Pronunciation variation in read and conversational Austrian German

Barbara Schuppler[1], *Martine Adda-Decker*[2], *Juan A. Morales-Cordovilla*[1]

[1]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria
[2]LPP-CNRS, Univ. Paris Sorbonne Nouvelle, Paris

b.schuppler@tugraz.at, madda@univ-paris3.fr, moralescordovilla@tugraz.at

Abstract

This paper presents the first large-scale analysis of pronunciation variation in conversational Austrian German. Whereas for the varieties of German spoken in Germany, conversational speech has been given noticeable attention in the fields of linguistics and automatic speech recognition, for conversational Austrian there is a lack in speech resources and tools as well as linguistic and phonetic studies. Based on the recently collected *GRASS* corpus, we provide (methods for the creation of) a pronunciation dictionary and (tools for the creation of) broad phonetic transcriptions for Austrian German. Subsequently, we present a comparative analysis of the occurrence of phonological and reduction rules in read and conversational speech. We find that whereas some rules are specific for the Austrian Standard variant and thus occur in both speech styles (e.g., the realization of /z/ as [s]), other rules are specific for conversational speech (e.g., the realization of /aː/ as [oː]. Overall, our results show that less words are produced with the citation form for conversational Austrian German (37.8%) than for other languages of the same style (e.g., Dutch conversations: 56%).

Index Terms: Austrian German, Pronunciation Variation, Conversational Speech, Speech Reduction, Forced Alignment

1. Introduction

Whereas for the varieties of German spoken in Germany, noticeable attention has been given to conversational speech in human and automatic speech recognition (e.g., [1, 2]), for conversational Austrian there is a lack both in speech resources and (speech technology) tools as well as in linguistic and phonetic studies. The European project *DIRHA*[1], which aims at creating a voice enabled automated home, and the Austrian project *Cross-layer pronunciation modeling for conversational speech*, have the objective to fill this gap. Based on the recently created first corpus of read and conversational Austrian German (AG henceforth) (*GRASS* corpus [3]), the goals of the present study are (1) to provide a pronunciation dictionary covering read and conversational AG (2) to analyze the frequency of phonological and reduction rules in read and conversational speech and (3) to provide an automatic method for the comparison of pronunciation variation in large corpora of different speech styles.

1.1. Previous phonetic studies on Austrian German (AG)

General descriptions of the pronunciation characteristics of standard (as spoken by trained broadcast speakers) AG (as compared to German spoken in Germany) can be found in the German Pronunciation Dictionary [4]. The first corpus-based pronunciation dictionary (ADABA) was created by Muhr (2007) [5, 6], using single words recordings by trained Austrian speakers. Also, the studies [7, 8, 9, 10, 11] present characteristics of standard AG. The studies by Moosmüller et al. [12] and by Klaß [13] focus on consonant realizations, whereas vowel realizations are studied in [14, 15]. Table 1 provides an overview of characteristics along with sample words and their citation forms in German and standard AG.

A relatively large amount of studies investigated the properties of certain regional variants. For instance, for the province of Styria, more than 30 speakers were interviewed [16] . Detailed acoustic analysis of /e/ showed that the second formant is significantly higher than for standard AG. Another study by Stelzl (2009) investigated the frequency of several phonological rules of speakers in the area of the Styrian town Murau [17]. For the Viennese dialect, one line of research focused on monophthongation processes (e.g., [18]) and its spreading to other Austrian cities (e.g., [15, 19, 20]).

The studies cited so far are based on either impressionistic or acoustic analyses of a small number of speakers in either non-spontaneous or semi-spontaneous (interview) conditions. Recently, the *Graz corpus of Read and Spontaneous Speech (GRASS)* has been created [3]. It contains read sentences, elicited commands and free conversations from 38 speakers. Thus, [3] is an ideal material for investigation of the distribution of phonological rules, reduction phenomena and regional variants for the different speech styles. For the conversational speech part, we expect to find higher frequencies for some of the phenomena mentioned in Table 1 and additional different phenomena to what has previously been documented for AG.

