GRASS: The Graz Corpus of Read and Spontaneous Speech

Barbara Schuppler, Martin Hagmüller, Juan A. Morales-Cordovilla, Hannes Pessentheiner

Signal Processing and Speech Communication Laboratory, Graz University of Technology
Inffeldgasse 16c, A-8010 Graz, Austria
{b.schuppler, hagmueller, moralescordovilla, hannes.pessentheiner}@tugraz.at

Abstract

This paper provides a description of the preparation, the speakers, the recordings, and the creation of the orthographic transcriptions of the first large scale speech database for Austrian German. It contains approximately 1900 minutes of (read and spontaneous) speech produced by 38 speakers. The corpus consists of three components. First, the Conversation Speech (CS) component contains free conversations of one hour length between friends, colleagues, couples, or family members. Second, the Commands Component (CC) contains commands and keywords which were either read or elicited by pictures. Third, the Read Speech (RS) component contains phonetically balanced sentences and digits. The speech of all components has been recorded at super-wideband quality in a soundproof recording-studio with head-mounted microphones, large-diaphragm microphones, a laryngograph, and with a video camera. The orthographic transcriptions, which have been created and subsequently corrected manually, contain approximately 290 000 word tokens from 15 000 different word types.

Keywords: Austrian German, Read Speech, Conversational Speech

1. Introduction

Both, research in the field of linguistics and speech technology require the existence of large speech corpora, recorded at sufficiently high quality and transcribed at least at the orthographic level, which later will be used for the (semi-automatic) generation of further annotation layers (e.g., phonetic, morphological, syntactic and/or prosodic level). Whereas for the varieties of German spoken in Germany, large corpora of read and spontaneous speech have been created (e.g., the Kiel Corpus of Spontaneous Speech (IPDS, 1997) and the Verbmobil Corpus (Weilhammer et al., 2002)), for Austrian German, the available annotated speech material is very limited. For instance, the interview material by Moosmüller (1998) contains both read and spontaneous speech, but only from five speakers. The SpeechDat-At database contains telephone speech from many ($= 1000$) speakers; the spontaneous speech part, however, is restricted to the spontaneous elicitation of single words (Baum et al., 2000). The speech collected for the ADABA database (Muhr, 2008) is restricted to read speech spoken by trained speakers; Muhr (2000) collected dialogues for different scenarios between speakers from Austria (12 speakers), Germany and Switzerland in order to provide authentic listening material for second language learning. These recordings, however, contain a substantial amount of background noise. None of the mentioned corpora for Austrian German contain broadband recordings with 48 kHz, nor do they contain a sufficiently large amount of read and spontaneous speech for performing acoustic analysis which can be incorporated into models of human and automatic speech recognition.

Recently large corpora of spontaneous speech have been created among others for English (Pitt et al., 2005), Dutch (Ernestus, 2000; Schuppler, 2011) and French (Torreira et al., 2010). These corpora, however, lack read speech of the same speakers from the same region in the same recording condition, which is required in order to draw conclusions about speaking style. Finally, read speech is not only necessary as a reference in linguistic and phonetic studies but also when building up a speech recognition system, for instance, for the training and/or adaptation of acoustic models.

The *Graz corpus of Read And Spontaneous Speech* (GRASS) is designed to be suitable for both linguistic and phonetic studies as well as for the development of automatic speech recognition (ASR) and dialogue systems, comprising the following technical and content-related characteristics:

1. High-quality super-wideband recordings which enable the simulation of different acoustic environments by filtering the speech material with different measured room impulse responses.
2. Phonetically balanced sentences and digits from each speaker, as well as read and elicited commands and keywords as needed for certain dialogue-system applications.
3. Sufficient speech material from free conversations in order to model pronunciation variation and spontaneous dialogue phenomena (hesitations, fillers, overlapping speech).
4. High quality orthographic transcriptions which allow the (semi-automatic) generation of further phonetic and morpho-syntactic annotation layers.

