LRE manuscript No.
(will be inserted by the editor)

Alcohol Language Corpus

The first public corpus of alcoholized German speech

Florian Schiel · Christian Heinrich · Sabine Barfüsser

Received: date / Accepted: date

Abstract The Alcohol Language Corpus (ALC) is the first publicly available speech corpus comprising intoxicated and sober speech of 162 female and male German speakers. Recordings are done in the automotive environment to allow for the development of automatic alcohol detection and to ensure a consistent acoustic environment for the alcoholized and the sober recording. The recorded speech covers a variety of contents and speech styles. Breath and blood alcohol concentration measurements are provided for all speakers. A transcription according to SpeechDat/Verbmobil standards and disfluency tagging as well as an automatic phonetic segmentation are part of the corpus. An Emu version of ALC allows easy access to basic speech parameters as well as the us of R for statistical analysis of selected parts of ALC. ALC is available without restriction for scientific or commercial use at the Bavarian Archive for Speech Signals.

Keywords speech corpus · alcohol detection · intoxication · speaker features and forensic phonetics

1 Introduction

It is a widely accepted hypothesis that alcoholic intoxication as other factors such as fatigue, stress and illness influence the way a person speaks. Quite a number of studies during the last decades have investigated this hypothesis from different points of view: looking for reliable acoustic [18,8] or behavioristic [14,3,33,36] features that may indicate intoxication, studying the physiological effects of alcohol on the articulators [37] or even pursuing forensic questions [18,4,17,22] such as in the infamous case of the captain of the Exxon Valdez [15]. Unfortunately, all these studies have in common that the analyzed empirical speech data are not available for other research groups.

To our knowledge up to this point nobody has ever seriously claimed to be able to detect the grade of intoxication from the speech signal by means of automatic methods

F. Schiel, Chr. Heinrich, S. Barfüsser
Bavarian Archive for Speech Signals, Ludwig-Maximilians-Universität München, Schellingstr. 3, 80799 München, Germany
Tel.: +49-89-21802758
Fax: +49-89-21805790
E-mail: schiel|heinrich|bine@bas.uni-muenchen.de

alone. However, if researchers are ever to develop such a method, they will first need a corpus of intoxicated speech produced not only in the lab but also in a possible real life situation.

This article describes a new speech resource at the Bavarian Archive for Speech Signals (BAS)[1] containing speech recordings from sober and intoxicated speakers. The Alcohol Language Corpus (ALC) was recorded over a time period of 30 months (2007-2009) in close cooperation with the Institute of Legal Medicine, Munich, and the German 'Bund gegen Alkohol und Drogen im Strassenverkehr'[2] (BADS). ALC comprises alcoholized and sober speech of 162 male and female German speakers aged between 21 and 64 who were tested by breath and blood samples, recorded outside the laboratory and with a variety of speech styles.

There were three main motivations to produce ALC:

1. Forensic speech sciences:
   Former investigations of alcoholized speech report differing and partly inconsistent findings on how intoxication affects speech. Most of these studies analyzed fewer than 40 speakers, mostly male, under lab conditions and with read speech, single words or vowels (e.g. [18],[14],[8],[3],[17],[33]). Furthermore, in most studies the amount of intoxication was measured by breath alcohol detectors (BRAC = breath alcohol concentration) or estimated from the intake of beverages.
   In ALC 162 female and male speakers have been recorded in real live conditions and all intoxicated speakers were tested with BRAC and – more reliably – by taking blood samples (BAC = blood alcohol concentration). Therefore ALC should provide a statistically sound basis to answer some of the still debated questions (see also Table 7 for the basic numbers of ALC).
2. Phonetic sciences dealing with speaker characteristics / biometrics:
   In the last decade a number of studies have identified phonetic cues and feature sets for speaker profiling. For instance age, gender, dialect, fatigue and other pathological states, but also emotion have been investigated in speech (e.g. [31], [13], [35], [41]). However, the interaction of such speaker characteristics has not been addressed thoroughly. More specifically, how does alcoholic intoxication affect the phonetic cues for other speaker characteristics?
   Since ALC covers both genders and a variety of age groups, it should offer a first empirical basis to investigate some of these unknown relations.
3. Alcohol detection in the automotive environment:
   Alcoholic intoxication (AI) has always been and still is one of the major causes for traffic accidents ([34]). AI can be measured by (ordered by descending reliability): measuring BAC, measuring BRAC and a variety of psychological tests (mainly about reaction time and motor control). All these tests can only be applied either in random checks on drivers or after an accident has already happened. Currently there are no known practical methods to routinely check on the AI of a driver pre-emptively. The fact that an increasing number of functions in the automobile are and will be controlled by the speech of the driver raises the question whether this speech input may be used to detect possible alcoholic intoxication, and thus prevent driving under the influence of alcohol.
   ALC is recorded in the automotive environment and covers speech styles (command

[1] BAS is located at the Ludwig-Maximilians-Universität, München, Germany, www.bas.uni-muenchen.de/Bas

[2] 'Union against alcohol and drugs in traffic' ([6])

& control) typical for car applications. As such ALC can provide the training and test materials necessary to train automatic alcohol detection systems.
Alcohol detection differs from classic pattern recognition tasks when the training or enrollment data matches the test data and the subject is sober when producing both. In the alcohol detection application, the subject is sober when producing the enrollment data and either sober or intoxicated in the test situation. Please refer to [30] for a more detailed discussion of this problem.

Aside from these primary motivations the resulting corpus may be used for other investigations/applications such as:

- automatic speech recognition in the automobile
- human machine dialogue design in the automotive environment
- discourse analysis

The remaining article is structured as follows: In Section 2 and 3 we will give some considerations regarding the corpus design and describe the recorded speech items of ALC followed by Section 4 which describes the recording procedure including all factors that might have an influence on the speech signal and how they have been registered for the corpus. Section 5 gives an overview about the transcription and tagging schema. In section 6 the post-processing of the raw data will be outlined including the automatic segmentation into words and phonemic segments while Section 7 gives a brief description of the resulting Emu database. Section 8 lists speaker and recording statistics as well as information about accessibility before we conclude with a list of some of the ongoing projects based on ALC in section 9.