1.2. Forced Alignment as a tool for speech style comparison

Due to a growing interest in conversational speech in the last decade, large conversational speech corpora have been created for many languages (e.g., for English [21], for Dutch [22]), and for French [23]). In order to facilitate the annotation and phonetic analysis of such large speech corpora, automatic methods have been developed. One ASR based method, which showed good results for the analysis of pronunciation variation and speech reduction at the broad segmental level, is *forced alignment*. It has been used for instance as a tool to model reduced variants in German [24], and to analyze temporal speech reduction in French [25]. Whereas, some have used forced alignment to give an overview of the frequency of certain phonological and reduction rules in a given corpus (e.g., for Dutch [26, 27]), others have adapted the method to study the conditions for specific realizations of one sound (e.g., [28] studied /l/ variation in English, [29] studied /t/ variation in Dutch). In this paper, we will use forced alignment for studying pronunciation variation in different AG speech styles.

[1]For more information see: www.dirha.fbk.eu.

Table 1: **Summary of characteristics of standard Austrian German (AG) as compared to the German citation form (GC).** Subsegmental characteristics, which are not analyzed in the current paper, are marked with an asterisk.

Characteristics (Plosives)	GC	AG	Characteristics	GC	AG
Lenis plosives are voiceless *backen* ‘baking’	bˈakən	pˈakən	∗ Different use of long and short vowels *Spaß* ‘fun’	ʃpˈas	ʃpˈaːs
∗ Low aspiration of fortis plosives *Tisch* ‘table’	tʰˈɪʃ	t’iʃ	Devoicing of alveolar fricatives *Sonne* ‘sun’	zˈɔnə	sˈɔnə
Intervocalic lenition of fortis plosives *weiter* ‘further’	vˈaːɪtɐ	vˈaːɪdɐ	Monophthongations *keine* ‘none’	kʰˈaːɪnə	kˈaːnə
Spirantization of lenis plosives *Barbara*	bˈaʀbaʀa	pˈaːvaʀa	Assimilation of nasals *fünf* ‘five’	fˈynf	fˈymf
Inter-vocalic plosive deletion before /n/ *gegen* ‘against’	gˈeːgən	gˈeːŋ	Word-final syllable *-ig* as /ik/ *lustig* ‘funny’	lˈʊstɪç	lˈʊstɪk

2. Materials and methods

2.1. Speech material

Our study is based on the *Graz corpus of Read And Spontaneous Speech (GRASS)* [3]. It contains for each speaker: 62 phonetically balanced sentences, 20 commands elicited with pictures and half an hour of conversation ($\sim$ 6 500 tokens/speaker). Since conversations were between friends or family members who were freely talking about everyday topics (in the absence of an experimenter), the style of the speech is informal and casual. The corpus was collected with speech technology applications in mind and thus fulfills the requirements for automatic processing (e.g., [27]): the recordings took place in a soundproof studio with both head-mounted and large-membrane microphones at 48kHz. The orthographic transcriptions were created using PRAAT [30] on speech stretches below six seconds. They contain detailed annotations of hesitations, disfluencies and other vocal and non-vocal noises.

From this corpus, we extracted read sentences, elicited commands and conversational speech from 12 (out of 38) speakers. The 12 speakers were gender balanced, with a similar average age per group (age_{male}=33, $\text{age}_{\text{female}}$=32). They were born in one of the eastern provinces of Austria and they all have lived most of their adulthood in Graz. In the conversational speech, there were two gender-mixed pairs, two between women and two between men. From the conversations we excluded interjections, disfluencies, response tokens (e.g., ’mhm’, ’aha’) as well as broken words. This leaves us with a total of 22 260 word tokens for our analyses.