This paper is organized as follows. First, we describe the data collection, (i.e., the speaker characteristics, the equipment and the recording procedure). Second, we describe the creation of the orthographic transcriptions. Then, Section 4. provides a short summary of the content of each of the components of the corpus. In the last section of the paper, we will give insight into two current research projects which make use of GRASS.

2. Data Collection

As a first step, we carried out a pilot with two speakers in which we tested the equipment, the recording procedure, and the resulting recording quality. After having applied the necessary modifications, we recorded the 38 speakers within two weeks following the here presented final set-up and procedure.

2.1. Speakers

The GRASS corpus contains speech produced by 38 speakers (balanced male and female, between 20 and 60 years old). They are moderately educated (at least high school diploma). Speakers were born and grew up in Austria (with the exception of the western provinces of Austria) and they currently live in Graz. These restrictions concerning the regional background of the speakers are necessary in order to reduces the dialectal variation. Table 1 provides an overview of speaker characteristics (age, gender, regional background, education, foreign languages). In addition to the characteristics mentioned in Table 1, we documented information about the speakers' size, the educational and regional background of their parents, the working area (e.g, technology, social, languages) of the speakers and their parents, their musical education, and whether they received some sort of professional pronunciation training.

Table 1: **Information about the speakers** (M= Male, F= Female). 'Years of Education' refers to the years after the obligatory secondary education. 'L1' stands for mother tongue. In 'Foreign Languages': *two* stands for two foreign languages and *more* for more than two foreign languages learned by the speaker.

		Gender	
	Total	M	F
Total	38	19	19
Year of birth (Y)			
$Y >= 1985$	12	6	6
$1985 > Y >= 1978$	20	10	10
$Y < 1978$	6	3	3
Region of childhood			
Carinthia	3	1	2
Salzburg	3	3	0
Styria	23	10	13
Upper Austria	6	3	3
Vorarlberg (Styrian parents)	2	1	1
Size in # of inhabitants			
City (> 120 000)	9	4	5
Town (16 000 - 120 000)	2	0	2
Village (4 000 - 15 000)	12	7	5
Village (< 3 000)	15	8	7
Years of education			
4 - 6	5	3	2
7 - 10	14	5	9
11 - 14	19	11	8
L1 Parents			
German = L1	34	18	16
German $\neq$ L1	4	1	3
Foreign languages			
only English	11	9	2
two	9	4	5
more	18	6	12
Experience abroad			
less than 3 months	15	10	5
3-6 months	7	4	3
6-12 months	5	1	4
more than 12 months	11	4	7

2.2. Equipment and Sound Quality

Figure 1 shows the setup of the equipment for the conversational speech component (left panel) and for the other components (right panel). We recorded the speech of all speakers in the recording studio of the SPSC Laboratory of the Graz University of Technology with a close talking headset (AKG HC-577L: i.e., HM1 and HM2 in Figure 1) and a large diaphragm microphone (AKG C414 BXLS: i.e., FM1 and FM2 in Figure 1) with attached pop screen. Additionally, all speakers were recorded with a laryngograph (i.e., a device that measures the impedance of the larnyx, which depends on the contact area of the vocal folds; recordings can be used as ground truth for F0 estimation). Finally, most of the conversations (14 out of 19) were recorded with a video camera (Canon Legria HF M31 HD Camcorder). These recordings might be used in the future for the study of gestures, which is relevant for the development of multimodal dialogue systems.

The .wav files have the format RIFF (little-endian), mono WAVE audio, uncompressed PCM 16 bit, with a sampling frequency of 48kHz. We acquired at 48 kHz sampling rate and then generated a version at 16 kHz. The average SNR over all speakers over the read speech and the commands components resulted to be 49.7dB. For the conversational speech, the SNR differed strongly between the different speakers. For the recordings with the head-mounted microphones, the lowest SNR was 35.8dB (a female speaker) and the highest was 52.8dB (a male speaker), with an average of 46.2dB for HM1 and 46.4dB for HM2; For the recordings with the large diaphragm microphone the SNR resulted to be lower in average (31.3dB for FM1 and 35.7dB for FM2), which is due to the speakers movements during the conversation. As expected, the SNR of the recordings with the laryngograph were even lower with an average of 28.4dB over all conversations. For the calculations of the SNR values presented in this section, we followed the approach presented in Hänsler and Schmidt (2004).