2 Corpus Design with regard to previous studies

There are some inherent questions to be answered when dealing with speech from intoxicated persons before starting the actual data collection:

1. How to measure the intoxication?
   Most previous studies applied breath alcohol concentration (BRAC) detectors as being used by law enforcement; only a few studies report real blood alcohol measures (e.g. [17]). BRAC values tend to correlate with the blood alcohol level but are not 100% reliable (and are therefore not admissible as evidence before court in most countries). In a pilot study we analyzed the BAC and BRAC test results of 152 intoxicated persons and found a Pearson correlation of 0.89. The BAC varied from 0.00023 to 0.00175[3]; the maximum difference between BRAC and BAC was 0.00076. From the distribution we estimated that the chance for a deviation between BRAC and BAC of more than 0.0001 is about 0.29.
   We therefore decided to apply BAC tests for all experiments in ALC.
2. Which persons are to be tested?
   Reviewing the literature we found that in most cases only the speech of adult male persons or students was analyzed limiting the potential use of such research even if it is true that the majority of felonies under the influence of AI are committed by males. Since the purpose of ALC is not solely forensic but should also address the impact of intoxication on both sexes and different age groups, we decided to collect speech from both genders over a broader range of age.

[3] According to German law (2010) a BAC level of above 0.0005 is regarded as illegal in traffic.

3. How many speakers?
   Most of the published findings were based on the data of less than 40 persons.
   In case that we can measure only one feature value per participant – for instance the long term fundamental frequency – and still want to yield significant results for both genders we need at least 60 participants per gender.[4] Hence the target number of participants in ALC should be 120 or more, equally distributed to both genders.
4. What type of speech should be analyzed?
   Most earlier studies use read speech recorded in the lab (often the well-known story 'The Northwind and the Sun'). Only a few studies looked into semi-spontaneous speech (e.g. [4,14]). Forensic speech and application speech in the automotive environment will be rather dominated by spontaneous speech, commands, place names and digit strings. Therefore a greater variety of speech styles including listings, digit chains, command&control, spontaneous monologue and dialog speech is desirable. Which leads us directly to the next question:
5. How to evoke realistic speech from intoxicated persons?
   Ethical considerations prohibit eavesdropping on the conversation of persons without their consent - even more so if they are intoxicated. Standard lab tests where stimuli are prompted to persons tend to be in a very artificial environment and may therefore influence the behavior of intoxicated persons. Screen prompted speech may be suitable for tongue-twisters, but how to elicit real spontaneous speech? Most studies so far have used screen prompted stimuli or even stimuli read from paper.
   ALC contains real dialogues between two persons, question answering, picture comments and situational prompting ([23]) aside from prompted texts to achieve a more realistic and broader set of speech styles in ALC.
6. Which acoustic environment?
   The acoustic environment should be as realistic as possible, while on the other hand we need some control about the acoustics to ensure we do not simply measure differences in the acoustic environment instead of the recorded speech signal. In the case of ALC we encountered another problem, namely the fact that we had to record at different locations in Germany to elicit speech in different dialects. As a compromise we chose to record in the automotive environment, which can be kept constant across the sober and intoxicated recording as well as across different recording locations. This also had the benefit of yielding field recordings that may be used for different investigations into voice control in the car.

The next two sections will give the details of the recorded content and the recording procedure used in ALC, which more or less directly result from the considerations above.

3 Recorded Speech

ALC contains a variety of speech styles: *read, spontaneous* and *command&control speech* in various forms. Table 1 lists all recording types for the intoxicated case (set A =

[4] The number 60 roughly represents the degrees of freedom where the F statistic gets flattened; that is to say, the F-value does not change very much for degrees of freedom above 60, and therefore testing for significance does not improve much more above that number [20].

Table 1 ALC recording types and their respective numbers in set A and N.

speech type	*item type*	*intoxicated/control (A)*	*sober (N)*
read speech	digit string	5	10
	tongue twister	5	10
	read command	4	9
	address	5	10
	spelling	1	1
spontaneous speech	picture description	2	4
	question answering	1	1
	spontaneous command	5	10
	dialogue	2	5
sum		30	60

'alcoholized') and the sober case (set N = 'non-alcoholized').[5] While designing the read speech part, combinations of sounds were emphasized that have been reported as being affected by alcoholic intoxication (e.g. [19]), such as /s/ in contrast to /ʃ/, /t/ in contrast to /k/, voiceless plosives /p/, /t/, /k/ in contrast to their voiced counterparts /b/, /d/, /g/ as well as the nasals /m/ and /n/.

Digit strings are represented by telephone, credit card and license plate numbers. Tongue twisters were added to the read speech part to verify the hypothesis that intoxicated speakers increase their articulation errors. The selected tongue twisters are of rare types that are not generally known to avoid the case where speakers are able to speak them by heart. Read commands were taken from a real automotive voice control application. Addresses are real addresses selected from a geo database which are either difficult to pronounce (e.g. *'Schwester-Hermenegildis-Strasse'*) or contain interesting sound combinations as pointed out above (e.g. *'Madapaka-Betegindis-Strasse'*). In the spelling recording type, subjects spell the names of German cities.

The picture description, question answering and dialogues have a maximum recording time of 60secs. Speakers are not forced to fill the 60sec time slot to avoid unnatural silence intervals. Each speaker described 6 examples taken from a collection of psychological test pictures. Then she/he answered/discussed the following questions/topics:
'What was the nicest present you ever received?'
'Tell me about your last vacation.'
'What do you think of Christmas?'
'Discuss the previous intoxication experiment.'

Particularly the question answering and the dialogues evoke spontaneous speech that comes fairly close to real-life-situations.

Spontaneous commands are control commands from the same scenario as the read command items formulated by the speaker herself following directions on screen. For details about the Situational Prompting technique see [23].

