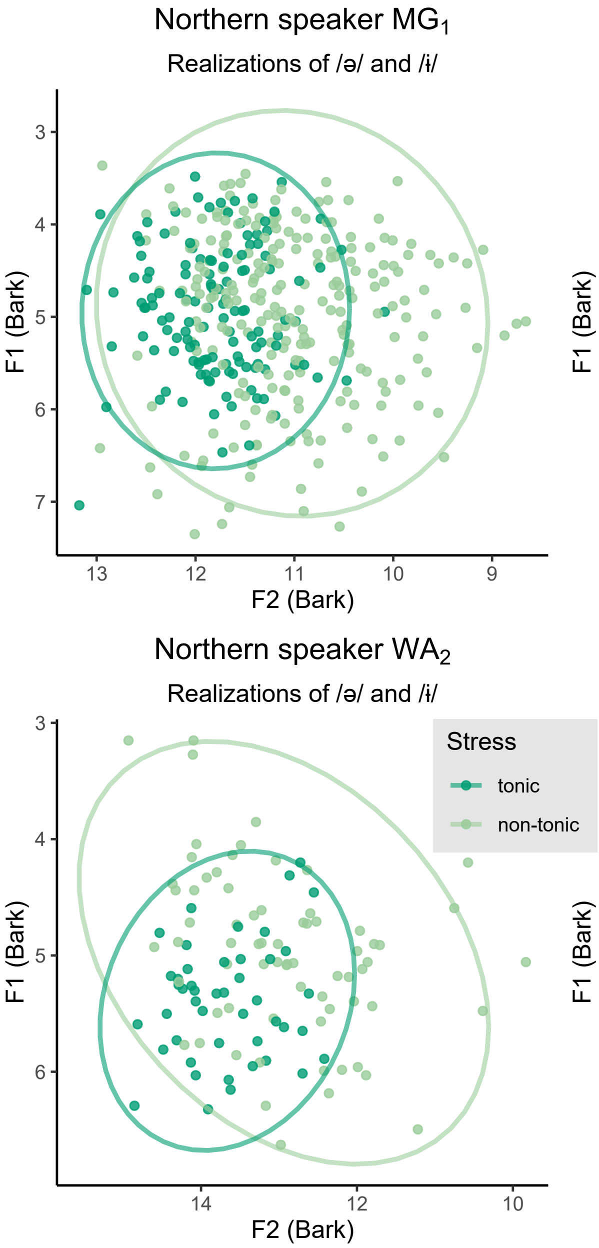

Figure 14: Three Northern and one Southern speaker: Scatter plots of the distributions of /ə/$_{\text{single}}$ (a.k.a. /ə/ and /ɨ/) realizations, grouped by whether word-stress position was tonic or not. Confidence ellipses with a probability level of .95 have been added to the plots.

to be much difference, except maybe for the subtle observation that among the three Northern speakers, posttonic tokens take up a somewhat more narrow interval on the F2 scale. Apart from that, pretonic and posttonic distributions are highly overlapping, mostly so, again, for Southern speaker MT_1. On average, atonic tokens were displaying similar behaviour as pretonic and posttonic ones, but were left out of the plots, for simplicity, as their comparatively sparse numbers of occurrences would render quite unintuitive confidence ellipses.

On the whole, it can be stated that the acoustic quality of the Northern-Aromanian

