INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Exploring graph theory methods for the analysis of pronunciation variation in spontaneous speech

Bernhard C. Geiger[1,2], *Barbara Schuppler*[2]*

[1]Know-Center GmbH, Austria
[2]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

geiger@ieee.org, b.schuppler@tugraz.at

Abstract

Given the development of automatic speech recognition based techniques for creating phonetic annotations of large speech corpora, there has been a growing interest in investigating the frequencies of occurrence of phonological and reduction processes. Given that most studies have analyzed these processes separately, they did not provide insights about their co-occurrences. This paper contributes with introducing graph theory methods for the analysis of pronunciation variation in a large corpus of Austrian German conversational speech. More specifically, we investigate how reduction processes that are typical for spontaneous German in general co-occur with phonological processes typical for the Austrian German variety. Whereas our concrete findings are of special interest to scientists investigating variation in German, the approach presented opens new possibilities to analyze pronunciation variation in large corpora of different speaking styles in any language.

Index Terms: acoustic reduction, Austrian German, conversational speech, graph theory methods, pronunciation variation

1. Introduction

Given the development of automatic speech recognition (ASR) based techniques for creating phonetic annotations, there has been a growing interest in performing large-scale corpus studies investigating the frequency of certain phonetic (reduction) processes and phonological rules as well as the conditions under which they occur [1]. For instance, van den Heuvel and Cucchiarini [2] found that in Dutch, /R/ deletion tends to be more frequent after schwa than after full vowels, independent of vowel length and lexical stress. Schuppler et al. [3] showed that the absence vs. presence of Dutch word-final /t/ is conditioned by segmental context, bigram frequency and the morphological properties of the words. More recently, Wu et al. [4] presented an analysis of phonological and extra-linguistic factors affecting schwa deletion in word-initial syllables of polysyllabic words in continuous French. Finally, large scale phonetic corpus studies have not only been used to investigate one specific process, but also to analyze global tendencies of pronunciation variation in different speaking styles (e.g., [5, 6, 7]).

Mentioned studies report the frequencies of occurrence of certain phonological and reduction processes, they do, however, not report which of these rules and processes tend to ”co-occur”, where with ”co-occur” we refer here to ”occurring together on a word and/or utterance level”. The main aim of this paper is to introduce graph theory methods to analyze co-occurrences of processes at word and utterance level and to illustrate its potentials for the analysis of pronunciation variation. We do so for a not yet extensively well documented language variety, i.e., Austrian German (AG), and base our study on a large corpus of spontaneous, casual conversations.

The number of acoustic phonetic corpus studies investigating pronunciation variation in AG is limited. With respect to consonant realizations, the literature documents plosive devoicing, lenition and spirantization [8, 9, 10], devoicing of alveolar fricatives and reductions and deletions of /R/ [11]. With respect to AG realization of vowels, Moosmüller [12] found that in also in standard AG (as reported for Middle-Bavarian Dialects), several long and short vowel pairs are distinguished by durational rather than spectral characteristics. Other studies focused on regional dialects (e.g., [13, 14]). For the Viennese dialect, several studies reported monophthongation processes and their spreading to other Austrian cities [15, 16, 17, 18]. Another well documented (regional) process is /l/-velarization in word-initial position and in diminuitives ending on <–erl> [19, 20]. So far there has only been one study providing an overview of frequencies of occurrence for a large number of phonological- and reduction processes in read and spontaneous AG [7]. Based on automatically created broad phonetic transcriptions, the study found that only 37.8% of word tokens in spontaneous speech were produced canonically. There is no other study taking into account not only rules typical for AG, but also processes that are frequent in spontaneous German in general (e.g., schwa-deletion in unstressed syllables).