2.2. Forced alignment with multiple pronunciation variants

The forced alignment is carried out using a HTK system [31] with standard German acoustic phone models and our AG lexicon with multiple pronunciation variants of all observed word types. The input data (speech files and their manual transcriptions) are processed by the alignment system determining for each transcribed word the most likely pronunciation variant and the corresponding phone segment boundaries. The 35 acoustic models are continuous density 3-state monophones with 5 Gaussians per state. The models have been trained on a total of 5 000 utterances from 50 German speakers of the BAS read speech corpus [32]. The acoustic parameterization can be described as follows: 16 kHz sampling frequency, frame shift and length of 10 and 32 ms, 1024 frequency bins, 26 mel channels and 13 cepstral coefficients with cepstral mean normalization. After adding delta and delta-delta features, each final MFCC vector has 39 components.

2.2.1. Pronunciation lexicon

The pronunciation lexicon was created as follows: First, for each word type (#= 4 364) a canonical pronunciation (German standard) was created with the Balloon tool [33], providing syllabic and morphological boundaries, as well as primary and secondary stress. Second, these pronunciations were corrected manually. Errors mainly concerned proper names, foreign words and compounds.

In a third step, 32 phonological and reduction rules were applied to the canonical pronunciations. Table 2 gives an overview. The first part of the table (ID 1.1- 1.3) shows general coarticulation, assimilation and reduction rules which are also typical for spontaneous German spoken by speakers originating from Germany. These rules include those mentioned by Wesenick et al. [34] and by Schiel, F. [35], which we narrowed by specifying them for specific syllable structures and stress patterns. The second part of the table shows rules which were formulated on the basis of the literature on standard AG (cf. section 1.1). Several of these rules overlap with those formulated for the develpment of text-to-speech engines for standard AG [36, 37] and for the Viennese dialect [38].

Finally, variants were created manually for highly frequent words (mostly pronouns and verbs) whose typical Austrian pronunciation cannot be created by rules, or where the application of a rule to all words in the lexicon would create unrealistic variants (*nein* ‘no’ with the citation form /nˈaɪn/ pronounced as nˈaː). The final pronunciation lexicon had 6.8 variants per word type.

2.2.2. Validation

To validate the acoustic models and our pronunciation variants, a phonetically trained transcriber corrected part of the forced aligned material. Instructions to the transcriber were to correct the phone sequence (i.e., to substitute, delete or insert phone labels). The material for validation consisted of two minutes of each conversation and 8 read utterances of each speaker (total: 12 951 and 2 434 phones respectively). Overall, there was a 18.5% discrepancy between the phone labels of the forced alignment and the manually corrected ones. This was mainly due to substitutions (16.9% in read and 15.1% in conversational speech), with a small number of insertions (1.4% in read and 2.1% in conversational speech) and very few deletions (0.3% in read and 1.0% in conversational speech). These low numbers for insertions and deletions show that our set of reduction rules indeed cover most of the reduced variants in GRASS. These deviations between automatic and manual transcriptions are smaller than previously reported (24.3% [39]) and well within the range of reported inter-annotator discrepancies on manual transcriptions (5.6% - 21.2%) [40]. We may thus consider our data as a reliable basis for the study of pronunciation variation.

3. Results and discussion

To analyze the frequency of phonological and reduction rules we proceeded as follows: When generating the pronunciation variants for the lexicon, we logged the rules that contributed to their creation. Then, the pronunciation variant which best matched the speech signal was chosen by the forced alignment procedure. Then the frequency of occurrence of the rules was computed separately for the read and the conversational speech. Finally, we used a Welch's t-test to test whether pronunciation variation differed significantly in read and conversational speech. Henceforth, we will show t- and p-values for significant results.

According to our expectations, our results show that significantly ($t = -53.2$, $p < .0001$) more words are produced with the citation form in AG read sentences (67.9 %) than in conversational AG (37.8%). The low percentage of word tokens produced with the German citation form in standard Austrian read speech reflects that the German citation forms do not well represent standard AG. It is thus not surprising, that for conversational AG we find fewer canonical pronunciations than previously for other germanic languages (e.g., 56% canonical pronunciations in Dutch conversations [27]).

Table 2 shows a summary of our analysis on the frequencies of occurrence of the phonological- and reduction rules. We assign each rule a rule-ID and illustrate it with an example. Rule-IDs ending in 'v', 'c,' and 's' affect vowels, consonants and syllables respectively. For each rule, we give the numbers for word types and word tokens in order to show whether a rule is general or specific for a small number of word types.