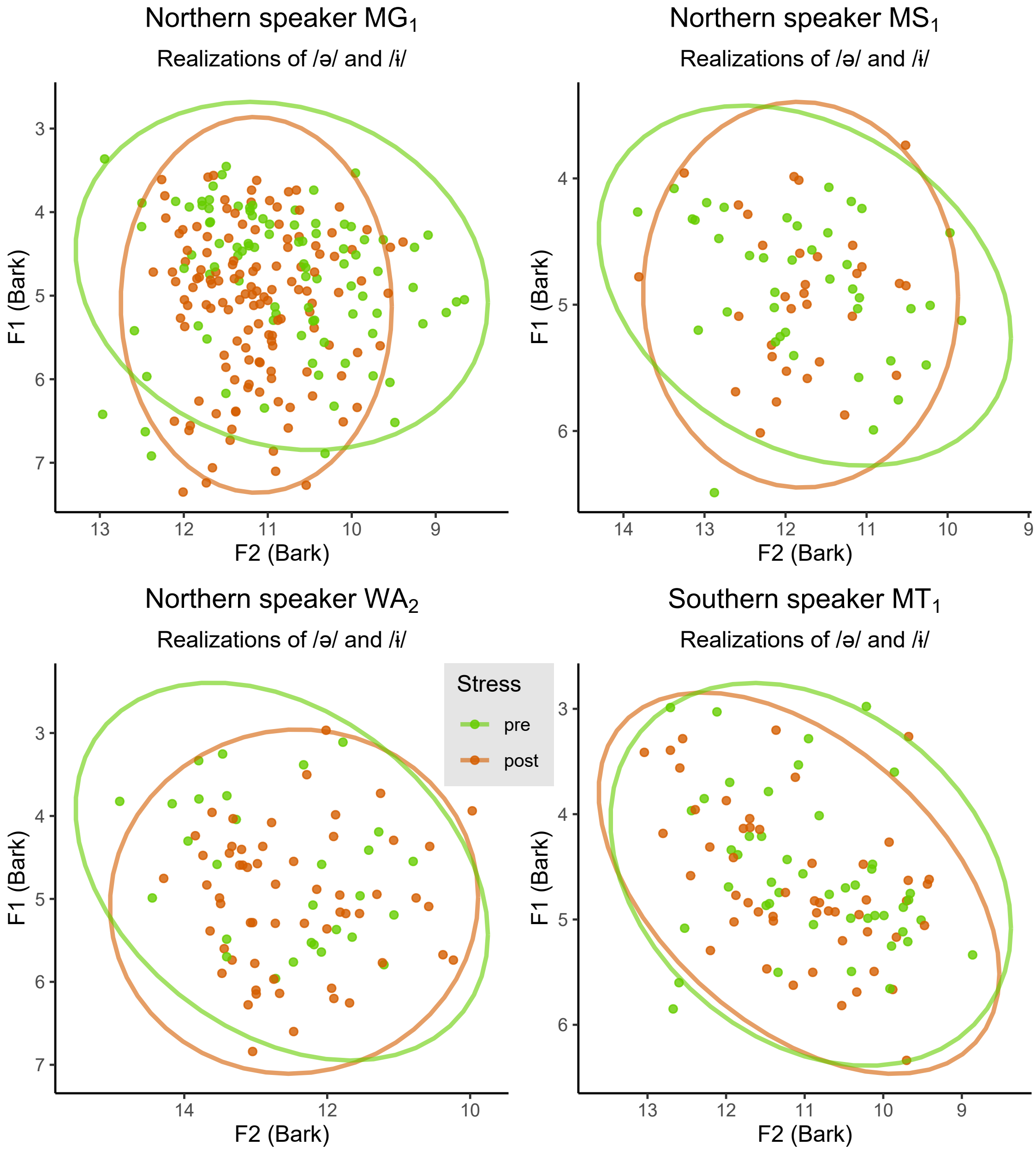

Figure 15: Three Northern and one Southern speaker: Scatter plots of the distributions of pre- and posttonic $/\text{ə}/_{\text{single}}$ (a.k.a. /ə/ or /ɨ/) realizations, grouped by whether word-stress position was pretonic or posttonic. Confidence ellipses with a probability level of .95 have been added to the plots.

vowel phoneme $/\text{ə}/_{\text{single}}$ displays moderate variation for a syntactically motivated factor like lexical category and large variation for phonetic factors like coarticulation context and word stress. The variation linked to word stress displays a clear pattern across individual speakers, namely that stressed vowel realizations appear as more fronted and somewhat more open than unstressed ones. This could reasonably motivate labelling $/\text{ə}/_{\text{single}}$ as /ɛ/ instead, interpreting realizations around [ɛ] as the default and realizations around [ə] as centralized allophonic variants caused by reduction in unstressed positions.