In some languages, there is a situation of diglossia, where speakers clearly switch between standard and dialect in different speaking styles (e.g., in Arabic [21] and in Swiss German [22]). Earlier work by Auer [23] stated that diglossia was also present in rural AG, whereas other studies suggested rather a continuum of speaking styles between standard (Austrian) German and dialectal AG [24, 25]. Chosen style depends, among many factors, on the situation, the interlocutor (e.g., type of relationship) and the background of the speaker (education, region, etc.). Speakers change in style not only given the situation, but also within a conversation, sometimes even within a sentence (e.g., changing to clearly pronounced citation forms to mark emphasis). Given these continuous spectrum of styles, we believe that the analysis of spontaneous AG requires not only the analysis of sounds expected to be different from German German (as identified in impressionistic phonetic studies), but also an analysis along with regionally-independent German reduction processes.

In this paper, general reduction processes typical for spontaneous German (here referred to as *SpG rules*) are considered along with regional phonological processes typical for AG (here referred to as *SpAG rules*). Their co-occurrences are analyzed by means of graph theory methods, facilitating to gain a complete picture of pronunciation variation in Austrian German.

*Authors are listed in alphabetical order and contributed equally to this work.

10.21437/Interspeech.2023-1398

2. Materials and Methods

2.1. GRASS corpus

This study is based on the conversational component of the Graz Corpus of Read and Spontaneous Speech (GRASS) [26, 27]), which contains casual conversations from a total of 38 speakers (19f, 19m) from eastern Austria. They all have a university-level education and live in one of the mayor cities. The 19 speaker pairs were family members, friends, couples or colleagues, who had known each other for years. They were recorded with head-mounted microphones, while sitting together in a recording studio for one hour, without interruption and without an experimenter being present. Moreover, there was no restriction in terms of chosen topic or speaking behavior, leading to a casual speaking style of high degree of pronunciation variation [7], frequently occurring disfluencies, overlapping speech [28] and laughter [29].

2.2. Forced alignment with pronunciation variants

GRASS was entirely manually annotated at the orthographic level (including labels for speaker noises, laughter, disfluencies, etc.) and segmented into utterances of no longer than 8s. For the current study, utterances containing laughter and/or only breathing and smacking noises (without lexical token) were excluded, leaving a total of 181.230 word tokens produced in 30.091 utterances by 38 speakers. These utterances were automatically segmented using a KALDI-based [30] forced alignment [31, 32], using a lexicon with multiple pronunciation variants per lexical entry. The variants were created by applying 37 phonological and reduction rules to the canonical transcriptions (i.e., as created with a G2P online tool [33] for standard German). These rules make use of the syllabic structure of the words (syllabic boundaries and stress) and are mainly based on those listed in [7]. They reflect AG pronunciation typical for eastern Austria (given the GRASS speakers' origin). Additionally, manual pronunciation variants were created for very high frequent words, especially auxiliary verbs and pronouns, for which a sequence of rules would not yield to AG typical dialectal pronunciation. The resulting lexicon contained on average $5.57 - 6.18$ variants per word. During the forced alignment, the ASR system chose which pronunciation variant best matched the acoustic signal and the broad phonetic annotation (phone sequence and boundaries) was stored to a Praat textgrid file [34]. The phone labels of a small subset of the material (2 min from 6 conversations, i.e., a total of 12.951 segments) were corrected manually and used for a qualitative validation of the segmentation.

2.3. SpG and SpAG rules analyzed

During the generation of the variants, each rule applied was logged for all variants created. The original set of 37 rules was optimized for the variant generation process. For our analysis, we selected a subset of them and merged them:

1. *canonAG*: the word's full form, as given by the canonical German pronunciation including phonological substitution rules for standard AG (devoicing of alveolar fricatives, word-final <*-ig*> as /ik/, wordinitial <*Ch*> as /k/).
2. */R/-reduction*: deletion or vocalization of /R/ in coda position (e.g., *Garten* "garden" /g'artən/ pronounced as [g'aːtən].
3. *plosive-deletion*: in consonant clusters and in word-final position (e.g., *wichtig* "important" /v'ɪçtik/ spoken as [v'ɪçti]).
4. *schwa-deletion*: in unstressed syllables (e.g., *Garten* "garden" /g'artən/ pronounced as [g'aːtn]).
5. */C/-deletion*: in syllable-final position (e.g., *ich* "I" /ɪç/ as [ɪː]).
6. *vowel-diphthongation*: of stressed vowels (e.g., *kannst* "can you" /k'ɑnst/ pronounced as [k'aʊnst]).
7. *vowel-monophthongation*: of stressed vowels (e.g., *weit* "far" /v'aɛt/ as [v'eːt]).
8. *vowel-substitution*: of stressed vowels (e.g., *fragen* "to ask" /fr'agən/ as [fr'oːgn]).
9. *manualAG*: manually created pronunciations of highly frequent words (e.g., *wir* "we" /w'ɪə/ pronounced as [m'a]).