3.1. Assimilations and deletions for all varieties of German

First, we analyze those assimilation and reduction rules which can be expected in all varieties of German (cf. Table2, ID 1.1.c-1.3.c). Our data shows that progressive and regressive assimilation of voice and place of articulation of consonants are approximately as frequent in read speech (53.4% of the tokens) as in conversational speech (51.6%). Also schwas, which precede /n/ in unstressed syllables, are nearly as often absent in read (57.4%) as in conversational speech (63.2%). This suggests, that these rules from the phonological literature are speech-style independent in AG. In average over all segmental contexts, plosive- and schwa deletion occurs nearly twice as often in conversational than in read speech. Overall, we find schwa deletion to be similarly frequent in AG (38.1%) as previously reported for conversational Dutch (41.0%) [27] and for fast spoken American English (43%) [41]. Thus, schwa deletion is as frequent in AG as in other Germanic languages.

3.2. Substitutions and deletions typical for AG

This section presents our results for those pronunciation rules which are specific for AG. Lines 2.1.c - 2.10.s in Table 2 show the rules in order of decreasing difference between read and conversational speech (i.e., the later a rule appears in the table the more speech style specific it is). Our data shows that several rules concerning the AG pronunciation are similarly frequent in read as in spontaneous speech: the lenition of fortis plosives, the spirantization of lenis plosives (ID 2.1.c), the devoicing of alveolar fricatives (2.2.c) and the realization of the word-final syllable *-ig* as /ik/ (2.3.c). The high frequency of these rules in read speech is as expected given that they are frequently mentioned in the literature on standard AG (e.g., [9]) and have even been incorporated in text-to-speech engines [36, 37, 38].

Lines 2.4.c to 2.6.c in Table 2 show rules which are significantly more frequent in conversational speech than in read speech: vocalization or deletion of /R/ ($t = -3.09$, $p < .01$), reduction of the syllable *-gen* /gən/ to /ŋ/ ($t = 3.16$, $p < .01$) and vocalization or deletion of /l/ ($t = 8.08$, $p < .0001$). The most frequent among these rules is vocalization or deletion of /R/, which occurs in 49.1% of the word tokens in read speech, compared to 58.2% in conversational speech. These are very high numbers given that the canonical pronunciation lexicon already includes vocalization of /R/ in coda-position. Our results are in line with [9], who mentioned that vocalization of /R/ is its most frequent realization, also in his non-spontaneous Austrian speech material.

Finally, lines 2.7.c - 2.9.c show rules which hardly occur in read speech (between 5.3% and 0.0% of the tokens) but which are still relatively frequent in conversational AG (between 24.1% and 33.3% of the tokens): the deletion of word-final [ç] or [x] ($t = 29.2$, $p < .0001$), the substitution of full vowels and diphthongs ($t = 48.5$, $p < .0001$), and the deletion of the morpheme *ge-* ($t = 15.9$, $p < .0001$). We consider these rules as speech-style dependent, or rather, typical for conversational AG. In read speech, these rules are limited to a small number of word types. For instance, word-final [ɕ] and [x] - deletion exclusively occurs with the words *ich* 'I' with the citation form /ˈiːç/ and *noch* 'yet' with the citation form /nˈoːx/. In our read speech material, *ich* was pronounced in 8 out of 114 tokens as [ˈiː], and *noch* was pronounced in two out of 18 tokens as [nˈoː].

3.3. Not-generalizable variants for highly frequent words

This section presents our results concerning not-generalizable pronunciation variants which only occur in highly frequent words. In general, we observed that these not-generalizable pronunciation variants occur significantly ($t = 65.9$, $p < .0001$) less frequently in read speech (17.9%) than in conversational speech (49.7%). Even though this rule only affects 104 word types in the conversational speech, 25.4% of all word tokens in the conversational speech are produced with one of these not-generalizable variants.