5 Discussion and Conclusions

At first, the results presented in Section 4.1 seemed to indicate less overlap between the vowel types /ə/ and /ɨ/ for speakers of the Northern rather than the Southern dialectal varieties of Aromanian, captured by higher Pillai scores on average. However, this rather paradoxical result could largely be attributed to a strong association between the variables vowel type and word-stress position, with /ə/ tokens appearing way more frequently in unstressed (a.k.a. atonic, pretonic or posttonic) positions than /ɨ/, whereas /ɨ/ showed up somewhat more frequently than /ə/ in stressed (a.k.a. tonic) positions. This association, in combination with the fact that stress is associated with more variation in the distributions of the vowels of interest in the Northern varieties, with stressed tokens being more likely to be realized as [ɛ], was causing those rather misleading initial measurements. A more reliable comparison could thus be acquired by focusing exclusively on stressed vowels, also considering that reduction towards [ə] is very common for unstressed vowels across many languages. The distributions of stressed vowels, indeed, appeared to confirm the trend assumed by theoretical classifications, namely that Southern speakers would show a clear distinction between the types /ə/ and /ɨ/ with corresponding acoustic qualities of approximately [ə] and [ɨ], whereas in the speech of Northern speakers, a contrast along those lines would be absent. After discovering strong associations between vowel type, on the one hand, and a rough encoding of syntactic role as well as adjacent phonetic segment types, on the other hand, it remains a bit doubtful whether the observed Southern-Aromanian contrast between /ə/ and /ɨ/ is context-independent enough to actually constitute a phonemic one (see Section 4.3). These doubts are underlined by the examination of the impact of a specific combination of syntactic and phonetic context on the resulting overlap measurements, which turned out to be substantial. An impression of the Aromanian vowel space and the mean locations of all seven vowel types within it, grouped by indexical speaker properties, pointed to expectable differences across sex and dialect (see Section 4.2). It could also confirm tendencies already suspected based on the overlap measurements discussed earlier. Most remarkably, in the speech of Northern speakers, stressed /ə/ tokens tended to be realized even more open and slightly more fronted (closer to [ɛ]) than their stressed /ɨ/ counterparts (closer to [e]), on average. On the whole, though, the common assumption of a merged vowel phoneme /ə/$_{\text{single}}$ for the Northern-Aromanian dialect group still seems to be justified, as suggested by the measures for degree of distribution overlap collected for the respective speakers. In Section 4.4, a more detailed statistical analysis could quantify the extents to which various factors were contributing to the observed variation within that phoneme. Even though the largest effect sizes were associated with context of coarticulation, the most systematic, speaker-independent, but dialect-specific relationship could be discerned for the influence of word stress: The much more concentrated and dense distribution of stressed tokens in more fronted acoustic positions, could justify interpreting [ɛ] as the default realization of the Northern-Aromanian /ə/$_{\text{single}}$ phoneme, rendering an alternative denotation as /ɛ/ quite plausible.

With regard to the main research question, it can be stated that the vowel types /ə/ and /ɨ/ are, indeed, acoustically merged in the speech of the majority of examined native speakers from Andon Poçi, Stjar and Gjirokastër, representing the Northern-Aromanian dialectal varieties. Acoustic mergedness obviously also entails phonemic irrelevance for the hypothesized vowel contrast. The judgement has to be a little more nuanced when it comes to native speakers of Southern-Aromanian varieties. As most of the featured individuals from Turia—none from Kutsufliani since their speech portions were too sparse for the tonic-only condition—did tend to acoustically distinguish between the two vowel types when only considering tonic word-stress positions, the underlying theoretical assumption about a phonemic relevance of that vowel contrast cannot be dismissed that easily, for their dialect region. Nonetheless, taking into account the significant association between vowel type and syntactic as well as phonetic context, it seems unlikely that the moderate acoustic distinctness between the two vowel types actually translates into a phonemic one as well. In addition to that, the fact that the theoretically ideal phonological mapping to either /ə/ or /ɨ/ is often ambiguous and contradictory among existing written sources, notably, even with respect to one of the two minimal pairs proposed by Beis and Dasoulas (2017), also supports a conclusion along the lines laid out by Schlösser (1985), namely that even if acoustic differences between the two vowel types do exist, under certain conditions, those types should rather be viewed as allophonic variants of the same underlying phoneme, i.e. /ə/$_{\text{single}}$.