We refer to categories 2 - 4 as *SpG* (rules that would occur in spontaneous German of any variety), and to 5 - 9 as *SpAG* (rules that would occur in spontaneous AG.

2.4. Graph analysis

We visualize the co-occurrence of rules in a weighted, directed graph $\mathcal{G} = (V, E, W)$, where the vertices in V correspond to rules, where an edge $(a, b) \in E \subseteq V^2$ indicates that rule b co-occurs with rule a, and where the matrix W collects the edge weights w_e for all $e \in E$. We analyze the corpus both on word- and utterance-level, yielding two graphs $\mathcal{G}^{(\mathrm{w})}$ and $\mathcal{G}^{(\mathrm{u})}$.

For $\mathcal{G}^{(\mathrm{w})}$, an edge (a, b) exists if there is at least one word to which rules a and b are applied. The edge weight $w^{(\mathrm{w})}_{(a,b)}$ is computed by the relative fraction of times that rule b is applied to words to which rule a has been applied, i.e.,

$$w^{(\mathrm{w})}_{(a,b)} = \frac{\sum_{t\in\mathcal{C}} \mathbb{I}(a \text{ and } b \text{ applied to } t)}{\sum_{t'\in\mathcal{C}} \mathbb{I}(a \text{ applied to } t')\mathbb{I}(b \text{ applicable to } t')} \quad (1)$$

where $\mathcal{C}$ denotes the set of words t in the corpus and where $\mathbb{I}(A)$ is 1 if A is true and 0 otherwise. To make this clear, consider the following example: Suppose that rule a is applied to 60 words in the corpus. Of these 60 words, rule b can be applied to 40 words, while it cannot be applied to the remaining 20 words (i.e., "cannot be applied" refers to that either a word does not contain the phoneme affected by the specific rule or that in a word the phoneme does not occur in a segmental and/or syllabic context where the rule applies). Rule b has been applied jointly with rule a 10 times. The edge weight of (a, b) is thus $w^{(\mathrm{w})}_{(a,b)} = 10/40 = 0.25$. Suppose further that rule b has been applied to only 10 words in the corpus. Then, since for these 10 words also rule a has been applied, the edge (b, a) from b to a has weight $w^{(\mathrm{w})}_{(b,a)} = 10/10 = 1$. For $\mathcal{G}^{(\mathrm{u})}$, an edge (a, b) exists if there is at least one utterance in which rules a and b are applied at least once (regardless of whether they are applied to the same or to different words in the utterance). The edge weight $w^{(\mathrm{u})}_{(a,b)}$ for (a, b) is the average relative fraction of times rule b is applied whenever rule a is applied, where the average is taken over all utterances.

For visualization purposes, we additionally plot the vertices of $\mathcal{G}^{(\mathrm{w})}$ with a radius proportional to the number of times a rule is applied, i.e.,

$$r^{(\mathrm{w})}_a = \frac{\sum_{t\in\mathcal{C}} \mathbb{I}(a \text{ applied to } t)}{\sum_{t'\in\mathcal{C}} \mathbb{I}(a \text{ applicable to } t')}. \quad (2)$$

For $\mathcal{G}^{(\mathrm{u})}$, we average (2) over all utterances. Specifically, if in utterance u_i with 15 words rule a can be applied to 10 words and has been applied to three words, then for this utterance the ratio is 3/10. The vertex size $r^{(\mathrm{u})}_a$ in the graph is then the average of these ratios over all utterances.