The pronunciation variants of this category affect highly frequent words (e.g., pronouns and verbs). They combine substitutions and deletions: For instance, the word *nein* meaning 'no' with the citation form /nˈaɪn/ is pronounced as [nˈa] 271 times out of the 340 occurrences in the conversational speech material. The word *wir* meaning 'we' is produced 183 times in its citation form /wˈiɐ/ and 173 times as [mˈa]. Some variants of this category also combine substitutions, deletions and insertions. For instance, the modal verb *könnte* 'could' is 6 times realized in its citation form kˈɔntə, 4 times as kˈenat, and 11 times with another reduced variant.

4. Concluding remarks

In this paper, we have presented the first quantitative analysis of pronunciation variation in read and conversational AG based

Table 2: **Results: Rules and their frequencies.** Column 'Tokens' and 'Types' contain the absolute number of word tokens/word types to which a rule could apply and the relative frequency (%) with which a pronunciation variant created with the respective rule was found in *GRASS* compared to the total number of word tokens/word types for which a pronunciation variant was created. 'RS' stands for read speech, 'CS' for conversational speech.'v' in Rule IDs stands for vowels, 'c' for consonants, and 's' for whole syllables.

				Tokens		**Types**	
ID		Citation form	*GRASS*	RS	CS	RS	CS
0.0	Citation form			67.9%	37.8%	70.4%	55.1%
				5 519	16 741	757	2 647
1	**Assimilations and deletions for all varieties of German**						
1.1.c	Progressive or regressive assimilation			53.4%	51.6%	68.3%	71.4%
	erworben 'received'	ɛɐvˈɔɐbn	ɛɐvˈɔɐbm	127	433	28	185
1.2.v	Deletion of unstressed schwa			15.1%	38.1%	22.6%	51.7%
	bereitet 'caused'	bˈəʀaɪtət	bʀˈaɪtət	275	3 566	79	914
	Specifically in context /@n/			57.4%	63.2%	65.9%	70.9%
	halten 'to hold'	hˈaltən	hˈaltn	597	2 235	597	737
1.3.c	Plosive deletions in all positions			15.9%	27.8%	25.3%	37.9%
	wichtig 'important'	vˈɪçtɪk	vˈɪɕtɪ	385	3 824	118	821
2	**Substitutions and deletions typical for Austrian German**						
2.1.c	Lenition of fortis and spirantization of lenis plosives			27.7%	32.1%	43.5%	33.9%
	weiter 'further'	vˈaːɪtɐ	vˈaːɪdɐ	127	997	37	204
2.2.c	Devoicing of alveolar fricatives			55.4%	52.7%	65.7%	81.9%
	Sonne 'sun'	zˈɔnə	sˈɔnə	346	1 627	67	424
2.3.c	Word-final -ig realized as /ik/			83.3%	76.9%	84.5%	95.1%
	lustig 'funny'	lˈʊstɪç	lˈʊstɪk	70	240	28	96
2.4.c	Vocalization or deletion of /r/			49.1%	58.2%	50.0%	49.0%
	Garten 'garden'	gˈaʀtən	gˈaːtən	181	747	20	140
2.5.s	Syllable /gən/ > /ŋ/			36.0%	47.4%	41.2%	49.1%
	ehemaligen 'former'	ˈeːəmaːlɪgən	ˈeːəmaːlɪŋ	44	369	14	109
2.6.c	Vocalization or deletion of /l/			14.3%	39.8%	16.1%	24.2%
	Resultat 'result'	ʀezʊlˈtaːt	ʀezʊɪˈtaːt	65	772	23	232
2.7.c	Deletion of syllable-final [ç] or [x]			0.6%	24.1%	1.9%	22.6%
	dich 'for you'	dˈɪç	dˈiː	314	980	36	89
2.8.v	Full vowel substitutions			5.3%	33.3%	12.5%	46.9%
	sagen 'to say'	zˈagən	sˈoːgn	87	4 385	34	546
2.9.s	Deletion of morpheme *ge-*			0.0%	30.9%	0.0%	43.0%
	gegangen 'ran'	gəgˈaŋən	gˈaŋən	83	253	16	77
3	**Not-generalizable variants for highly frequent words**						
	All manually created variants			17.9%	49.7%	51.4%	79.4%
	wir 'we'	wˈiɐ	mˈa	399	11 445	38	104