In any case, it should be kept in mind that in the present corpus study, two geographically quite small clusters of speaker origin places have been picked for representing the two broader Aromanian dialect regions: The aerial distance between Andon Poçi and Stjar is barely 27 km (Gjirokastër lies roughly in-between the two). Between Turia and Kutsufliani, it is just 6 km. Of course, the entire geographical regions associated with either dialect region each extend over much further distances. In consideration of the high degree of local variation that can be observed all across the Aromanian dialect spectrum, as pointed out by Kahl (2005) and indicated by the findings of Kharlamova (2019), it would thus be necessary to investigate whether the quantitative tendencies discovered by this study can be reproduced when applying the same methodology to the speech of speakers from other places of origin, located in less proximity to the ones examined here.

References

Austin, P. K. (2009), *Encyclopedia of the world's endangered languages*, JSTOR.

Bara, M., Kahl, T. and Sobolev, A. (2005), *Die südaromunische Mundart von Turia (Pindos): Syntax, Lexik, Ethnolinguistik, Texte*, Biblion Verlag.

Beis, S. and Dasoulas, F. (2017), 'Prótasi gia to sýstima grafís tis vlachikís glóssas', *Glossologia* (25), 51–69.

Bobicev, V., Mărănduc, C. and Perez, C. A. (2017), 'Tools for building a corpus to study the historical and geographical variation of the Romanian language', *Proceedings of the LT4DHCSEE in conjunction with RANLP* pp. 10–19.

Boersma, P. (2001), 'Praat, a system for doing phonetics by computer', *Glot. Int.* **5**(9), 341–345.

Capidan, T. (1932), *Aromânii. Dialectul aromân. Studiu lingvistic.*

Caragiu-Marioțeanu, M. (1975), *Compendiu de dialectologie română (nord- şi sud-dunăreană)*, Editura ştiinţifică şi enciclopedică.

Dietrich, W., Kahl, T. and Sarros, G. (2001), *Volkskundliche Studie eines aromunischen Dorfes im Pindos-Gebirge (Panagía, Distrikt Tríkala)*, Verlag Gebrüder Kyriakidis.

Grosvald, M. (2010), 'Long-distance coarticulation in spoken and signed language: An overview', *Language and Linguistics Compass* **4**(6), 348–362.

Harrington, J. (2010), 'Acoustic phonetics', *The handbook of phonetic sciences* **2**, 81–129.

Harrington, J. (2018), 'Analysis of formants and formant transitions'. URL: *https://www.bas.uni-muenchen.de/~jmh/lehre/Rdf/EMU-SDMS/lesson8/08_Analysis-of-Formant-data2.html* [Accessed December 29, 2020].

Heeringa, W., Johnson, K. and Gooskens, C. (2009), 'Measuring Norwegian dialect distances using acoustic features', *Speech Communication* **51**(2), 167–183.

Kahl, T. (2005), 'Offene Fragen in der Erforschung des Aromunischen und seiner Dialekte', *Yazyki i dialekty malykh etnicheskikh grupp na Balkanakh* pp. 155–166.

Kahl, T. and Pascaru, A. (2018), 'Rrāmānj und Armānj. Wie einheitlich sind Kultur, Sprache und Identität der Aromunen?', *Quo vadis Romania* (50), 51–77.

