

ALC — Alcohol Language Corpus

Florian Schiel¹; Christian Heinrich¹; Sabine Barfüsser¹; Thomas Gilg²

Ludwig-Maximilians-Universität, München

¹Bavarian Archive for Speech Signals, ²Institute of Legal Medicine

¹Schellingstr. 3, 80799 München, Germany; ²Nussbaumstr. 26, 80336 München, Germany
schiel@bas-services.de; heinrich|bine@bas.uni-muenchen.de; Thomas.Gilg@med.uni-muenchen.de

Abstract

A number of forensic studies published during the last 50 years report that intoxication with alcohol influences speech in a way that is made manifest in certain features of the speech signal. However, most of these studies are based on data that are not publicly available nor of statistically sufficient size. Furthermore, in spite of the positive reports nobody ever successfully implemented a method to detect alcoholic intoxication from the speech signal. The Alcohol Language Corpus (ALC) aims to answer these open questions by providing a publicly available large and statistically sound corpus of intoxicated and sober speech. This paper gives a detailed description of the corpus features and methodology. Also, we will present some preliminary results on a series of verifications about reported potential features that are claimed to reliably indicate alcoholic intoxication.

1. Introduction

It is a widely accepted hypothesis that alcoholic intoxication as other factors such as fatigue, stress, illness influence the way a person speaks. Most people claim that they are capable of judging the state of intoxication of a known or unknown person by her speech even if they have never experienced the 'normal' speech of the same person. If this hypothesis is true, then

- sub-hypothesis 1 : perception tests should reveal a nearly 100% detection of intoxication from the speech signal alone
- sub-hypothesis 2 : it should be possible to detect the relevant features signaling intoxication from the speech signal automatically using methods of pattern recognition

Quite a number of studies during the last decades have tried to investigate that hypothesis from different points of view: looking for reliable acoustic (Künzel & Braun, 2003; Cooney et al., 1998) or behavioristic (Hollien et al., 2001; Behne et al., 1991; Sobell et al., 1982; Trojan & Kryspin-Exne, 1968) features that may indicate intoxication, studying the physiological effects of alcohol on the articulators (Watanabe et al., 1994) or even pursuing forensic questions (Künzel & Braun, 2003; Braun, 1991; Klingholz et al., 1988; Martin & Yuchtman, 1986) such as in the infamous case of the captain of the Exxon Valdez (Johnson et al., 1990).

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. However, if we assume that such a method exists, it is vital for future work that a corpus of intoxicated speech that not only covers lab speech but speech from a possible real life situation as well is freely available to the scientific community. There are some inherent problems to existing studies about speech from intoxicated persons that seem to apply in many cases:

1. How to measure the intoxication?

Most studies applied breath alcohol concentration

(BRAC) detectors as being used by law enforcement; only one study reported real blood alcohol measures (Klingholz et al., 1988). BRAC values tend to correlate with the blood alcohol level but are not 100% reliable (and is therefore not admissible as evidence before court in most countries). Whenever possible blood samples should be taken from test persons.

2. Which persons are to be investigated?

Reviewing the literature we found that in most cases only the speech of adult male persons was analyzed. Is the implication that only male adults are capable of being intoxicated? Gender studies are required here.

3. What type of speech should be investigated?

It is quite surprising that many studies simply used read speech (often the famous 'The Northwind and the Sun') as target speech. It is quite obvious that the main hypothesis stated above was meant for real-life-situation, command&control or even spontaneous speech. Which leads us to the next technical question:

4. How to evoke realistic speech from intoxicated persons?

Ethical requirements 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 different 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 read from paper.

5. Most of the published findings were based on the data of very few test persons.

Statistical surveys based on 3-15 test persons (in most cases only males) are probably not significant. Therefore it is maybe not surprising that a number of publications contradict each other in their findings, especially in the case of statistical surveys of standard

prosodic features.

6. No data of intoxicated speech has ever been made publicly available for other researchers to verify the published findings and/or perform other analysis on the same set of data.
7. No speech data have ever been analyzed from real life situations, such as a driver who tries to start his car under the influence of alcohol.

In this paper we announce and describe the Alcohol Language Corpus (ALC) project at the Bavarian Archive of Speech Signals (BAS) located at the Ludwig-Maximilians-Universität in Munich, Germany. ALC is a collaboration between the BAS, the Institute of Legal Medicine and the company BAS Services Schiel based in Munich. Its aims are two-fold:

- ALC will create a publicly available speech corpus of persons speaking with and without alcoholic intoxication. The corpus will cover a wide range of speech types as well as a large number of test persons across gender and age. The ALC corpus will be distributed by the BAS and ELDA.
- ALC will try to falsify the above stated hypothesis that alcoholic intoxication can be detected perceptively and/or by means of automatic pattern recognition.

In the following we will give a detailed description of the recording methodology and the recorded material, the recruiting technique and measurement of alcoholic intoxication as well as the recording technique within an automobile.

2. Recording Methodology

Test speakers undergo a systematic intoxication test supervised by staff of the Institute of Legal Medicine. These intoxication tests are organized on a regular basis by the 'Bund gegen Alkohol und Drogen im Strassenverkehr' (B.A.D.S.), a league fighting drugs and alcohol abuse in traffic¹. Beside the ALC recordings these intoxication tests are motivated to enhance the sensibility of legal professions and law enforcement about the influence of alcohol intoxication.

Beforehand each speaker chooses the blood alcohol concentration (BAC) she wants to reach during the intoxication test. The possible range is between 0.5 ‰ and 2.5 ‰. Using both Watson- and Widmark formula the amount of required alcohol for each person is estimated and converted into the corresponding amount of beer or wine the subject has to drink in order to reach her individual chosen BAC. After having consumed the estimated amount of alcohol within the maximum time period of one hour, the speaker has to wait another 20 minutes before undergoing two alcohol tests, a breath alcohol concentration (BRAC) test and a blood sample test.