on 22 260 word tokens from 12 speakers from the GRASS corpus. The first aim of the presented study was to incorporate major AG phonological and reduction rules into a pronunciation lexicon. For this purpose, we applied 32 rules to the standard German citation forms of the words and we added variants manually. The validation of the transcriptions created with our set of pronunciation variants showed good results (18.5% discrepancy, similar as for human labellers), suggesting that most of the variation found in the GRASS corpus is covered by our pronunciation lexicon.

The second aim of the present study was to analyze the frequency and distribution of phonological and reduction rules in read and conversational speech. We found that whereas some rules are specific for AG and thus occur in both speech styles (e.g., the realization of /z/ as [s]), other rules are specific for conversational speech (e.g., the realization of /aː/ as [oː]. Overall, our results show that less words are produced with the citation form for conversational AG (37.8%) than for other languages of the same speech style (e.g., Dutch conversations: 56% [27]).

The findings presented here will inform further studies on pronunciation variation in Austrian German, both in the field of linguistics and speech technology. We have created a pronunciation dictionary, which is – along with the derived probabilities for the occurrence of the variants – suitable for the incorporation into an ASR system for AG. Furthermore, our study confirms that forced alignment is a suitable tool for the comparison of pronunciation variation in different speech styles. Its use is not limited to the analysis of different speech styles, but it can be expanded to the analysis of regional, social and speaker-dependent aspects of pronunciation variation.

5. Acknowledgements

The work by Barbara Schuppler was supported by a Hertha-Firnberg grant from the FWF (Austrian Science Fund). The work of Juan A. Morales-Cordovilla was funded by the European project DIRHA (FP7-ICT-2011-7-288121). This work was also supported by the French Investisements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

6. References

[1] IPDS, "CD-ROM: The Kiel Corpus of Spontaneous Speech, vol i- vol iii," Corpus description available at http://www.ipds.uni-kiel.de/forschung/kielcorpus.de.html (last viewed 25/04/2011), 1997.

[2] A. Lücking, K. Bergman, F. Hahn, S. Kopp, and H. Rieser, "The Bielefeld speech and gesture alignment corpus (SaGA)," in *Proceedings of LREC 2010 Workshop: Multimodal Corpora- Advances in Capturing, Coding and Analyzing Multimodality*, 2010, pp. 92–98.

[3] B. Schuppler, M. Hagmüller, J. A. Morales-Cordovilla, and H. Pessentheiner, "GRASS: The Graz corpus of Read and Spontaneous Speech," in *Proceedings of LREC*, 2014.

[4] P. Wiesinger, "Die Standardsprache in Österreich," in *Deutsches Aussprachewörterbuch.*, E.-M. Krech, E. Stock, U. Hirschfeld, and L. C. Anders, Eds. Wien, Austria: Walter de Gruyter, 2009, pp. 229–258.

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

[6] ——, "The pronouncing dictionary of Austrian German (AGPD) and the Austrian phonetic database (ADABA): Report on a large phonetic resources database of the three major varieties of German," in *Proceedings of LREC*, 2008, pp. 3093–3100.

[7] S. Moosmüller and W. U. Dressler, "Hochlautung und soziophonologische Variation in Österreich," *Jahrbuch für Internationale Germanistik*, vol. 20, no. 2, pp. 82–90, 1988.

[8] S. Moosmüller, *Hochsprache und Dialekt in Österreich*. Böhlau, Wien, Austria, 1991.

[9] M. Bürkle, "Österreichische Standardaussprache: Vorurteile und Schibboleths." in *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen.*, R. Muhr, R. Schrodt, and P. Wiesinger, Eds. Wien, Austria: Verlag Hölderer-Pichler-Tempsky, 1995, pp. 236–249.