Kahl, T., Poci, H., Bara, M., Bara, N., Poci, S., Pascaru, A. and Aminian, M. (2019), 'Dóji frats – The two brothers', retrieved from www.oeaw.ac.at/VLACH, ID number: rram1234ALV0001a. Video URL: *https://www.youtube.com/watch?v=sW3O2W3qagM* [Accessed December 29, 2020].

Kahl, T., Poci, S. and Fichera, A. (2020), 'Noia şi noi – Noah and the Aromanians', retrieved from www.oeaw.ac.at/VLACH, ID number: rram1234ALV0003a. Video URL: *https:// www.youtube.com/watch?v=WeyfrcUqsag* [Accessed December 29, 2020].

Katsanis, N. and Ntinas, K. (1990), *Grammatikí tis Koinís Koutsovlachikís*, Archeío Koutsovlachikón Meletón.

Kharlamova, A. (2017), 'Affricates and affricated consonants in Aromanian spontaneous speech', *ExLing 2017* p. 57.

Kharlamova, A. (2019), 'Foneticheskiye osobennosti arumynskogo govora g. Selenitsa (Albaniya): rezul'taty predvaritel'nogo issledovaniya', *Indoyevropeyskoye yazykoznaniye i klassicheskaya filologiya* (23-2).

Kisler, T., Reichel, U. and Schiel, F. (2017), 'Multilingual processing of speech via web services', *Computer Speech & Language* **45**, 326–347.

Mayer, J. (2017), *Phonetische Analysen mit Praat. Ein Handbuch für Ein-und Umsteiger*. URL: *http://praatpfanne.lingphon.net/downloads/praat_ manual.pdf* [Accessed December 29, 2020].

Mazzoni, D. and Dannenberg, R. (2000), *Audacity [software]*. The Audacity Team, Pittsburg, PA, USA.

Namurov, A. (2001), 'The Aromanian language/dialect', *Yazyki mira. Romanskiye yazyki [Languages of the World. Romance Languages]* pp. 636–656.

Nitsiakos, V., Liakos, S. and Seviloglou, M. (2019), *O Lele [CD booklet]*. Ekdóseis Taxideftís.

Nycz, J. and Hall-Lew, L. (2013), 'Best practices in measuring vowel merger', *Proceedings of Meetings on Acoustics 166ASA* **20**(1), 060008.

Papadiotis, G. (1988), *Vláchika tragoúdia tou Metsóvou*, Topikoí politismoí ar. 1, Ekdóseis Gérou.

Papahagi, T. (1922), *Antologie aromânească*, Din Publicaţiile casei scoalelor, România nouă.

R Core Team (2020), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. URL: *https://www.R-project.org* [Accessed December 29, 2020].

Rădulescu, S. and Kahl, T. (2006), *The Aromanians of Andon Poçi. Songs and stories [CD booklet]*, Fundația Alexandru Tzigara Samurcaș.

Schlösser, R. (1985), *Historische Lautlehre des Aromunischen von Metsovon*, Buske.

Simionescu, R. (2011), 'Hybrid POS tagger', *Proceedings of language resources and tools with industrial applications workshop (Eurolan 2011 Summer School)* pp. 21–28. Cluj-Napoca, Romania.

Stanley, J. (2017), 'A tutorial on extracting formants in Praat'. URL: *https://joeystanley.com/blog/a-tutorial-on-extracting-formants-in-praat* [Accessed December 29, 2020].

Stanley, J. (2019), 'A tutorial in measuring vowel overlap in R'. URL: *https://joeystanley.com/blog/a-tutorial-in-calculating-vowel-overlap* [Accessed December 29, 2020].

Turculeț, A. (2010), 'Intonația interogativă la aromânii din R. Macedonia (cu câteva particularități segmentale ale informatorilor)'. *Fonetică și dialectologie.*

Weigand, G. (1888), *Die Sprache der Olympo-Walachen. I. Teil. Lautlehre*, Druck von Metzger und Wittig.

Weinfurt, K. P. (1995), *Multivariate analysis of variance*, American Psychological Association.

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York. URL: *https://ggplot2.tidyverse.org* [Accessed December 29, 2020].