Items are presented in a fixed randomized order except that all the command&-control type items (1/3 in each set) are grouped together at the end of each session, during which the engine of the car is switched on.

[5] A full listing of all screen prompts can be downloaded from *http://www.bas.uni-muenchen.de/Bas/BasALCPROMPTS*.

4 Recording Procedure

All speakers voluntarily participated in an intoxication test supervised by staff of the Institute of Legal Medicine. These intoxication tests are organized on a regular basis by the BADS. Beside the speech recordings for ALC these intoxication tests are intended to enhance the sensitivity of legal professions, medical personnel and law enforcement officers to the possible influence of alcoholic intoxication.

Each speaker participating in ALC signs a legal form stating that she/he gives her/his consent for the scientific and technical use of the recorded speech, under the condition that the corpus contents may not be associated with personal data.

Before the actual test each speaker chooses the blood alcohol concentration (BAC) she or he wants to reach during the intoxication test. The possible target range is between 0.3 ‰ and 1.5 ‰. To estimate the required amount of alcohol we use the Widmark formula ([40]):

$$c = \frac{V}{mr} \qquad V = cmr \tag{1}$$

where c is the alcohol concentration (in ‰), V is the amount of consumed alcohol (in g), m is the body mass (in kg) and r is the reduction factor, depending on gender, age and body mass.

To estimate r we apply the extended Watson formulas ([38]) for the body water content of females and males

$$g_{male} = 2.447 - 0.09516t + 0.1074h + 0.3362m \tag{2}$$

$$g_{female} = 0.203 - 0.07t + 0.1069h + 0.2466m$$

where t is the age (in years) and h is the body height (in cm), and combine g with the density of blood $\rho_b = 1.055 \frac{g}{cm^3}$ and the fraction of water in blood $f = 0.8$:

$$r = \frac{\rho_b g}{fm} \tag{3}$$

Inserting (3) in (1) yields the necessary amount of alcohol (in g):

$$V = \frac{c\rho_b g}{f} \tag{4}$$

Finally V has to be re-calculated into amounts of beer or wine respectively.

After having consumed the estimated amount of alcohol within the maximum time period of two hours, the speaker has to wait another 20 minutes before undergoing three tests: BAC, BRAC and speech recording.

We use two different BRAC testers of the same technology: Dräger Alcotest 7410, a pretest instrument with fuel cell as measuring principle and an internal conversion from mg/l BRAC to ‰ BAC, and an Envitec Alcotest, similar in construction. The BAC is determined by Head-Space Gaschromatography as used in forensic analytics but without ADH-method averaging over repeat determination.

To avoid any significant changes (saturation, decomposition) of the measured BAC the speaker is asked to perform the ALC speech test immediately after the alcohol tests, which lasts no longer than 15 minutes. After a minimum of two weeks later the speaker is required to undergo a second recording in sober condition, which takes about 30 minutes and includes two times as many prompts as the test in intoxicated condition. A randomly selected group of 10 male and 10 female speakers is recorded for a third

Table 2 Meta data registered of speakers and recordings

speaker data	*values*	*recording data*	*values*
gender	F,M	date&time	2009-03-15.12:45
speaker ID	(integer)	speaker ID	(integer)
dialect	(state of school)	recording car	C1,C2
height	(in cm)	BRAC	(float)
weight	(in kg)	BAC	(float)
smoker	yes,no	weather	sun,rain
drinking habits	light,moderate,heavy	emotional state	f1-f10
profession	(string)	emotional state in test	r1-r4
age	(integer)	-	-

time after another delay of at least one week under the exact same recording condition as the first test but without being intoxicated. This control group provides data to check for unknown factors that may influence the speech signal beside the effects of intoxication.

To factor out other influences, in all tests the speaker is interviewed beforehand about any pathological or psychological events that may affect her/his speech. If any such factors are evident, the test is either postponed or the speaker is not included in ALC at all.

All the recordings take place in one of two standard cars[6], to ensure the same acoustic environment for the different recording locations. The engine is switched off for 2/3 of the recordings and switched on for the application speech to create a realistic ambience for voice control commands. For security reasons no recordings are performed in the moving car. Each test, in intoxicated and sober state, is supervised by the same member of the ALC staff, who at the same time acts as the conversational partner for the dialogues. The recordings are controlled by SpeechRecorder ([10]) running on a laptop where the respective task is prompted on the display. For all text-prompted recordings (read speech), the text prompt is not visible before the speaker hits the record button. To compensate for early recording stops (that is, the speaker hits the stop button while still speaking) SpeechRecorder was configured to delay the recording by another 500msec. Speakers are not allowed to repeat a recording unless there is a technical problem. In cases where there are two or more versions of a recording item, the first recording containing a serious attempt is selected for the corpus.

The speech signal is captured by two microphones: one headset Beyerdynamic Opus 54.16/3 and one AKG Q400 mouse microphone, frequently used for in-car voice input, located in the middle of the front ceiling of the automobile. Both microphones are connected to an MAUDIO MobilePre USB audio interface where the analog signal is converted to digital and transferred to the laptop. The sampling rate is 44,1kHz, 16 bit, PCM.

Aside from the speech signal we collected a number of meta data about speakers and recording conditions to allow statistical cross testing for other factors than the main factor sober/intoxicated. Table 2 summarizes these meta data. Meta data are provided in SpeechDat compatible ([32]) speaker and session tables. A pronunciation dictionary lists the citation form of each word token found in ALC coded in SAM-PA ([39]).

[6] Opel (GM) Astra gasoline (C1), VW Passat diesel (C2).

5 Transcription and Tagging

All recordings are annotated and tagged using the web-based annotation tool Web-Transcribe ([11]) and applying SpeechDat transcription conventions (specified in [32]) extended by a subset of the German Verbmobil (e.g. [5]) conventions as summarized in Table 3.