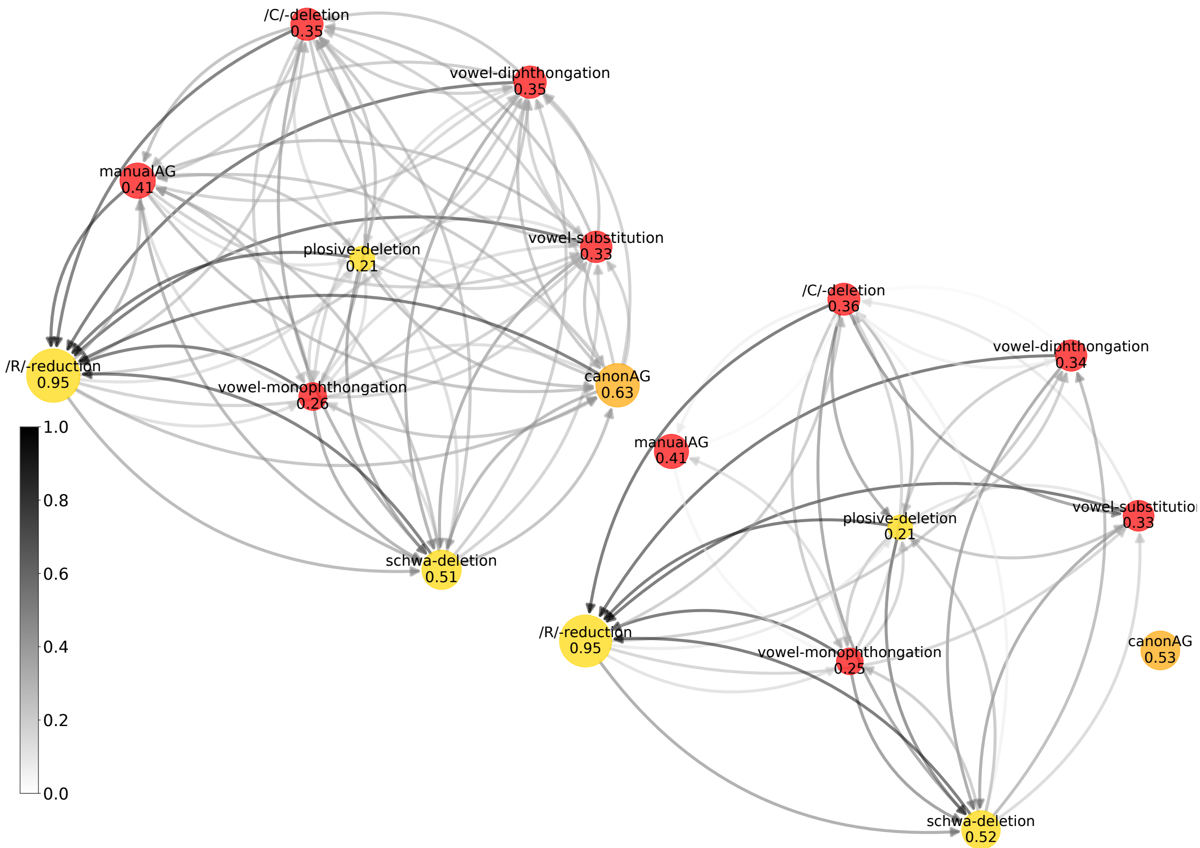

Figure 1: *Co-occurence graphs at word-level (right) and utterance-level (left) computed on the GRASS corpus. Node colors distinguish between canonical pronunciation (orange), general reduction for spontaneous German (yellow), and rules specific to AG (red). Node size and edge hue are proportional to the usage frequency and the co-occurrence of rules according to* (2) *and* (1)*, respectively.*

Graph analysis and visualization was implemented in Python using NetworkX 3.0, equations (1) and (2) were implemented manually.

3. Results and Discussion

Figure 1 shows the graphs $\mathcal{G}^{(\mathrm{w})}$ and $\mathcal{G}^{(\mathrm{u})}$ for the nine (merged) rules as calculated at the word- and utterance-level, where yellow circles indicate how often SpG rules were realized by the GRASS speakers, where red circles indicate SpAG rules, and where the orange circle indicates *canonAG*. Whereas $\mathcal{G}^{(\mathrm{w})}$ allows us to gain insights into the frequencies of occurrence and of co-occurrence of rules within the same word, which is largely determined by the word's length and segmental structure, graph $\mathcal{G}^{(\mathrm{u})}$ reveals additional insights given the larger, utterance-level context.

