Marginal Contrast Among Romanian Vowels: Evidence from ASR and Functional Load

Margaret E L Renwick, Ioana Vasilescu, Camille Dutrey, Lori Lamel, Bianca
Vieru

Marginal contrast among Romanian vowels: evidence from ASR and functional load

Margaret E.L. Renwick[1], Ioana Vasilescu[2], Camille Dutrey[3], Lori Lamel[4], Bianca Vieru[5]

[1]University of Georgia, Athens GA 30602, USA
[2],[4]LIMSI, CNRS, Université Paris-Saclay, Bât. 508, Orsay, France
[3]Laboratoire National de Métrologie et d'Essais, 29 Avenue Roger Hennequin, 78190 Trappes, France
[5] Vocapia Research, 28 Rue Jean Rostand, 91400 Orsay, France
`mrenwick@uga.edu`, `{ioana,lamel}@limsi.fr`, `camille.dutrey@lne.fr`, `vieru@vocapia.com`

Abstract

This work quantifies the phonological contrast between the Romanian central vowels [ʌ] and [ɨ], which are considered separate phonemes, although they are historical allophones with few minimal pairs. We consider the vowels' functional load within the Romanian inventory and the usefulness of the contrast for automatic speech recognition (ASR). Using a 7 hour corpus of automatically aligned broadcast speech, the relative frequencies of vowels are compared across phonological contexts. Results indicate a near complementary distribution of [ʌ] and [ɨ]: the contrast scores lowest of all pairwise comparisons on measures of functional load, and shows the highest Kullback-Leibler divergence, suggesting that few lexical distinctions depend on the contrast. Thereafter, forced alignment is performed using an existing ASR system. The system selects among [ɨ], [ʌ], ∅ for lexical /i/, testing for its reduction in continuous speech. The same data is transcribed using the ASR system where [ʌ]/[ɨ] are merged, testing the hypothesis that loss of a marginal contrast has little impact on ASR error rates. Both results are consistent with functional load calculations, indicating that the /ʌ/ - /i/ contrast is lexically and phonetically weak. These results show how automatic transcription tools can help test phonological predictions using continuous speech.

Index Terms: phonology, Romanian vowels, marginal contrast, frequency distribution, functional load, automatic speech recognition, pronunciation variants.

1. Introduction

A central question in phonological analysis is whether two sounds lie in phonemic contrast with one another, a state traditionally diagnosed by the presence of minimal pairs, in which a single sound change signals a lexical difference. Such a distinction implies that the two sounds must be represented as separate entities in the minds of native speakers, and the set of all contrastive sounds constitute a language's phonemic inventory. Sounds that do not contrast, for example because they lie in complementary distribution, are predictable allophones. Increasingly, however, close analysis of the relationships among sounds has revealed gradient, even *marginal*, phonological contrasts. Evidence comes from speech production [1, 2, 3], speech perception [4, 5] and via sounds' distribution in the language, particularly functional load [6, 7]. Functional load is estimated with measures that consider how languages use particular contrasts, i.e. how many lexical distinctions depend on a single pair of phonemes [8, 9]. It has been associated with language evolution, that is, the probability that two sounds will merge over time [10], language acquisition [11], or cross-language comparisons [12].

In this paper we analyze continuous speech processed by automatic speech transcription tools. Firstly, we investigate the relationship between the frequency distributions of sounds in the corpus and the functional load of contrasts they participate in, acknowledging the effects of contextual conditioning on their distributions. Secondly, we apply forced alignment and ASR experimental methods to gain insight about how crucial a particular contrast is in a given language. We focus on the vocalic inventory of Romanian, a language whose central vowels [ʌ] and [ɨ] are argued to be marginally contrastive on the basis of their distribution in the language; they are historical allophones, and despite the synchronic presence of a few minimal pairs, they are in nearly complementary distribution [13].