Table 3 Annotation tags used in ALC: the basic tag set is SpeechDat extended by a subset of German Verbmobil tags.

tag	meaning	example
#	wrong pronunciation or word fragment	mit dem #Tufenkopftopf ...
***	dialectal variant	*hamma gemacht ...
****	incomprehensible part	heut ist schönes ** Wetter
~	technical truncation	in dem Kupferkocht~
[spk] *[int]* *[sta]*	speaker noise temporary background noise stationary background noise	
-/.../-	correctional truncation	er ist -/verschw/- gegangen
+/.../+	repetition or stutter	als ob +/der/+ der Mann einen...
<"ah>, <hm> *<"ahm>, <hes>*	vocalic hesitation, nasalized hesitation mixed hesitation, residual class	
<Z>	word lengthening	und dann<Z> sind wir ...
<P> *<PP>*	short silence interval (< 1 sec) long silence interval (> 1 sec)	
wo_..._rd	interruption in word	Urlaubs_ <hm> _budget

The following additional guidelines were applied in the transcription:

- the orthographic transcription is as close to the spoken material as possible, even in cases of dialectal variation, pronunciation errors or word breaks
- no punctuation marks are used
- spelled words are transcribed with space-separated capital letters
- speech of the dialogue partner as well as cross-talk is not transcribed

Aside from the transcript the annotator counts irregularities which occur within a recording.[7] The irregularity count is supposed to be a gold standard for the detection of disfluencies: if this counter does not show significant differences between intoxicated vs. sober speech, it does not make sense to work on automatic means for the detection of such effects. The term 'irregularities' in our context refers to all phenomena within the speech signal that can be considered not to be part of error-free fluent speech:

- tagged silence interval if it can be considered as a hesitation
- abnormal word lengthening
- filled pause
- wrong pronunciation or word truncation
- correctional truncation
- repetition or stutter

[7] Due to budget constraints this was done only for a subset of ALC we consider to be worth investigating with respect to irregularities: tongue twister, picture description, question answering, dialogue, read control & command (set A: 14 items, set N: 29 items).

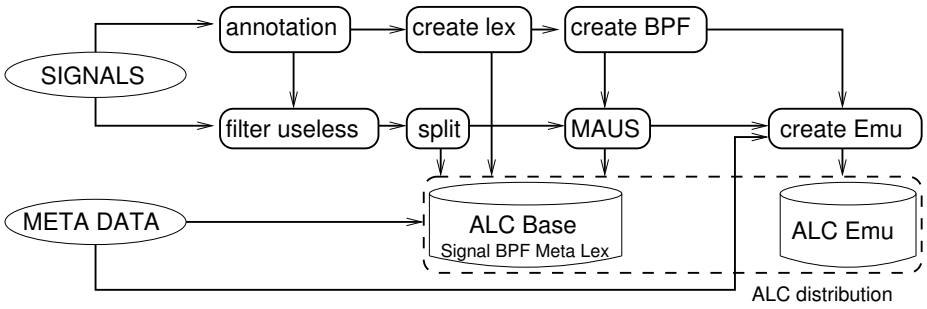

Fig. 1 Post-processing of the ALC corpus data

Where more than one repetition or stutter is observed after another, this group of repetitions or stutter is counted as one irregularity. Correctional truncations including other irregularities are also counted as one irregularity. Hesitations occurring before correctional truncations are dealt with separately and thus result in two counted irregularities; hesitations right after correctional truncations can be attributed to the truncation and in this case only one irregularity is counted.

Additional switches for each recording are set by the annotator for the perceived condition of the subject: *inconspicuous, lightly intoxicated, heavily intoxicated*; in cases where the recording contains no speech it is marked as *useless*.

Finally, in each recording the beginning and end of speech is marked on the time line to improve further automated processing. Thus, pauses that occur at the beginning and the end of a recording are not considered for further analysis nor marked in the transcription.

The described ALC annotation is performed as a one-pass process, that is no second manual verification of the annotation is applied. Unclear cases are marked as such by the individual annotator, and then discussed among annotators in regular meetings. Three different annotators participated in the ALC transcription.

6 Post-processing

Figure 1 depicts the data flow of the post-processing after the completed annotation and tagging.

After a consistency check on sound and annotation files word tokens are harvested from the annotation and cross-checked against the ALC pronunciation dictionary. If an unknown word token[8] is found, a citation form pronunciation coded in SAM-PA ([39]) is inserted into the lexicon either by lexicon lookup from PHONOLEX ([24]) or by applying the text-to-phoneme method BALLOON ([27]).

BAS Partitur Format files (BPF)[9] are created for each recorded sound based on the annotation and tagging described in Section 5. They comprise the tiers *orthography* (ORT), *pronunciation* (KAN, derived from the dictionary) and *recording segmentation* (TRN, derived from the annotation).

[8] including word fragments, dialectal variants and mispronunciations.

[9] For a detailed and up-to-date documentation on the BPF see http://www.bas.uni-muenchen.de/Bas/BasFormatseng.html

Table 4 Meta data labels in ALC Emu top hierarchy.

meta data	*description*	*values*
alc	alcoholized vs. non-alcoholized	a/na
sex	speaker gender	F/M
age	speaker age	21-64
acc	German accent	(state code)
drh	drinking habits	light/moderate/heavy
aak	BAC value	(float)
bak	BRAC value	(float)
ges	emotional state	f1-f10
ces	emotional state recording	r1-r4
wea	weather	SUN/RAIN
irreg	irregularity counts	(see Table 5)
specom	comment regarding speaker	(free text)
utt	utterance ID	*e.g. 0121012023*
o_utt	corresponding utterance ID	*e.g. 0122031059*
spn	speaker ID	*e.g. 012*
item	prompt item ID	*e.g. 023*
o_item	corresponding item ID	*e.g. 059*
type	speech type (Table 6)	R/E/M/D/L
content	content type (Table 6)	A/P/Q/N/R/C/S/T

The KAN and TRN tiers serve as basis for the automatic phonetic segmentation and labeling performed by the Munich AUtomatic Segmentation system (MAUS, [28]). In a validation on German face-to-face dialogue speech ([16]) the MAUS segmentation scored a label accuracy of 93.8% of the inter-labeler agreement, while the segmental boundary accuracy (deviations of $< 20msec$) was about 90.3% of inter-labeler agreement. As with all automatically performed segmentations the MAUS segmentation can serve to localize anchor points such as word boundaries and syllable nuclei, but not for fine-grained phonetic duration analysis such as voice onset time. MAUS segmentations of dialogue recordings should not be used without prior manual checking, since cross-talk may affect the segmentation quality.
No formal validation on the segmentation quality was performed in the ALC project due to budget reasons. However, informal random checks on longer spontaneous speech recordings and dialogue speech in intoxicated speech of ALC showed no deterioration compared to the segmentation of normal speech.