Full pronunciation form 53% of word tokens were produced in their full form (i.e., *canonAG*). This number is similar to quantitative analyses of casual, conversational speech of other Germanic languages (e.g., 56% canonical pronunciations in Dutch [35]), but higher than previously reported for the GRASS corpus (i.e., 37.8% in [7]). This can be explained by the fact that [7] considered *canonical* as the standard German pronunciation, whereas here we consider as *canonical* also standard AG (as documented for read speech of trained Austrian broad-cast speakers [36]). By definition, *canonAG* is isolated in the word-level graph. At the utterance level, the strongest edges point to */R/-reduction* (with a weight of 0.945) and *schwa-deletion* (0.501), while the weakest points to *plosive-deletion* (0.196). These values seem to correlate with $r_a^{(\mathrm{w})}$.

Despite **manualAG** being at the other end of the standard - dialect continuum than *canonAG*, it occurs similarly frequently in words where it applies (41%) and moreover, at the utterance level, all outgoing edge weights to all other rules are within a difference less than 0.03 to the edge weights of *canonAG*. These results indicate that *canonAG* and *manualAG*, in general, occur in similar utterance-contexts, which supports clearly the assumption of a continuity of speaking styles in AG rather than a diglossia (as suggested by Auer [23]).

SpAG vs. SpG In order to answer the question whether SpAG and SpG rules tend to mainly co-occur among themselves, we analyzed the assortativity of the graph at the word-level (−0.21) and utterance-level (−0.12), as well as the weighted modularity. Both indicate that SpAG and SpG do not form communities. This finding emphasizes the need to analyze (regionally occurring) phonological processes "together" with general reduction processes, which was the original motivation for this paper.

3.1. Spontaneous German reduction rules (SpG)

/R/-reduction The by far most frequent rule is */R/-reduction* with 95%, revealing that in most word tokens where /R/ appeared in coda position it was either reduced to [ɒ] or deleted.

That */R/-reduction* has a frequency close to 100% at both the word and utterance level also explains the large weights of incoming edges. In utterances with */R/-reduction*, *schwa-deletion* (0.516) and */C/-deletion* (0.36) are common, while *plosive-deletion* (0.215) is not; other rules co-occur with */R/-deletion* with a relative fraction of ranging around 0.3.

/R/-deletion was reported to occur in only 23% of tokens in AG read by trained speakers [11], and to be either deleted or vocalized in 49.1% when read by not-trained speakers [7]. Our results indicate that /R/-reduction increases with increasing spontaneity and casualness in AG.

Plosive deletion Plosives were deleted only in 21% of the word tokens in which the rule applied (i.e., in word-final position, in consonant clusters, in the prefixes). In those words where plosives were deleted, nearly always also /R/ was deleted (0.992) and schwas were deleted more frequently (0.793) than in general ($r^{(\mathrm{w})}_{\text{schwa-deletion}} = 0.52$). A further strong edge points to */C/-deletion* (0.467), and also all three SpAG vowel processes occur more frequently in a word in which a plosive was deleted than in other word tokens of the corpus. These results indicate that, at word level, plosive deletion occurs primarily in pronunciation variants that are reduced also by other reduction rules.

A comparison of these results with other studies on plosive deletion in German is difficult, as they either investigated plosives in different segmental contexts, or only plosives of one specific place of articulation. Based on spontaneous speech from the Kiel corpus, it was found that 10.5% of plosives (in any position and context) were deleted [37]. One comparable study to ours is the one presented in [7], which, based on a subset of 12 speakers of GRASS, reported 15.9% of plosive deletions in read AG and 27% in spontaneous AG, which is more than what we observe here on average for all 38 GRASS speakers.