2.3. Recording Procedure

We first recorded the conversational speech component (CS), then the elicited commands. Only after these (semi) spontaneous tasks, we recorded the read commands and the read sentences. We chose this order of events for several reasons. First, the pairs of speakers mostly arrived at the same time in the recording studio and often simply continued their conversation (which started on their way to the studio) with no interruption[1]. Second, this order guarantees

[1]Such an interruption would mean a change in topic, but most importantly a switch to a different speech style.

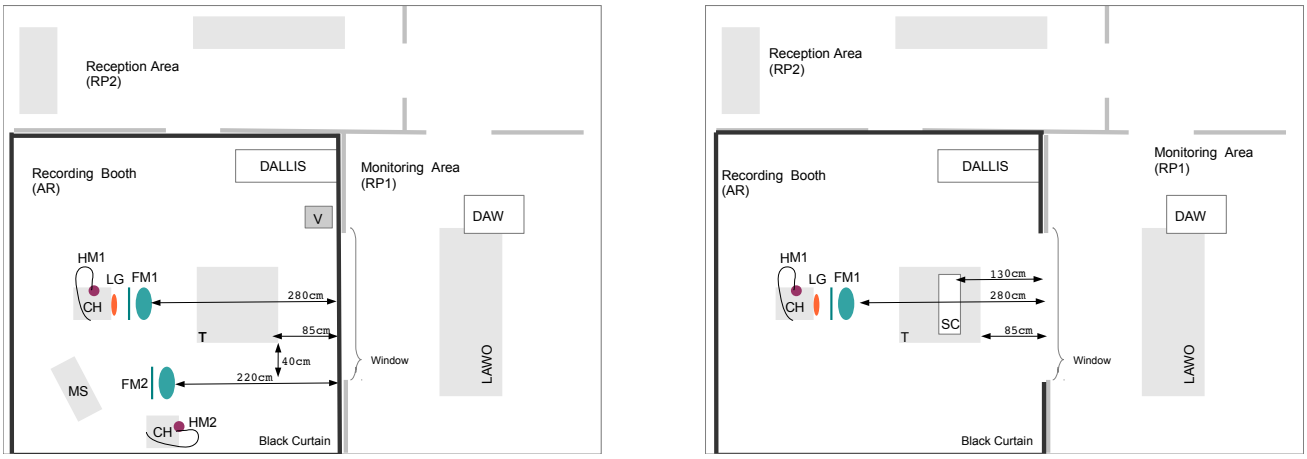

Figure 1: **Schematic setup of the equipment in the recording booth and in the monitoring area of the recording studio.** Left panel: setup during the Conversational Speech component, Right panel: setup during the Read Speech and the Commands Component. FM1 and FM2 = large diaphragm fixed microphone for speaker 1 and 2, including pop screen; HM1 and HM2 = talking head-set for speaker 1 and 2; LG = laryngograph; CH = chair; MS = music stand; T= table; SC = screen; DAW= digital audio workstation; LAWO = mixing table; V = video camera; DALLIS = microphone pre-amplifier and A/D converter.

that the elicited commands are not influenced by the reading material. Furthermore, this order of recordings ensures that speakers know as little as possible about the purpose of the recordings and about the setup in the recording studio.

3. Orthographic Transcription

For the RS and CC components of the corpus (see Section 4.2. and 4.3.), we used the original reading material to create a first transcription of the utterances and we subsequently corrected them.