7 ALC Emu Database

To simplify phonetic and linguistic analysis and to save the prospective user the filtering of unwanted versions, an Emu database of ALC is added to the corpus distribution ([7]). In contrast to the base corpus the Emu database contains only one validated recording for each prompt item. See also Table 7 for detailed figures derived of the ALC Emu database.

Emu annotation files are derived from the phonetic MAUS segmentation and stored on the phonetic layer. The segmental information is propagated up to the word layer, which carries an additional label describing the canonical pronunciation form of each

Table 5 Counter of irregularities based on ALC transcripts provided in ALC Emu for each individual recording (see Table 3 for details).

counter	*description*
1	number of irregularities (see section 5)
2	number of hesitations (filled pauses)
3	number of short pauses
4	number of long pauses
5	number of word lengthening
6	number of wrong pronunciations or word fragments
7	number of repetitions or stutter
8	number of correctional truncations
9	number of interruptions within word

Table 6 Speech type and content type classes used in ALC Emu (see also Table 1).

speech type	R	read speech (except lists)
	E	elicited speech (spontaneous commands)
	M	monologue
	D	dialogue
	L	read list
content type	A	address
	P	picture description
	Q	question answering
	N	number
	R	read command
	C	spontaneous command
	S	spelling
	T	tongue twister

word (cano)[10]. The word segments are then integrated into the utterance layer which also contains a complete set of meta data labels as listed in Table 4.

The label *irreg* contains a string of nine counters based on the manual transcript as described in Table 5. The labels *type* and *content* allow a rough classification of the recording into speech type and content classes as depicted in Table 6.

Since an Emu database can be queried across hierarchical layers this mechanism allows very elegant grouping and participation of the whole dataset according to meta data values. Via the R language [25] interface of Emu the same queries can be used to load labels, segmental information as well as derived feature signals (e.g. fundamental frequency, energy, formats) into R for further analysis.

8 ALC in Numbers

This section summarizes the most prominent figures of ALC and provides some statistics based on the actual corpus that might be useful for the prospective user.

8.1 Detailed Numbers

In its present state the ALC corpus covers the alcoholized and non-alcoholized speech of 77 female and 85 male speakers. 86% of all speakers were born in southern German

[10] All phonetic symbols coded in SAM-PA [39].

states and also attended the first 4 years of school there; 96% have an university degree; 22% are smokers.

The ALC distribution totals in 30Gbyte and is distributed on DVD-R via the ELRA or BAS.[11]

Table 7 lists the absolute numbers and percentages regarding speakers, recordings, duration, word tokens, lexicon size, phone tokens and tagging of the three data groups *alcoholized, non-alcoholized* and *control*. Please note that the numbers for the sub-groups *read, spontaneous* and *command & control* do not add up to 100% since the latter consists of both read and spontaneous speech.

The percentages given in Table 7 are either with regard to a total (marked as 100%) or with regard to the total number of word tokens in the respective sub-group. For instance 2.27% of all word tokens recorded in the alcoholized condition are hesitations. In case of the tag *irregularity* this base number differs from those of the remaining tags because the tag *irregularity* was applied only to a subset of recordings.

Tags showing a highly significant difference ($p < 0.0001$) between alcoholized and sober condition are marked with '*'.

8.2 Some Interesting Cases

The number of *word tokens* in the sub-group *spontaneous speech* reveals that in average speakers utter more words per recording item in sober condition than being intoxicated (set N: 50.84 vs. set A: 46.74, $t = 3.9, p = 0.0001688$ [12]) which correlates with findings about a higher speech rate of sober speakers reported in a recent study ([30]). This was rather unexpected since the common stereotype is that speakers speak more under the influence of alcohol. One possible explanation for both effects might be the experimental setting seen as a 'test situation', where speakers try to articulate as clearly as possible to camouflage their intoxication.

The reverse effect can be observed in the sub-category *command & control*, in which the speakers were asked to read or formulate commands to the automobile (set N: 5.00 vs. set A: 6.03 words per item). Again this is not what we expected to see, since the higher mental load required to think of new commands was expected to be diminished under the influence of alcohol and hence the number of words to be less than in the sober state.

Table 8 illustrates some selected numbers across genders and recording groups.[13] At first glance there seem to be some interesting differences in gender behavior:

- *number of word tokens in spontaneous speech:* while female speakers utter the same number of words in sober and intoxicated states, their male colleagues produce fewer words under the influence of alcohol.
- male speakers exhibit a higher proportion of *irregularities* per word token in both groups (set A: $\chi^2 = 13.5, p = 0.0002369$, set N: $\chi^2 = 12.3, p = 0.0004531$)
- the same is observed for *hesitations* (set A: $\chi^2 = 37.0, p < 0.0001$, set N: $\chi^2 = 95.3, p < 0.0001$)

[11] See the BAS catalog at http://www.bas.uni-muenchen.de/Bas/BasALCeng.html for details; BAS and ELRA distribution fees apply.

[12] Paired t-test based on the averaged words per recording item per speaker

[13] The percentages in Table 8 are given with regard to the number of word tokens in the respective sub-corpus; therefore the values are comparable across groups and genders.