2. Romanian: a brief description

Romanian is a Romance language, mother tongue of about 29 million speakers in Romania and the Republic of Moldova. Although more than 60% of the fundamental vocabulary is Latin, Slavic influence is prevalent in the Romanian lexicon. Today's Romanian is nourished by borrowings from modern Romance languages and English [14]. The Romanian vocalic system includes seven monophthongs ([a], [e], [i], [o], [u], [ʌ] and [ɨ]) that may appear in both stressed and unstressed position, in open or closed syllables, in lexical roots of any length, and in affixes. Vowels may appear adjacent to any consonant, though as we show, certain sequences are more or less likely. Within phonological forms, the root is typically followed by one or more morphological suffixes that place restrictions on the vowels that actually appear in word-final position [15]. Romanian inherited the five vowel qualities of Latin after the loss of length contrast, and developed two additional central vowels [ʌ] and [ɨ]. [ʌ] arose from Latin [a] in unstressed syllables, and was reinforced by Slavic and other East European idiomatic borrowings. [ɨ] and [ʌ] coexisted in free variation and then in complementary distribution, before becoming separate phonemes in the 16th century, as minimal pairs arose [16, 17]. Historically, [ɨ] operated as a stressed, typically pre-nasal allophone of [ʌ] [18]. [ʌ] and [ɨ] are distinct for native speakers, demonstrated by production and perception analyses of laboratory speech [13]. However, recently it has been emphasized that their contrast is severely diminished in continuous speech, leading to a phonetic near-merger [19].

3. Data

The corpus analyzed here consists of broadcast data gathered from various Romanian radio and television shows (7 hours, 141 speakers), from read speech and more spontaneous interactions such as debates. The number of speakers varies according to the source, ranging from 3 to 24 and including both male and female adult speakers. Recordings portray the standard version of the language based on the Southern dialect. Efforts were made to avoid sources with significant quantities of overlapping speech, foreign or regional accents and noisy background. The data were collect and annotated in the context of the Quaero program[1].

Table 1 sums up the characteristics of the corpus. The data were automatically aligned and segmented into words and phones using the system described in [20], for which they served as development and evaluation corpora.

Table 1: Summary of data sources and characteristics

Total sources	4
Duration	7h10'
Mean duration/file	4'56"
Total words	56,296
Total distinct words	9,032
Speaker count	141

4. Experimental approach

Three experimental approaches are used to gain insight into the phonological patterns of the /ʌ/-/ɨ/ contrast.

Frequency distribution of vowels. Comparing the relative frequencies of Romanian vowels can highlight restrictions on the vowels' distributions, as well as their phonological conditioning, in particular of [ʌ] and [ɨ]. Here, vowel type frequency was calculated across the 9,032 distinct words in the corpus, representing the 56k total lexical tokens.

Functional load and contextual frequency. The distributions of /ɨ, ʌ/ are quantitatively compared using three algorithms available via the Phonological Corpus Tools (PCT) toolkit [21]. The input to the toolkit is the list of unique words in the corpus, converted to IPA characters, and calculations are run as pairwise comparisons of Romanian vowels, using both type and token frequency measures where available. The weak relationship between [ɨ] and [ʌ] is hypothesized to show in several measures:

1. *Functional load* is correlated with the presence of minimal pairs; thus it is expected to be low for [ɨ], [ʌ] with respect to other pairs of sounds. It is calculated on the basis of both type and token frequency in the corpus, using a measure of the change in entropy that occurs when a contrast is neutralized (i.e. merged).

2. *Probability distributions* are based on the Kullback-Leibler divergence score, a calculation of relative entropy. It quantifies the dissimilarity of distributions between two sounds, such that a low score is produced when two sounds appear in the same environments, and a high score is produced when they appear in diverse environments (i.e. when their probability distributions are quite different). Because both [ɨ] and [ʌ] appear preferentially in different contexts, their distributions are expected to diverge. Kullback-Leibler scores are calculated according to (a) the left-hand segmental environment; (b) the right-hand environment; (c) both sides.