Schwa-deletion Schwas were deleted in 52% of the word tokens. Its main co-occurrence is with */R/-reduction*: in 100% of the tokens, where schwas were deleted and */R/-reduction* could apply, /R/ was indeed deleted or vocalized. Despite plosives being reduced in general only in 21% of tokens, in those where schwas were deleted this number rises to 39.3%. Among the vowel processes studied here, which all apply to full, stressed vowels, *vowel-diphthongation* (0.477) occurs most likely in words with deleted schwas, followed by *vowel-monophthongation* (0.366), and *vowel-substitution* (0.266). The co-occurrence at word-level with */C/-deletion* is very low (0.085). This is as expected, given that */C/-deletion* occurs mainly in mono-syllabic pronouns. At the utterance-level, */C/-deletion* occurs in 37.9% of the utterances in which schwas were deleted.

In the last two decades, numerous studies investigated schwa deletion and the segmental, linguistic, probabilistic and extra-linguistic factors affecting whether schwa is deleted or not (e.g., [38, 39, 4]). To the best of our knowledge, this is the first study reporting with which other rules it tends to co-occur at word and utterance level.

3.2. Phonological- and reduction rules for Austrian German (SpAG)

Vowel processes *Vowel-substitution* occurs in 33% , *Vowel-monophthongation* in 25% and *Vowel-diphthongation* in 34% of the word tokens where the rule applied. Since all of these three rules apply to the stressed vowel of a word, it is not surprising that they do not co-occur at word-level at all. At utterance level, however, they occur together: In utterances with *Vowel-substitution*, *Vowel-diphthongation* (0.369) occurs more frequently than *Vowel-monophthongation* (0.240). Given that the vowel substitutions incorporated (e.g., /a/ to /oː/) apply to speakers across different regions, whereas the other two rules are, according to earlier (impressionistic) studies, more frequent in certain regions (vowel-diphthongation in Styria, vowel-monophthongation in Vienna [15]), the on higher co-occurrence with diphthongation than with monophthongation may be caused by different speakers of different regional background. Furthermore, a phonetically observed monophthongation might either be caused by a phonological process (documented for Vienna), going along with a vowel lengthening (i.e., /aɛ/ to [eː]), but could also simply stem from a reduction at high speech rate. Further detailed acoustic analyses will be needed to untangle these possible origins for our observation.

/C/-deletion Syllable-final /C/ was deleted in on average 35% of the tokens where the rule applied. This rule very often occurs in (short) pronouns (e.g., *dich* "you" /dˈɪç/ as [dˈiː], and thus to all of those tokens no other rule applies. For the rest of the cases where syllable-final /C/ was deleted, the outgoing edge weights show that it most frequently co-occurred with */R/-reduction* (1.0), *schwa-deletion* (0.621) and *plosive-deletion* (0.579). Whereas */R/-reduction* and *schwa-deletion* are generally highly frequent also in other word tokens, plosives are deleted much more often in words with deleted /C/ (0.579) than in words where other rules were applied (incoming weights ranging between 0.144 and 0.341). Despite */C/-deletion* being a very salient characteristic of AG, it has only been (quantitatively) acoustically analyzed in one study [7], showing that it occurs in 24.1% of the tokens where the rule applies in spontaneous speech (12 speakers of GRASS), but only in 0.6% in read speech. For spontaneous German, Rodgers et al. [37] reported 8.75% of /C/s to be deleted, however without providing contextual information.

4. Conclusions

The aim of this paper was to introduce graph theory as a method to study pronunciation variation. Our analysis of 181.230 tokens of spontaneous Austrian German revealed that the most frequently observed rules are reduction processes that are also typical for spontaneous German (i.e., /R/ reduction and schwa deletion) and that they frequently co-occur with phonological processes typical for the Austrian German variety at word and utterance level.

Preliminary experiments suggest that graphs such as $\mathcal{G}^{(\mathrm{w})}$ and $\mathcal{G}^{(\mathrm{u})}$ created for individual speakers may yield insights into how these speakers apply pronunciation rules. We believe that the resulting graphs cluster in a manner that corresponds with speaker demographics (regional origin, age, gender, etc.). Future work shall investigate this further and propose "prototypical" pronunciation graphs for certain demographics. To conclude, on the example of spontaneous Austrian German, this paper showed that the approach presented opens new possibilities to analyze pronunciation variation in large corpora of different speaking styles.