Table 7 ALC corpus numbers: percentage of tags is in relation to the number of word tokens; tags with significant different proportions are marked with *.

group		alcoholized		non-alc.		control	
speakers	total	162	100%	162	100%	20	100%
	female	77	47.5%	77	47.5%	10	50.0%
	male	85	52.5%	85	52.5%	10	50.0%
	age 21-27	88	54.3%	89	54.9%	13	65.0%
	age 28-35	44	27.2%	44	27.2%	5	25.0%
	age 36-50	16	9.9%	16	9.9%	0	0.0%
	age 51-67	13	8.0%	13	8.0%	2	10.0%
	smoker	36	22.2%	36	22.2%	4	20.0%
	drink. habits light	73	45.1%	73	45.1%	7	35.0%
	drink. habits mod.	80	49.4%	80	49.4%	10	50.0%
	drink. habits heavy	9	5.6%	9	5.6%	3	15.0%
records	total	4860	100%	9720	100%	600	100%
	read speech	3240	66.7%	6480	66.7%	400	66.7%
	spontaneous speech	1620	33.3%	3240	33.3%	200	33.3%
	command & control	1458	30.0%	2916	30.0%	180	30.0%
duration	total	758min	100%	1499min	100%	98min	100%
	read speech	235min	30.9%	418min	27.9%	25min	25.8%
	spontaneous speech	524min	69.1%	1081min	72.1%	73min	74.2%
	command & control	70min	9.2%	125min	8.4%	7min	7.2%
words	total	103831	100%	218430	100%	14403	100%
	read speech	28105	27.1%	53695	24.6%	3304	22.9%
	spontaneous speech	75726	72.9%	164735	75.4%	11099	77.1%
	command & control	8796	8.5%	14566	6.7%	871	6.0%
lexicon	total	7634	100%	11732	100%	2048	100%
	read speech	1146	15.0%	1285	11.0%	201	9.8%
	spontaneous speech	6822	89.4%	10838	92.4%	1910	93.3%
	command & control	879	11.5%	886	7.6%	151	7.4%
phones	total	452698	100%	943086	100%	60769	100%
	read speech	148825	32.9%	290180	30.8%	17573	28.9%
	spontaneous speech	303873	67.1%	652906	69.2%	43196	71.1%
	command & control	49979	11.0%	91539	9.7%	5389	8.9%
tags	irregularity *	4809	5.93%	9441	5.25%	610	5.14%
	hesitation	2362	2.27%	4780	2.19%	341	2.37%
	short pause	4268	4.11%	8915	4.08%	401	2.78%
	long pause *	1956	1.88%	2868	1.31%	203	1.41%
	word lengthening *	511	0.49%	714	0.33%	32	0.22%
	pronunciation err. *	1791	1.72%	2746	1.26%	136	0.94%
	repetition / stutter	1254	1.21%	2575	1.18%	168	1.17%
	correction *	456	0.44%	1330	0.61%	96	0.67%
	word interruption *	173	0.17%	216	0.10%	3	0.02%

Table 8 Selected ALC corpus numbers normalized across genders and groups.

group		alcoholized		non-alc.	
gender		F	M	F	M
words per item	spontaneous speech	48.26	45.37	49.66	51.92
tags per word	irregularity	5.62%	6.23%	5.05%	5.42%
	hesitation	1.98%	2.55%	1.86%	2.48%
	short pause	4.10%	4.12%	4.12%	4.05%
	long pause	1.88%	1.89%	1.40%	1.23%
	word lengthening	0.56%	0.43%	0.35%	0.30%
	pronunciation error	1.74%	1.71%	1.36%	1.17%
	repetition / stutter	1.24%	1.17%	1.30%	1.07%
	correction	0.43%	0.45%	0.57%	0.65%
	word interruption	0.16%	0.17%	0.11%	0.09%

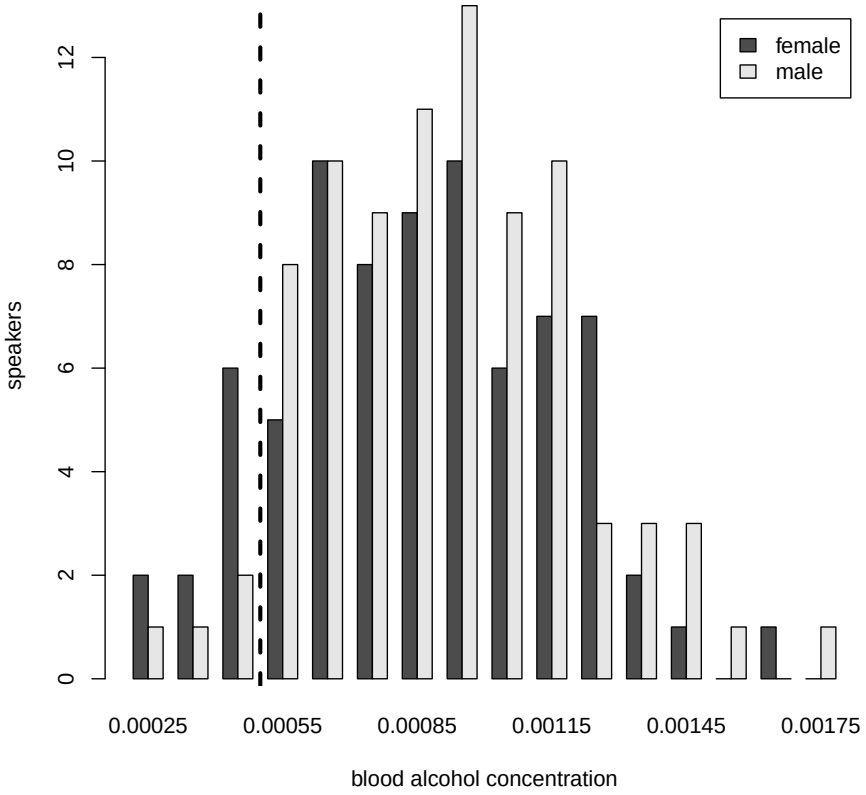

Fig. 2 Histogram of measured blood alcohol concentrations for both genders. The dashed line marks the legal boundary for intoxication in Germany.