[10] F. T. Stubkjær, "Überlegungen zur Standardaussprache in Österreich." in *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachpolitische Aspekte einer nationalen Variante des Deutschen.*, R. Muhr, R. Schrodt, and P. Wiesinger, Eds. Wien, Austria: Verlag Hölderer-Pichler-Tempsky, 1995, pp. 250–271.

[11] S. Moosmüller, "Aussprachevarianten im Österreichischen Standarddeutsch," in *Interpersonelle Kommunikation - Analyse und Optimierung*, I. Bose and B. Neuber, Eds. Frankfurt: Lang, 2011, p. in press.

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

[13] 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.

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

[15] ——, "The process of monophthongization in Austria (reading material and spontaneous speech)," in *Papers and Studies in Contrastive Linguistics*, 1998, pp. 9–25.

[16] E. Steiner and R. Vollmann, "Fragebuch zur Sprachdatenerhebung in der Steiermark," Karl-Franzens Univ. Graz, Tech. Rep., 2010.

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

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

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

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

[21] M. A. Pitt, K. Johnson, E. Hume, S. Kiesling, and W. D. Raymond, "The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability," *Speech Communication*, vol. 45, pp. 89–95, 2005.

[22] M. Ernestus, "Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface," Ph.D. dissertation, LOT, Vrije Universiteit Amsterdam, The Netherlands, 2000.

[23] F. Torreira, M. Adda-Decker, and M. Ernestus, "The Nijmegen Corpus of Casual French," *Speech Communication*, vol. 52, no. 3, pp. 201–212, 2010.

[24] M. Adda-Decker and L. Lamel, "Modeling reduced pronunciations in German," *Phonus 5, Institute of Phonetics, University of the Saarland*, pp. 129–143, 2000.

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

[26] C. Van Bael, "Validation, automatic generation and use of broad phonetic transcriptions," Ph.D. dissertation, Radboud Universiteit Nijmegen, Nijmegen, October 2007.

[27] 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.

[28] J. Yuan and M. Liberman, "Investigating /l/ variation in English through forced alignment," in *Proceedings of Interspeech*, 2009, pp. 2215–2218.

[29] 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.

[30] 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, last viewed 18-08-2010

[31] S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, "The HTK book (v. 3.2)," Cambridge University. Engineering Department, Tech. Rep., 2002.

[32] F. Schiel and A. Baumann, "Phondat1, corpus version 3.4," Bavarian Archive for Speech Signals (BAS), Intern. report, http://www.bas.uni-muenchen.de/Bas/BasFormatseng.html, 2006.

[33] U. D. Reichel, "PermA and Balloon: Tools for string alignment and text processing," in *Proceedings of Interspeech 2012*, 2012, p. paper no. 346.

[34] M.-B. Wesenick, "Automatic generation of German pronunciation variants," in *Proceedings of the ICSLP*, 1996, pp. 125–128.

[35] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in *Proceedings of the ICPhS 1999*, 1999, pp. 607–610.

[36] F. Neubarth, M. Pucher, and C. Kranzler, "Modeling Austrian dialect varieties for TTS," in *Proceedings of Interspeech*, 2008, pp. 1877–1880.

[37] C. Kranzler, F. Pernkopf, R. Muhr, and F. Neubarth, "Text-to-speech engine with Austrian German corpus," in *Proceedings of SPECOM*, 2009.

[38] M. Pucher, D. Schabus, J. Yamagishi, F. Neubarth, and V. Strom, "Modeling and interpolation of Austrian German and Viennese dialect in HMM-based speech synthesis," *Speech Communication*, vol. 52, pp. 164–179, 2010.

[39] C. Cucchiarini and D. Binnenpoorte, "Validation and improvement of automatic phonetic transcriptions," in *Proceedings of ICSLP*, Denver, USA, 2002, pp. 313–316.

[40] A. Kipp, M. Wesenick, and F. Schiel, "Pronunciation modeling applied to automatic segmentation of spontaneous speech," in *Proceedings of Eurospeech*, 1997, pp. 1023–1026.

[41] J. M. Dalby, "Phonetic structure of fast speech in American English," in *Phonetics and Phonology*. Bloomington: Indiana University Linguistic, 1986.