For the conversational component (see Section 4.1.), six linguistically educated transcribers created orthographic transcriptions manually. The transcribers received two specific training units of three hours each. In the first unit, they got familiar with the guidelines and practiced them with some minutes of conversational speech. In the second unit, they corrected the transcriptions created until then under supervision of the first author. Only then, they created the orthographic transcriptions of all conversations. Finally, these transcriptions were corrected by a transcriber other than the one who created the first version of the transcription.

Transcribers used the open-source software PRAAT (Boersma, 2001), where for each speech file a *TextGrid* was created with separate tiers for each speaker. The details of the guidelines are strongly motivated by our previous research on automatic methods for the creation of phonetic segmentations for which orthographic transcriptions are the basis (Gubian et al., 2009; Schuppler et al., 2008; Schuppler et al., 2011). Since we also plan to create phonetic transcriptions automatically for the *GRASS* corpus, we follow the recommendations of mentioned studies. Thus, speech was segmented into very short chunks (max 4 seconds) and a very detailed annotation of laughter and other speaker noises, backchannels (e.g., *hm*), fillers (e.g., *eh, ah, uh*), broken words, overlapping speech and disfluencies. The chosen set of symbols is similar as presented in the guidelines of the BAS project (Schiel et al., 2012).

Table 2: Number of conversations for each kind of relationship between the dialogue partners.

Relationship	**Total**	**M-M**	**F-F**	**M-F**
Colleagues	3	3	0	0
Friends	10	2	4	4
Couple	3	0	0	3
Family-members	3	1	2	0
Total	19	6	6	7

Finally, transcribers created and shared a dictionary in order to unify the spelling of words, particles and multi-word expressions for which no standard spelling exists. More details about the guidelines for the orthographic transcription can be found in Appendix A.

4. The Components of the Corpus

The numbers of each component presented in the following section are based on the orthographic transcriptions described in the previous section.

4.1. The Conversational Speech Component (CS)

In total, 19 conversations were recorded, each of approximately one hour length. There were both mixed pairs and gender-homogeneous pairs (of the 19 pairs, 6 between men, 6 between women and 7 mixed). All conversations were between pairs of speakers who have known each other for several years. Table 2 shows details concerning the different relationships between the speakers.

In the studio, a small table with provocative pictures concerning the topic 'Living in Graz' was placed close to the speakers. Speakers were instructed that they could start their conversation by using these cards if they wanted to, but that in principle, they could talk about whatever topic they like. They were told that recordings would be transcribed afterwards, but that during the recordings nobody

Table 3: Summary of characteristics of the utterances produced in the Conversational Speech Component.

	# utterances
Total # of utterances	55 5571
Utterances containing lexical items	48 960
Utterances spoken in overlap	20 845
Utterances with laughter	3 593
Utterances with other speaker noises	6 165
Utterances with unintelligible speech	508
Utterances with broken words	1 567

would listen. The two speakers were left without watch nor mobile phone in the recording room for one hour. Only one quarter of the pairs of speakers started with the cards provided. Since the speakers knew each other very well (i.e., good friends, family members), they spoke freely and casually and forgot about the studio situation after only a short warming up period of several minutes. The casual speech style is also reflected by the high frequency of laughter, overlapping talk and disfluencies, shown in Table 3. Since speakers chose their conversation topics freely, the conversational speech component contains a broad lexicon covering many different topics, resulting in more than 276 000 word tokens from 14 590 word types.

4.2. The Commands Component (CC)

All speakers produced 15 commands and 5 keywords while being presented an image indicating which object inside an apartment shall be operated by a voice controlled system. In total, the recorded elicited commands and keywords contain 1 720 word tokens from 464 word types. Furthermore, speakers read 15 commands of the type *Open the window, please! Turn off the light!* and 10 keywords of the type *Wake up!* as used in common voice control system. In total, the read commands and keywords contain 3 853 word tokens from 270 different word types.