3. *Mutual information* is calculated between a sound and those to its left and right. It is useful in understanding the status of /ɨ/, whose distribution is more restricted than other vowels in the system. We predict an increased relationship of mutual information between /ɨ/ and the contexts where it was conditioned, such as before [m,n,r]; but such relationships are not expected for /ʌ/. Mutual information is calculated using token frequency in the following six ordered-pair bigrams: {[ɨ, ʌ]}[n], {[ɨ, ʌ]}[m], {[ɨ, ʌ]}[l], {[ɨ, ʌ]}[r], {[ɨ, ʌ]}#, #{[ɨ, ʌ]}.

Forced alignment and automatic speech transcription experiments. Two experiments are run to test the functionality for ASR of the /ɨ/-/ʌ/ contrast:

1. *Forced alignment of pronunciation variants for /ɨ/*, performed with an ASR system for Romanian [20]. The speech recognizer selects among [ɨ], [ʌ] or ∅ (suppression of the vowel) for lexical /ɨ/ on development data (3.5h of the broadcast corpus). The variants are selected to estimate the relation between the effective realization of [ɨ] and [ʌ] and their historical relation through maintenance [13], suppression [18], or centralization of [ɨ] [19]. We use both large context dependent (CD) and context independent (CI) models, as a method to investigate the dependency of the alignment on the system's configuration [22].

2. *Automatic speech transcription experiments*. The initial phonemic setting used for the ASR system includes 29 symbols: 20 consonants, 2 glides, 7 vowels and a special symbol for silence [20]. We evaluate the effect of the /ɨ/-/ʌ/ contrast on ASR performance, by comparing performance with and without the distinction. A weak contrast should result in a low word error rate increase[2]. Additionally, we consider two other vowel pairs proven to be frequent and to participate in morphological oppositions: [a]/[e] and [e]/[ʌ] [13]. The rationale behind this choice is that loss of the latter contrasts would increase the WER. Acoustic models are trained for each phone set using the respective pronunciation dictionaries in which the three vowels pairs are consecutively merged: [ʌ]/[ɨ], [a]/[e] and [e]/[ʌ]. The same mappings are applied to pronunciation dictionaries used to automatically transcribe the data set. The training of the language models is as in [20]. All acoustic models are built in a semi-supervised manner as reported in [24, 25] via flat start using about 370 hours of untranscribed audio. Acoustic models are triphone-based left-to-right 3-state Hidden Markov Models (HMM) with Gaussian mixture observation densities. The models are word position dependent. A neural network of 4 hidden layers with 10M parameters is trained with the cross-entropy criterion. The input consists of Perceptual Linear Predictive features [26] with 9 spliced frame vectors. A Linear Discriminant Analysis transform is applied to reduce the input feature vector to 360. Constrained Maximum Likelihood Linear Regression [27] speaker-dependent transforms are applied during training and testing. The target outputs consist of 10k HMM tied-states.

---

[1] http://www.quaero.org

[2] The commonly used metric for ASR systems evaluation is the Word Error Rate (WER), which takes into account three error types with respect to a reference transcription: substitutions of one word by another, insertions and deletions [23].
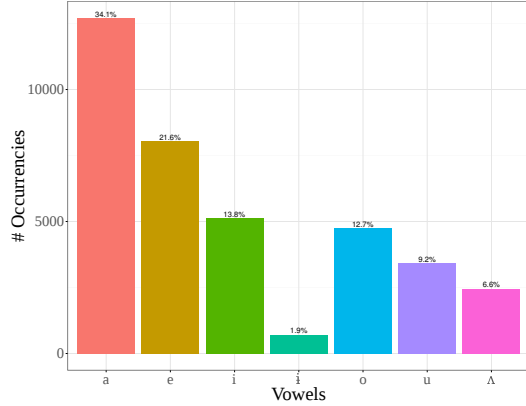
Figure 1: Frequency distribution of the vowels. Left to right: /a, e, i, ɨ, o, u, ʌ/

5. Results

5.1. Frequency distribution

A frequency analysis illustrates which vowels are most common across the corpus, and which sounds tend to co-occur, setting the stage for calculations of functional load and mutual information. If two segments co-occur more often than predicted by their individual frequencies, phonological conditioning may be the cause. As shown in Figure 1, across the 9,032 distinct words in the corpus, [a], [e] and [i] are the most frequent timbres, whereas taken together [ɨ] and [ʌ] account for less than 10% of the vowels. These results are globally consistent with [13], in particular with regard to the very low representation of the central vowels [ɨ] and [ʌ].