- male speakers produce fewer silence intervals >1sec (*long pause*) than female speakers but only in the non-intoxicated state ($\chi^2 = 11.6, p = 0.0006454$)
- on the other hand female speakers exhibit more *pronunciation errors* than their male colleagues only in the non-intoxicated state ($\chi^2 = 16.5, p < 0.0001$)
- finally, male speakers seem to correct themselves more often than female speakers again only when being sober ($\chi^2 = 5.8, p = 0.01575$)

Whether these preliminary findings can be exploited in any way to detect intoxication from the speech input automatically remains to be clarified.

8.3 Alcohol Concentration

The *measured alcoholization BAC* in the intoxicated recordings ranges from 0.00023 to 0.00175; the histograms in Figure 2 depicts the distribution of measurements for both

genders. Since both distributions appear to be uni-modal, measured speech features (speaker independent) may be tested for correlation against the BAC values.

9 Conclusion and Ongoing Projects

A new corpus of speech recordings under the influence of alcohol has been presented. The corpus is available to everyone who is interested to repeat published findings about alcoholized speech or conduct new investigations.

Aside from already distributed copies to other researchers the ALC corpus as described in the previous sections is being used for a number of ongoing phonetic studies at the BAS. Investigated features to separate intoxicated from sober speech are *longterm fundamental frequency* (F0), *F0 in lexically accented, tense vowels*, *F0 trajectories*, several *rhythm parameters* based on the *CVCV...* speech pattern, *long term formant values* and *formant trajectories* and a variety of *disfluencies in spontaneous and read speech*. Selected samples from ALC are being used in perception experiments to yield a baseline of what humans achieve in a simple discrimination task and to verify the common stereotype that intoxication correlates with audible speech markers.

Detailed results from these studies are being published elsewhere; here we give just a coarse summary of some preliminary results.

Previous studies of longterm F0 in intoxicated speech were inconsistent: some authors reported falling, some rising F0; some suggested a non-linear behavior: first falling then rising with increasing BAC.
A study based on 46 speakers of the ALC corpus reveals that F0 as well as F0 range rises significantly in the intoxicated case, although some speakers show opposing behavior. We found no indication for a non-linear relationship of F0 and BAC. F0 in tense vowels did not discriminate better than the overall F0, but significant differences were confirmed for the vowels /a:/, /E:/, /e:/ and /i:/. Female speakers tend to shift their long term F0 in a more consistent way than male speakers ([9]).

To our knowledge rhythm features, except for speech rate, have not been investigated in intoxicated speech so far. Speech rate has been reported to rise as well as to fall in earlier studies on male speakers.
Rhythm analysis based on the automatic phonetic segmentation showed significant differences in the *standard deviation of the duration of vowels clusters* (δV.sd, [26]), the *average durational difference of consecutive vowel clusters* (nPVI-V, [12]) and the *short pause rate per syllable* based on 82 speakers of the ALC corpus ([29]). A newer study based on the energy patterns (*RMS rhythmicity*) of 128 speakers ([30]) indicates that speech rate is in fact decreasing significantly with intoxication.

Tagged disfluencies were investigated on a subset of 93 speaker of the ALC corpus: *Filled pauses* tend to have a significant longer duration in intoxicated speech, at least in spontaneous speech; for read speech the results are inconclusive across genders. Furthermore the *rate of filled pauses*, the *rate of pauses longer than 1sec*, the *number of wrong pronunciations* and the *occurrence of stutter* show significant differences ([1]). This is consistent with earlier studies where the rate of filled pauses, silence intervals and pronunciation errors were reported to rise with intoxication.

Acknowledgements ALC was made possible by funding of the Bavarian Archive for Speech Signals, the Institute for Legal Medicine (Prof. Thomas Gilg) and the 'Bund gegen Alkohol und Drogen im Strassenverkehr'. The recording software SpeechRecorder was supplied by Klaus

Jänsch and Christoph Draxler. The authors would like to thank all colleagues of the Institute of Phonetics and Speech Processing at the Ludwig-Maximilians-Universität (Prof. J. Harrington) for their valuable support as well as the students and technicians who were directly involved in this endeavor, namely Veronika Neumeyer, Indra Dhillon and Christian Gruttauer.