5. Acknowledgements

The work of B. C. Geiger was funded by grant P-32700-NB from FWF (Austrian Science Fund). We thank X. Kogler and S. Wepner for sharing their scripts to parse the pronunciation variants in GRASS. The Know-Center is funded within the Austrian COMET Program.

6. References

[1] M. Ernestus, I. Hanique, and E. Verboom, "The effect of speech situation on the occurrence of reduced word pronunciation variants," *Journal of Phonetics*, vol. 48, pp. 60–75, 2015.

[2] H. van den Heuvel and C. Cucchiarini, "/r/-deletion in Dutch: Rumors or reality?" in *r-atics: Sociolinguistic, phonetic and phonological characteristics of /r/*, H. Van de Velde and R. van Hout, Eds. Brussels: Etudes and Travaux - ILVP/ULB, 2001, vol. 4, pp. 185–198.

[3] B. Schuppler, W. van Dommelen, J. Koreman, and M. Ernestus, "How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level." *Journal of Phonetics*, vol. 40, pp. 595–607, 2012.

[4] Y. Wu, M. Adda-Decker, and L. Lamel, "Schwa deletion in word-initial syllables of polysyllabic words: Investigations using large French speech corpora," *Journal of Monolingual and Bilingual Speech*, vol. 2, no. 2, p. 269–289, 2020.

[5] M. Adda-Decker and N. D. Snoeren, "Quantifying temporal speech reduction in French using forced speech alignment," *Journal of Phonetics*, vol. 39, no. 3, pp. 261–270, 2011.

[6] C. Van Bael, R. H. Baayen, and H. Strik, "Segment deletion in spontaneous speech: A corpus study using mixed effects models with crossed random effects," in *Proceedings of Interspeech*, 2007, pp. 2741–2744.

[7] B. Schuppler, M. Adda-Decker, and J. A. Morales-Cordovilla, "Pronunciation variation in read and conversational Austrian German," in *Proceedings of Interspeech*, 2014, pp. 1453–1457.

[8] S. Moosmüller and C. Ringen, "Voice and aspiration in Austrian German plosives," *Folia Linguistica*, vol. 38, pp. 43–62, 2004.

[9] S. Moosmüller, C. Schmid, and J. Brandstätter, "Standard Austrian German," *Journal of the International Phonetic Association*, vol. 45, no. 3, p. 339–348, 2015.

[10] D. Klaaß, "Untersuchungen zu ausgewählten Aspekten des Konsonantismus bei österreichischen Nachrichtensprechern," *Duisburger Papers on Research in Language and Culture*, vol. 74, pp. 7–277, 2008.

[11] A. Jackschina, B. Schuppler, and R. Muhr, "Where /aR/ the /R/s in Standard Austrian German?" in *Proceedings of Interspeech*, 2014, pp. 1698–1702.

[12] S. Moosmüller, *Vowels in Standard Austrian German. An Acoustic-Phonetic and Phonological Analysis*. Habilitation, University of Vienna, 2007.

[13] J. Stelzl, "Soziophonologische Variation in der Kleinstadt Murau," Institut für Sprachwissenschaft, Masterthesis. University of Graz, Austria, 2009.

[14] E. Steiner and R. Vollmann, "Fragebuch zur Sprachdatenerbebung in der Steiermark," Karl-Franzens University Graz, Tech. Rep., 2010.

[15] M. S., "Diphthongs and the process of monophthongization in Austrian German: A first approach," in *Proceedings of Eurospeech*, 1997, pp. 787–790.

[16] R. Vollmann, "Phonetics of informal speech: The Vienniese monophthongation." *Studia Phonetica Posnaniensia*, vol. 5, pp. 87–100, 1996.

[17] R. Vollmann and S. Moosmüller, "'natürlichesDriften' im Lautwandel: die Monphthongierung im österreichischen Deutsch." *Zeitschrift für Sprachwissenschaft*, vol. 20, no. 1, pp. 42–65, 2001.