Table 2: Distribution of the central vowels /ɨ/ (î, â) and /ʌ/ (ă) according to contexts where they appear. Examples of the most frequent word in each context are given.

Ctx.	/ɨ/		/ʌ/	
CVC	30.5%	România	26.6%	astăzi
VVC	0.3%	neîncadrabili	2.8%	două mii doisprezece
CVV	1.3%	mâine	1.8%	său
#VC	67.2%	în	0.4%	ăsta
#VV	0.7%	îi	0.1%	ăia
CV#	0.0%	–	65.8%	să
VV#	0.0%	–	2.5%	două

The vowels [ɨ] and [ʌ] are the rarest elements of the system, despite their presence in frequent function words, and morphological role in the nominal and verbal declensions. Across the corpus their distribution is also highly complementary, as shown in Table 2: while [ɨ] is common in word-initial position, [ʌ] rarely appears there; conversely, [ʌ] is common word-finally, in its role as a marker of nominal declension (e.g. [ˈfatʌ] 'girl' vs. [ˈfata] 'the girl') or verb form ([sʌ ˈfakʌ], 'may he do it'); but [ɨ] never appears finally. Within the CVC context where both appear, [ɨ] overwhelmingly occurs before nasals or [r] (74.48% of CVC [ɨ] tokens), due to its emergence as a product of pre-nasal raising [18], while [ʌ] can precede a wider range of segments ([ʌ] precedes [m,n,r] in 43.46% of CVC cases).

5.2. Functional load and contextual frequency

Although the central vowels are declared separate phonemes on the basis of minimal pairs like [rɨu] 'river' vs. [rʌu] 'bad', the fact that minimal pairs are very rare suggests that the functional load of their contrast is low — a situation that could account for the vowels' near-merger in continuous speech [19]. Results of these calculations are shown in the tables below; although either type or token frequency-based calculations are shown, results were similar using both measures.

As shown in Table 3, the *functional load* of the /ʌ, ɨ/ contrast is the lowest of all pairwise comparisons, indicating that a minimal number of lexical contrasts would be lost if the two sounds merged. As for the *Kullback-Leibler divergence*, Tables 4 and 5 show that Kullback-Leibler divergence between /ʌ, ɨ/ is the highest of all comparisons, based both on the left-hand and right-hand contexts for all vowel segments. This indicates that, more than any other vowel pair, these two appear in complementary environments. Finally the results for *mutual information* isolate certain contexts where either /ʌ/ or /ɨ/ is expected, for historical reasons (see Table 6). In those environments, there is considerable mutual information between the conditioned vowel and the context, and much less mutual information between the same context and the unconditioned vowel; for instance, the mutual information of /ɨ/ with a preceding word boundary is high, while /ʌ/ shares negative mutual information with the same context (and vice versa in word-final position).

Table 3: Functional load, calculated with respect to type frequency, change in entropy algorithm

	e	i	o	u	ɨ	ʌ
a	0.0428	0.0143	0.0071	0.0135	0.0025	0.0456
e		0.0257	0.0032	0.0107	0.0011	0.0545
i			0.0018	0.0061	0.0007	0.0078
o				0.0025	0.0011	0.0018
u					0.0032	0.0036
ɨ						**0.0004**

Table 4: Left-hand context comparison, Kullback-Leibler divergence; calculated with token frequency

	e	i	o	u	ɨ	ʌ
a	0.901	1.087	1.151	0.815	4.359	1.132
e		0.648	0.933	1.295	5.241	1.123
i			1.648	0.802	4.006	1.770
o				1.682	3.814	2.721
u					3.354	2.198
ɨ						**7.567**

Table 5: Right-hand context comparison, Kullback-Leibler divergence; calculated with token frequency

	e	i	o	u	ɨ	ʌ
a	1.969	2.215	0.728	1.209	1.761	2.688
e		1.207	1.253	1.519	4.061	2.442
i			2.199	2.262	4.317	2.818
o				0.567	2.366	1.936
u					2.387	1.194
ɨ						**5.430**