4.3. The Read Speech Component (RS)

Each of the 38 speakers read approximately 62 phonetically balanced sentences, which were taken from the Kiel Corpus of Read Speech (IPDS, 1997), and 4 telephone numbers. Additionally, the speakers read 10 utterances with a spontaneous speech like structure (i.e., sentence fragments), containing word tokens which were also expected to occur frequently in the conversational speech component of the corpus. We collected these sentences to be able to draw better comparison with the conversational speech component. In total, the read speech component consists of 2 744 utterances with 19 511 word tokens from 1 660 word types.

5. Ongoing Work Based on GRASS

5.1. Creation of a DIRHA Database

The European Project 'Distant-speech Interaction for Robust Home Applications' (DIRHA) aims to create a voice enabled automated home system using a network of microphones. Important components of this project are: automatic speech recognition (ASR), speech/non-speech detection (SDET) and speech localization (SLOC). GRASS is part of a multi-language effort to create a simulated database that provides realistic data that reflect a home situation with many microphones distributed over different rooms. Our approach is to filter clean speech utterances with the room impulse responses measured in an apartment from many source positions to the different microphones installed. In addition, background noises recorded in the same apartment are added to the reverberant signal. This approach allows the simulation of a speaker controlling the automation equipment in a realistic environment.

Table 4: Information about speakers of the **DIRHA Subcorpus** (M= Male, F= Female).

		Gender	
	Total	M	F
Total	21	10	11
Year of birth			
$Y >= 1985$	2	2	4
$1985 > Y >= 1978$	5	6	11
$Y < 1978$	3	3	6
Region of childhood			
Carithia	1	0	1
Styria	12	5	7
Salzburg	2	2	0
Tyrol	2	0	2
Upper Austria	1	1	0
Vorarlberg	2	1	1
Lower Austria	1	1	0
Size in # inhabitants			
City ($>$ 120 000)	0	0	0
Town (16 000 - 120 000)	1	1	0
Village (4 000 - 15 000)	6	4	2
Village ($<$ 4 000)	14	5	9

For this purpose, we recorded in addition to the main corpus a sub-corpus with a modified setup that was specifically targeted to be used for the simulated database in the DIRHA project (Cristoforetti et al., 2014). While most of the configuration is the same, this sub-corpus differs to the main corpus in the following details. The sub-corpus was produced by 21 speakers (10 male, 11 female) between 24 and 55 years. Half of the speakers overlap with the speakers in the main corpus. All speakers were born and grew up in Austria. Table 4 provides an overview of speaker characteristics (age, gender, regional background). The DIRHA subcorpus does not contain any conversational speech (CS), instead the speakers were asked to describe a picture in their own words. The commands component (CC) are the same, while the read speech component (RS) consists of 20 unique phonetically balanced sentences and the German 'North wind and sun' passage. The material was recorded with large diaphragm microphone (Neumann U89i) at a distance of 5-10 cm to ensure the highest possible SNR. The lowest SNR was 48.6 dB (a male speaker) and the highest was 69.5 dB (a female speaker), with an average of 60.3dB. The other parameters of the recording setup are the same as in the main corpus.

5.2. Pronunciation Variation Modeling

The presented GRASS corpus is currently used to study the conditions under which certain pronunciation variants occur in the different speech styles. Pronunciation variation is very frequent in conversational speech. For Dutch, for instance, it has been shown that 61.3% percent of the words are not produced with their citation form in spontaneous dialogues (Schuppler et al., 2011). For German, a study by Mitterer (2008) reports that 8% of the segments are deleted in spontaneous German. In general, for the different German varieties, pronunciation variation has so far not yet been analyzed in a large corpus of conversational speech produced by speakers originating from Austria.
Within the Austrian project *Cross-layer pronunciation modeling for conversational speech* the GRASS corpus is used to model statistically which factors affect the variation observed (effects of styles, regions, speaker specific characteristics, segmental context and higher-level linguistic properties). The general aim of these studies is to increase our understanding of every-day speech processes as well as to improve pronunciation modeling techniques incorporated in ASR systems. This project will also deliver a pronunciation dictionary for Austrian German and a broad phonetic segmentation of the GRASS corpus.