[18] J. Luttenberger and J. Fanta-Jende, ""Viennese" monophthongs: present - marked - given? on intra-individual variation of the realization of /ae/ and /ao/ diphthongs in Standard German pronunciation in rural Austria," *The Phonetician*, vol. 117, pp. 45–58, 2020.

[19] C. Schmid, S. Moosmüller, and C. Kasess, "Sociophonetics of the velarized lateral in the Viennese dialect," in *Proceedings of ICPhS*, Glasgow, Scottland, 2015.

[20] J. Luttenberger, N. Weihs, and M. Pucher, "L-velarization in Austrian German dialect," in *Proceedings of the NWAV49*, Austin, 2021.

[21] C. Horn, "Diglossia in the arab world - educational implications and future perspectives," *Open Journal of Modern Linguistics*, vol. 5, pp. 100–104, 2015.

[22] A. Stepkowska, "Diglossia: A critical overview of the Swiss example," *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, vol. 129, pp. 199–209, 2012.

[23] P. Auer, "Europe's sociolinguistic unity, or: A typology of european dialect/standard constellations," *Perspectives on variation: Sociolinguistic, historical, comparative*, vol. 7, pp. 7–42, 2005.

[24] B. Soukup, "The strategic use of Austrian dialect in interaction: A sociolinguistic study of contextualization, speech perception and language attitudes," Ph.D. dissertation, Georgetown University, December 2007.

[25] A. Kim and K. Korecky-Kröll, "Prepositional phrases in German in Austria – identifying patterns of variation," *Open Linguistics*, vol. 7, no. 1, pp. 476–510, 2021.

[26] B. Schuppler, M. Hagmüller, J. A. Morales-Cordovilla, and H. Pessentheiner, "GRASS: the Graz corpus of Read And Spontaneous Speech," in *Proc. LREC*, 2014, pp. 1465–1470.

[27] B. Schuppler, M. Hagmüller, and A. Zahrer, "A corpus of read and conversational Austrian German," *Speech Communication*, vol. 94, pp. 62–74, 2017.

[28] B. Schuppler, M. Hagmüller, and A. Zahrer, "A corpus of read and conversational Austrian German," *Speech Communication*, vol. 94C, pp. 62–74, 2017.

[29] B. Ludusan and B. Schuppler, "To laugh or not to laugh? the use of laughter to mark discourse structure," in *Proc. of SIGDIAL*, 2022, pp. 76–82.

[30] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi Speech Recognition Toolkit," *Proc. of IEEE ASRU Workshop*, 2011.

[31] S. Wasserfall, *Automatic speech segmentation using Kaldi*. Master Thesis from Graz University of Technology, 2020.

[32] J. Linke, S. Wepner, G. Kubin, and B. Schuppler, "Using Kaldi for automatic speech recognition of conversational Austrian German," *arXiv*, p. doi: 10.48550/ARXIV.2301.06475, 2023.

[33] U. D. Reichel, "PermA and Balloon: Tools for string alignment and text processing," in *Proc. of Interspeech*, 2012.

[34] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 314–345, 2001. [Online]. Available: http://www.praat.org,lastviewed18-08-2018

[35] B. Schuppler, M. Ernestus, O. Scharenborg, and L. Boves, "Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions," *Journal of Phonetics*, vol. 39, pp. 96–109, 2011.

[36] R. Muhr, *Österreichisches Aussprachewörterbuch – Österreichische Aussprachedatenbank*. Frankfurt/M., Wien u.a. 525 S. mit DVD: Peter Lang Verlag, 2007.

[37] J. E. J. Rodgers, P. Helgason, and K. J. Kohler, "Segment deletion in the Kiel corpus of spontaneous speech *," *aipuk*, vol. 32, pp. 121 – 176, 2010.

[38] C. Kuipers and W. van Donselaar, "The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch," *Language and Speech*, vol. 41, pp. 87–108, 1997.

[39] I. Hanique, B. Schuppler, and M. Ernestus, "Morphological and predictability effects on schwa reduction: The case of Dutch word-initial syllables," in *Proc. of Interspeech*, 2010, pp. 933–936.