5.3. Assessing functionality through ASR experiments

5.3.1. Pronunciation variants for [ɨ]

Introducing pronunciation variants in a recognizer system's lexicon may improve the system's performance [22]. On the linguistic side, allowing an automatic selection of variants chosen to reflect contextual variation may validate and/or quantify

Table 6: Mutual information, calculated within ordered-pair bi-grams using token frequency

Context	ɨ	ʌ	/ɨ/ – /ʌ/ Difference
___n	4.335	–1.139	5.474
___m	1.381	–1.007	0.374
___r	0.524	–1.43	1.954
___l	–1.141	–2.395	1.254
#___	2.481	–5.683	8.164
___#	–6.419	2.503	–8.922

the frequency of a phonetic event in connected speech [28, 20]. Here they are used to highlight the relationship between the canonical form as encoded in the pronunciation dictionary and the actual realizations of [ɨ]. The observed variation may match historical "footprints" (e.g. [ɨ] aligned as [ʌ] for an historical [ʌ]) and suggest ongoing mutations. Although with CI models more variants are selected, results show similar trends with both CI and CD models (Table 7): [ɨ] is maintained or suppressed in word initial position before a nasal consonant, that is #**VC** context. Traditionally [ɨ] in word initial position is associated with pre-nasal centralization, whereas word internal [ɨ] corresponds to an emergence of [ɨ] from [ʌ], originally as a contextually predictable allophone. In this ASR test, word internal position elicits maintenance, [ʌ] selection, or suppression. Centralization effects (i.e. transcription of /ɨ/ as [ʌ]) are evident for 10% of the occurrences with CI models and only 2% with CD models, and the difference can mainly be explained by the fact that CI models are much "wider" models and the system can more easily choose another phoneme. The CD models capture the local contexts of [ɨ] and [ʌ], which is the likely explanation for why only 2% of lowering is observed. Taken together the results are consistent with analyses of [ɨ]'s historical emergence and restricted distribution [18].

Table 7: Pronunciation variants for /ɨ/ with CI (above) and CD (below) models. Word position: word initial (#**VV**, #**VC** and word internal (C**VC**, V**VC**, C**VV**), where V=/ɨ/. Variants: maintenance ([ɨ]), suppression (∅), substitution ([ʌ])

CI models			
Word pos./var.	/ɨ/	/ʌ/	∅
#**VV**	4	2	13
#**VC**	709	138	564
C**VC**	410	92	62
V**VC**	3	0	1
C**VV**	10	0	7
Total	1136 (56%)	232 (10%)	647 (32%)
CD models			
#**VV**	8	0	11
#**VC**	1132	30	221
C**VC**	516	11	20
V**VC**	3	0	1
C**VV**	15	1	0
Total	1674 (85%)	42 (2%)	253 (12%)

5.3.2. Relating functional load to ASR results

In this experiment, four sets of acoustic models were tested: one including a classic vocalic inventory with 7 vowels and three in which each of the paired vowels (successively, [ʌ]/[ɨ], [a]/[e] and [e]/[ʌ]) are merged to a single symbol. In the data set, the pair [ʌ]/[ɨ] is the less represented: only 7% of words

are affected by the [ʌ]/[ɨ] fusion. By contrast, 70% of words are affected by [e]/[a] and 59% by [e]/[ʌ] fusion. The results in Table 8 highlight the correlation between relative frequency and functional load, and the the effect on ASR functionality. Although the baseline system (with a 7-vowel inventory) and the system based on [ʌ]/[ɨ] fusion obtain similar WER (around 16%), mergers of both [e]/[a] and [e]/[ʌ] increase WER (17.6% and 16.6% respectively). More specifically, 1.2% of the total number of words containing [e] or [a] are better recognized by the baseline system vs. the 6-vowel system with [e]/[a] merged. The same tendency is seen for [e]/[ʌ]: 0.5% of words containing [e] or [ʌ] were better recognized by the baseline system compared to the system where the two vowels were merged. As for merging [ʌ]/[ɨ], 0.02% (14 words) are better recognized by the system when the opposition is reduced to a single symbol. These findings confirm the low relevance for the vocalic system of the /ʌ/-/ɨ/ contrast, as opposed to frequent and morphologically salient contrasts involving /e/-/a/ and /e/-/ʌ/.