6. Corpus Availability

The corpus-webpage, which can be found at `www.spsc.tugraz.at`, provides more details about the collection of the GRASS corpus along with audio and transcription examples as well as information about ongoing work (i.e., additional recording sessions, the creation of phonetic, prosodic, and morpho-syntactit transcription layers.) This webpage also informs about how to obtain a copy of the corpus and tools for searching the corpus.

7. Acknowledgements

The work of Barbara Schuppler was funded by a Hertha-Firnberg grant (T572-N23) from the Austrian Science Fund (FWF). The work of the other authors was partly funded by the European project DIRHA (FP7-ICT-2011-7-288121) and the K-Project ASD, which is funded in the context of COMET Competence Centers for Excellent Technologies by BMVIT, BMWFJ, Styrian Business Promotion Agency (SFG), the Province of Styria - Government of Styria and The Technology Agency of the City of Vienna (ZIT). The programme COMET is conducted by Austrian Research Promotion Agency (FFG).

8. References

Baum, Micha, Erbach, Gregor, and Kubin, Gernot. (2000). SpeechDat-AT: A telephone speech database for Austrian German. In *Proceedings of the LREC Workshop: Very Large Telephone Databases (XLDB)*, pages 51–56.

Boersma, Paul. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):314–345.

Cristoforetti, Luca, Ravanelli, Mirco, Omologo, Maurizio, Sosi, Alessandro, Abad, Alberto, Hagmüller, Martin, and Maragos, Petros. (2014). The DIRHA simulated corpus. In *Accepted for LREC 2014*.

Ernestus, Mirjam. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch. A Corpus-Based Study of the Phonology-Phonetics Interface*. Ph.D. thesis, LOT, Vrije Universiteit Amsterdam, The Netherlands.

Gubian, Michele, Schuppler, Barbara, van Doremalen, J.J.H.C., Sanders, Eric, and Boves, Lou. (2009). Novelty detection as a tool for automatic detection of orthographic transcription errors. In *Proceedings of the 13-th International Conference on Speech and Computer SPECOM-2009*, pages 509–514.

Hänsler, Eberhard and Schmidt, Gerhard. (2004). *Acoustic Echo and Noise Control: A Practical Approach*. Wiley-IEEE Press.

IPDS. (1997). 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).

Mitterer, Holger. (2008). How are words reduced in spontaneous speech? In *Proceedings of ISCA Tutorial and Research Workshop On Experimental Linguistics*, pages 165–168.

Moosmüller, Sylvia. (1998). The process of monophthongization in Austria (reading material and spontaneous speech). In *Papers and Studies in Contrastive Linguistics*, pages 9–25.

Muhr, Rudolf. (2000). *Österreichisches Sprachdiplom Deutsch. Lernzielkataloge*. öbv und hpt, Wien.

Muhr, Rudolf. (2007). *Österreichisches Aussprachewörterbuch – Österreichische Aussprachedatenbank*. Peter Lang Verlag, Frankfurt/M., Wien u.a. 525 S. mit DVD.

Muhr, Rudolf. (2008). 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*, pages 3093–3100.

Pitt, Mark A., Johnson, Keith, Hume, E., Kiesling, Scott, and Raymond, William D. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, 45:89–95.

Schiel, Florian, Draxler, Christoph, Baumann, Angela, Ellbogen, Tania, and Steffen, Alexander. (2012). The production of speech corpora, version 2.5. Technical report, Bavarian Archive for Speech Signals, University of Munich.

Schuppler, Barbara, Ernestus, Mirjam, Scharenborg, Odette, and Boves, Lou. (2008). Preparing a corpus of Dutch spontaneous dialogues for automatic phonetic analysis. In *Proceedings of Interspeech*, pages 1638–1641.

Schuppler, Barbara, Ernestus, Mirjam, Scharenborg, Odette, and Boves, Lou. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39:96–109.

Schuppler, Barbara. (2011). *Automatic Analysis of Acous-*

tic Reduction in Spontaneous Speech. Ph.D. thesis, Radboud University Nijmegen, The Netherlands.