References

1. Barfüsser S, Schiel F (2010): Disfluencies in alcoholized speech. IAFPA Annual Conference 2010, Trier, Germany.
2. Baumeister B, Schiel F (2010): On the Effect of Alcoholisation on Fundamental Frequency. IAFPA Annual Conference 2010, Trier, Germany.
3. Behne D M, Rivera S M, Pisoni D B (1991): Effects of Alcohol on Speech: Durations of Isolated Words, Sentences and Passages. In: Research on Speech Perception, No 17, pp. 285-301.
4. Braun A (1991): Speaking while intoxicated: Phonetic and forensic aspects. Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence, pp. 146-149.
5. Burger S, Weilhammer K, Schiel F, Tillmann H G (2000): Verbmobil Data Collection and Annotation. In: Wahlster W (Ed.): Verbmobil: Foundations of Speech-to-Speech Translation:537-549. Springer.
6. Bund gegen Alkohol und Drogen im Strassenverkehr. URL www.bads.de/Alkohol/statistik.htm. Cited 2009-03-23.
7. Cassidy St, Harrington J (2001): Multi-level annotation in the EMU speech database management system. Speech Communication 33(1-2), pp. 61-77.
8. Cooney O M, McGuigan K, Murphy P, Conroy R (1998): Acoustic analysis of the effects of alcohol on the human voice. In: The Journal of the Acoustical Society of America, p. 2895.
9. Dhillon I (2009): Variation der Grundfrequenz in alkoholisierter Sprache. Thesis magister artium, Ludwig-Maximilians-Universität München.
10. Draxler Chr, Jänsch K (2004): SpeechRecorder – a Universal Platform Independent Multi-Channel Audio Recording Software. In: Proc. of the LREC. Lisbon, Portugal.
11. Draxler Chr (2005): WebTranscribe - An Extensible Web-Based Speech Annotation Framework. In: Proc. TSD 2005, pp. 61-68.
12. Grabe E, Low E L (2004): Durational Variability in Speech and the Rhythm Class Hypothesis. In: Gussenhoven C, Warner N (eds): Papers in Laboratory Phonology 7, Berlin, New York, Mouton de Gruyter.
13. Hansen J H L, Patil S (2007): Speech under stress: analysis, modeling and recognition. In: Müller C (Ed.), Speaker Classification I, LNAI 4343, pp. 108-137.
14. Hollien H, De Jong G, Martin C A, Schwartz R, Liljegren K (2001): Effects of ethanol intoxication on speech suprasegmentals. In: The Journal of the Acoustical Society of America, pp. 3198-3206.
15. Johnson K, Pisoni D B, Bernacki R H (1990): Do voice Recordings Reveal whether a Person is Intoxicated? A Case Study. In: Phonetica, vol. 41, pp. 215-237.
16. Kipp A, Wesenick B, Schiel F (1997): Pronunciation Modeling Applied to Automatic Segmentation of Spontaneous Speech. In: Proceedings of the EUROSPEECH, pp. 1023-1026.
17. Klingholz F, Penning R, Liebhardt E (1988): Recognition of low-level alcohol intoxication from speech signal. In: Journal of the Acoustical Society of America, vol. 84, 1988, pp. 929-935.
18. Künzel H J, Braun A (2003): The effect of Alcohol on Speech Prosody. In: Proc. of the ICPhS. Barcelona, pp. 2645-2648.
19. Künzel H J, Braun A, Eysholdt U (1992): Einfluß von Alkohol auf Sprache und Stimme. Kriminalistik Verlag Heidelberg.
20. Leisch F (2009): personal communication with Friedrich Leisch, Computational Statistics, Ludwig-Maximilians-Universität München.
21. Levit M, Huber R, Batliner A, Nöth E (2001): Use of prosodic speech characteristics for automated detection of alcohol intoxication. In: Bacchiani M, Hirschberg, J, Litman D, Ostendorf M (Eds.): Proc. of the Workshop on Prosody and Speech Recognition 2001, Red Bank, NJ, pp. 103-106.
22. Martin C S, Yuchtman M (1986): Using speech as an Index of Alcohol-Intoxication. Research on Speech Perception, No. 12, pp. 413-426.

23. Mögele H, Kaiser M, Schiel F (2006): SmartWeb UMTS Speech Data Collection: The SmartWeb Handheld Corpus. In: Proc. of the LREC 2006, Genova, Italy, pp. 2106-2111.
24. Large German Pronunciation Dictionary PHONOLEX. URL www.bas.uni-muenchen.de/Bas/BasPHONOLEXeng.html. Cited 2012-12-01.
25. R Development Core Team (2005): R, a language and environment for statistical computing. Reference index version 2.9.0. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org.
26. Ramus F, Nespor M, Mehler J (1999): Correlates of linguistic rhythm in the speech signal. Cognition, Volume 73, Number 3, pp. 265-292, Elsevier.
27. Reichel U D, Schiel F (2005): Using Morphology and Phoneme History to improve Grapheme-to-Phoneme Conversion. In Proc. of the EUROSPEECH, pp. 1937-1940.
28. Schiel F (1999): Automatic Phonetic Transcription of Non-Prompted Speech. In: Proc. of the ICPhS. San Francisco, August 1999, pp. 607-610.
29. Schiel F, Heinrich Chr (2009): Laying the Foundation for In-car Alcohol Detection by Speech. In: Proceedings of the INTERSPEECH 2009, Brighton, UK, pp. 983-986.
30. Schiel F, Heinrich Chr, Neumeyer V (2010): Rhythm and Formant Features for Automatic Alcohol detection. In: Proceedings of the INTERSPEECH 2010, Chiba, Japan, pp. 458-461.
31. Schötz S (2007): Acoustic Analysis of Adult Speaker Age. In: C. Müller (Ed.): Speaker Classification I, LNAI 434, Springer, pp. 88-107, 2007.
32. SpeechDat Deliverable SD1.3.2 : Specification of orthographic transcription and lexicon conventions. URL www.speechdat.org/speechdat/deliverables/public/SD132V24.PDF cited 2010-12-01.
33. Sobell L C, Sobell M B, Coleman R F (1982): Alcohol-Inducted Disfluency in Non-alcoholics. In: Folia Phoniatrica, No. 34, pp. 316-323.
34. Statistisches Bundesamt, Wiesbaden, Germany (2007): Alkoholunfälle 2007. E.g. URL www.bads.de/Statistikdaten/Alkohol/AlkVU%202007.pdf. Cited 2010-12-01.
35. Traunmüller H (1997): Perception of speaker sex, age, and vocal effort. Phonum 4: 183 - 186.
36. Trojan F, Kryspin-Exner K (1968): The Decay of Articulation under the Influence of Alcohol and Paraldehyde. Folia Phoniatrica, No. 20, pp. 217-238.
37. Watanabe H, Shin T, Matsuo H, Okuno F, Tsuji T, Matsuoka M, Fukaura J, Matsunaga H (1994): Studies on Vocal Fold Injection and Changes in Pitch Associated with Alcohol Intake. Journal of Voice, pp. 340-346.
38. Watson P E, Watson R, Batt R D (1980): Total body water volumes for adult males and females estimated from simple anthropometric measurements. The American Journal of Clinical Nutrition 33: Jan 1980, pp. 27-39.
39. Wells J C (1997): SAMPA computer readable phonetic alphabet. In: Gibbon D, Moore R, Winski R (eds.): Handbook of Standards and Resources for Spoken Language Systems. Berlin and New York: Mouton de Gruyter. Part IV, section B.
40. Widmark E M P (1932): Die theoretischen Grundlagen und die praktische Verwendbarkeit der gerichtlich-medizinischen Alkoholbestimmung. Verlag Urban und Schwarzenberg, Berlin Wien.
41. Ke Wu (1991): Gender recognition from speech. Part I: Coarse analysis. The Journal of the Acoustical Society of America, Oct 1991, Volume 90, Issue 4, pp. 1828-1840.