Table 8: WER for baseline ASR system (7 vowels) compared to 6-vowel systems ([ʌ]/[ɨ] vs. [e]/[a] vs. [e]/[ʌ] fusion), and calculations of functional load and Kullback-Leibler divergence.

rates	FL	K-L right	K-L left	WER (%)
baseline	–	–	–	**15.4**
ʌ/ɨ	0.0004	5.43	7.56	**15.5**
e/a	0.0428	1.96	0.90	**16.8**
e/ʌ	0.0545	2.24	1.12	**16.0**

6. Discussion

We addressed the question of the marginal contrast between central vowels [ʌ] and [ɨ] in Romanian. We applied a methodology based on frequency and functional load calculations, coupled to forced alignment and automatic speech transcription experiments, conducted on a 7 hour corpus of continuous speech. It has the advantage of not being dependent on the particular acoustic characteristics of the speech analyzed. Instead it permits estimation of the relevance of phonemic oppositions, within the lexicon and in their use in the language, because automatic speech recognition takes into account both distributional facts and acoustic distinctions. The results shows that /ʌ/-/ɨ/ is the weakest contrast among the Romanian vowels. They are the least frequent, with the lowest functional load and with the highest score measured by the Kullback-Leibler divergence algorithm. The forced alignment of contextual variants for [ɨ] is consistent with historical and contrastive analysis of the sound's emergence and distribution. ASR experiments show that a 6-vowel system where the /ʌ/-/ɨ/ contrast is removed by merging the two vowels performs as well as a classical 7-vowel system. More generally, this methodology shows how automatic transcription tools can help test phonological predictions using continuous speech. Further work concerns a more accurate investigation of the cognitive representations of the two vowels, through interative exploration and integration of data from phonological distributions, phonetic realization, and perception.

7. Acknowledgements

8. References

[1] J. Scobbie and J. Stuart-Smith, "Quasi-phonemic contrast and the fuzzy inventory: Examples from scottish english," in *Contrast in Phonology: Theory, Perception, Acquisition*, B. E. D. Peter Avery and K. Rice, Eds. Berlin: de Gruyter, 2008, pp. 87–114.

[2] J. I. Hualde, "Quasi-phonemic contrasts in spanish," in *WCFL:Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 2004, pp. 374–398.

[3] M. Nadeu and E. L. Renwick, "Variation in the lexical distribution and implementation of phonetically similar phonemes in catalan," *Journal of Phonetics*, no. 58, pp. 22–47, 2016.

[4] A. Boomershine, K. Currie Hall, E. Hume, and K. Johnson, "The impact of allophony versus contrast on speech perception," in *Contrasts in phonology: Theory, perception, acquisition*, P. Avery, B. E. Dresher, and K. Rice, Eds. Berlin: De Gruyter Mouton, 2008, pp. 145–171.

[5] J. C. Mora, J. L. Keidel, and J. Flege, "Effects of spanish use on the production of catalan vowels by early spanish-catalan bilinguals," in *The Phonetics-Phonology Interface: Sounds, Representations, Methodologies*, J. Romero and M. Riera, Eds. Amsterdam: John Benjamins, 2010, pp. 33–53.

[6] K. C. Hall, *A Probabilistic Model of Phonological Relationships from Contrast to Allophony*. Columbus, OH: The Ohio State University Ph.D. Dissertation, 2009.

[7] ——, "A typology of intermediate phonological relationships," *The Linguistic Review*, vol. 30, pp. 215–275, 2013.

[8] D. Surendran and P. Niyogi, "Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals," in *Competing Models of Linguistic Change: Evolution and Beyond. In commemoration of Eugenio Coşeriu (1921-2002)*, O. Nedergaard Thomsen, Ed. Amsterdam & Philadelphia: Benjamins, 2006, pp. 43–58.