Torreira, Francisco, Adda-Decker, Martine, and Ernestus, Mirjam. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52(3):201–212.

Weilhammer, Karl, Reichel, Uwe, and Schiel, Florian. (2002). Multi-tier annotations in the Verbmobil Corpus. In *Proceedings of LREC*, pages 912–917.

A Instructions for the Manual Creation of Orthographic Transcriptions

For the creation of the orthographic transcriptions, transcribers used the set of symbols shown in Table 5. While creating the transcriptions, the transcribers compiled a lexicon with the words which do not already have an entry in the ADABA lexicon (Lexicon of standard Austrian German (Muhr, 2007)). This lexicon is shared among the transcribers (see column 'Lexicon' in Table 5). The general instructions for the creation of the orthographic transcriptions were:

1. Create a Praat-TextGrid (Boersma, 2001) for each speechfile (.wav) with two different interval tiers for the two different speakers. *Speaker 1* is annotated in the first tier, *Speaker 2* in the second tier.
2. Separate utterances with boundaries; since previous studies have shown that the resulting chunks shall ideally not be longer than 4s, in any case, chunks shall not be longer than 8s.
3. Don't include overlapping speech, non-speech noises, broken words and laughter in longer chunks, but separate them in smaller chunks.
4. Spell spoken words with their standard orthography and not close to the pronunciation. An exception are those words, which do not exist in standard orthography or which were produced in a variant which can not be derived from the citation pronunciation.
5. For low-frequent or not-standard words, check the lexicons; if it does not exist there, enter the word with the orthography you used.
6. Capitals are not used at the beginning of sentences.
7. Separate punctuation marks from the words with a white space. Use "." only at the end of declarative sentences, "!" at the end of commands and exclamations and "?" exclusively at the end of interrogative sentences. Use "," at intonational phrase boundaries, and not at all positions where the rules for German orthography would require a comma. Use "–" only in those words which are officially written with a dash.
8. Annotate every audible speech and non-speech noise in the recordings.

Table 5: **Symbols used for the orthographic transcriptions**: ADABA = Lexicon of Austrian German, ERG = Lexicon with additional German words, DIAL= Lexicon with dialect words, PART = List of small particles, FSP = foreign words, MWEX= multi-word expressions.

Lexical Item	**Example**	**Lexicon**
Standard Austrian German words	*ich gehe von zu Hause weg*	ERG
Dialect words	$< *DIAL >$*Kretzn*	DIAL
High frequent multi-word expressions	*ja geh bitte*	MWEX
Spelling of letters	$\$G$ $\$K$ $\$K$	–
Abbreviations, letters not spoken separately	UNI	ERG
Proper names of people, places, etc.	*Sankt Michael*	ERG
Numbers not written with digits	*#einhundertdreizehn*	ERG
Neologisms, invented by the speaker	*Genussvermeider*	ERG
Foreign words	$< *IT > saluti$	FSP
Hesitations and disfluencies	**Example**	
Repetition: word (group) produced more than once	*und dann hat* \+ *hat* \+ *er* + *und dann* \+ + *und dann* \+ *hat er*	
Slip of the tongue	kervehrt\v	PART
Broken word	gebra*	PART
Other types of speech and non-speech	**Example**	
Imitation of accent or other person	und\i was\i hast\i du\i	
Imitation of an animal, vehicle, etc.	tschu \L tschu \L	PART
Whispering of an utterance	er hat eh \F schon \F wissen \F	
Non-speech produced by the speakers' vocal folds	<laughter>, <singing> <sigh>, <cough>, <smack> <breathingIN>, <breathingOUT>	
Non-speech noise while producing a word	<laughter>und <laughter>dann hat er	
Non-speech other than mentioned above	<noise>	
Overlapping speech of two speakers	\\ja, hm, ja das \\ \\<laughter>\\	
Artifacts in the recordings	<# artefact>	
Other noises not covered with mentioned symbols	<# noise>	