[9] C. F. Hockett, "The quantification of functional load: A linguistic problem," *Memorandum. (U.S. Air Force)*, 1966.

[10] A. Wedel, A. Kaplan, and S. Jackson, "High functional load inhibits phonological contrast loss: A corpus study," *Cognition*, no. 2, pp. 179–186, 2013.

[11] S. F. Stokes and D. Surendran, "Ambient frequency, and functional load as predictors of consonant development in children," *Journal of Speech, Language, and Hearing Research*, vol. 48, pp. 577–591, 2004.

[12] Y. Oh, F. Pellegrino, C. Coupé, and E. Marsico, "Cross-language comparison of functional load for vowels, consonants, and tones," in *Proceedings of Interspeech 2013*, 2013, pp. 374–398.

[13] M. Renwick, *The Phonetics and Phonology of Contrast: The Case of the Romanian Vowel System*. Berlin, Boston: De Gruyter Mouton, 2014.

[14] D. Trandabat, E. Irimia, V. Barbu Mititelu, D. Cristea, and D. Tufis, "The romanian language in the digital age," in *META-NET White Paper Studies*. Springer, 2012.

[15] I. Chitoran, *The phonology of Romanian: a Constraint-Based Approach*, ser. Studies in Generative Grammar 56. Berlin; New York: de Gruyter Mouton, 2002.

[16] E. Vasiliu, *Fonologia istorică a dialectelor dacoromâne [Historical phonology of the Dacoromanian dialects]*. Bucureşti: Editura Academiei Republicii Socialiste România, 2014.

[17] A. Rosetti, *Istoria limbii române I: de la origini pînă la începutul secolului al XVII-lea [Histoire de la langue roumaine I: des origines jusqu'au début du 17ème siècle]*. Ediţie definitivă [Edition définitive]. Bucureşti, Editura ştiinţifică şi enciclopedică, 1986.

[18] A. Avram, "Vocalele [ă] si [î] [les voyelles [ă] si [î]]," *Studii de fonetică istorică a limbii române [Études de phonétique historique de la langue roumaine]*, pp. 25–111, 2012.

[19] M. Renwick, I. Vasilescu, C. Dutrey, L. Lamel, and B. Vieru, "A phonologically weak contrast can induce phonetic overlap," Poster presented at "LabPhon15: Speech Dynamics and Phonological Representation", 2016.

[20] I. Vasilescu, B. Vieru, and L. Lamel, "Exploring pronunciation variants for romanian speech-to-text transcription," in *Proceedings of SLTU-2014*, 2014, pp. 161–168.

[21] K. C. Hall, B. Allen, M. Fry, S. Mackie, and M. McAuliffe, "Phonological corpustools, version 1.1.0. [computer program]," 2015, https://github.com/PhonologicalCorpusTools/CorpusTools/releases.

[22] M. Adda-Decker and L. Lamel, "Pronunciation variants accros systems, languages and speaking style," in *Proceedings of MPV-1998*, 1998, pp. 1–6.

[23] L. Lamel and J. Gauvain, "Speech recognition," in *OUP Handbook on Computational Linguistics*, R. Mitkov, Ed. Oxford: Oxford University Press, 2003, pp. 305–322.

[24] J. Gauvain, L. Lamel, and G. Adda, "Lightly supervised and unsupervised acoustic model training," *Computer Speech and Language*, vol. 16, pp. 115–129, 2002.

[25] L. Lamel and B. Vieru, "Development of a speech-to-text transcription system for finnish," in *Proceedings of SLTU 2010*, 2010, pp. 62–67.

[26] H. Hermansky, "Perceptual linear predictive (plp) analysis of speech," *Journal of the Acoustical Society of America*, vol. 87, pp. 1738–1752, 1990.

[27] M. Gales, "Maximum likelihood linear transformations for hmm-based speech recognition," *University of Cambridge*, 1997.

[28] M. Adda-Decker, P. Boula de Mareuïl, and L. Lamel, "Pronunciation variants in french: schwa and liaison," in *Proceedings of ICPhS99*, 1999.