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 in ‰ is determined by Head-Space Gaschromatography as used in forensic analytics but without ADH-method averaging over repeat determination.

Immediately after the tests, the speaker is asked to perform the ALC speech test which will last no longer than 15 minutes to avoid any significant changes (saturation, decomposition) of the measured blood alcohol level. At least 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. To factor out other influences, in both tests the speaker will be interviewed about any pathological or psychological events that may affect her speech. If any such factors are evident, the test is either postponed or the speaker is not admitted to ALC at all. The recordings take place in an automobile, to ensure the same acoustic environment for the different recording locations. The engine is switched off except for the application speech where the running engine creates a realistic ambience for control commands. The test is supervised by a member of the ALC staff, who at the same time acts as the conversational partner for the dialogues. The recordings are controlled with a laptop by the speaker herself where the respective task is prompted on the display. We run two recording automobiles in parallel to cover as many recordings as possible on each intoxication test; care is being taken that every speaker is recorded in the same automobile and with the same dialogue partner in both tests.

3. Technical Setup

The speech signal is recorded with two different microphones: one headset Beyerdynamic Opus 54.16/3 and one AKG Q400 mouse microphone, Austria, 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 recording platform is SpeechRecorder (Draxler & Jansch, 2004), the sampling rate 44,1kHz, 16 bit, PCM. As stated above, a part of the recording is performed while the engine is running. For security reasons there will be no recordings in the moving car.

4. Content

To give consideration to all questions listed in the introduction and to test the feasibility of realistic applications it is necessary to yield recordings of different types of speech, such as read speech, command speech, spontaneous monologues and dialogues.

For the recordings in *intoxicated condition* the read speech part consists of five numbers (telephone, credit card), five addresses, two tongue twisters and three selected sentences of speech training (which for simplification are both referred to as tongue twisters in the following) and five control commands as being used in automobiles. While designing the read speech part there was paid great attention to the combination of sounds that have been reported as being prone to alcoholic intoxication

¹see www.bads.de for more information about B.A.D.S.

(e.g. (Künzel/Braun/Eysholdt, 1992)), such as the alveolar voiceless fricative alternating with the post-alveolar voiceless fricative, the alveolar voiceless plosive alternating with the velar voiceless plosive as well as all voiceless plosives alternating with their voiced counterparts.

Spontaneous speech is covered by five control commands that must be formulated by the speaker herself following directions on screen (situational prompting (Mögele/Kaiser/Schiel, 2006)), three monologues and two dialogues with the recording supervisor, which are initiated by pictures and questions; the length of the monologues and dialogues is restricted to 60 sec each. Particularly the monologues and dialogues evoke rather spontaneous speech that comes fairly close to real-life-situations. This totals in 30 recording items for the test in intoxicated condition which will at the most last for 15 minutes to avoid significant changes as stated earlier.

For the recordings in sober condition there are two times as many items of each category as for the test in intoxicated condition which makes a total of 60 items and a maximum recording duration of 30 minutes.

5. Annotation

The recordings are annotated with WebTranscribe (Draxler & Jänsch, 2004) using SpeechDat (Speechdat, 1997) conventions:

- spellings are marked with capital letters and blanks in between.
- no punctuation marks
- wrong pronunciation or word fragments are marked with a '#' in front of the transcribed intended word²
- dialectal variants are marked with a '*' in front of the corresponding term transcribed in Standard German
- incomprehensible parts are marked with a '**'
- three noise markers: [spk] for speaker noise, [int] for temporary background noise and [sta] for stationary background noise.
- technical truncations of words are marked with '~' at the position where the word is truncated

Speech of the dialogue partner is not transcribed as well as cross talk overlapping the speech of the speaker.

To cover the ALC requirements a number of Verbmobil (Verbmobil, 1997) tags have been added:

- four classes of hesitations: vocal <"ah>, nasal <hm>, mixed <"ahm>, residual class <hes>
- repetitions or stutter are marked with +/repetition/+, e.g. '... als ob +/der/+ der Mann einen...'
- correctional truncations are marked with -/unfinished material/-, e.g. '... haben wir -/fünf/- sechs Tage gebraucht ...'

²not SpeechDat convention

- word elongations are marked with <Z>, e.g. '... und dann<Z> sind wir ...'
- pauses shorter than one second are marked with <P>, longer pauses with <PP>; silence at the beginning or end of the recording is not marked
- interrupted words are marked with ' _ ' at the beginning and end of the spoken word fragments, e.g. 'Urlaubs_<hm> _budget'

Additional switches for each recording are set by the annotator for the perceived condition of the subject: inconspicuous, lightly intoxicated, heavily intoxicated, and whether the recording is useless in case there is no speech.

Finally, in each recording the longest vocal part matching an /a:/, /e:/ or /E:/ steady-state sound is marked with segmental boundaries for automated analysis.

Annotation is a one-pass process, that is no second manual verification of the annotation is performed. Unclear cases may be marked as such by the individual annotator, which are then discussed among annotators in regular meetings.

6. Speaker Recruitment

6.1. Speaker Types

Speakers are recruited from several different areas: graduated law students and other lawyers, judges, police enforcement and public prosecutors. Speakers must be at least 22 years old and agree to participate on their own will. They must further sign a written statement that their recorded speech may be used for scientific investigations as well as for product development granted that the speech recording is absolutely anonymous (even for the distributor of the speech corpus). For a successful performed test the speaker will receive a small incentive after the second recording.

6.2. Number and Distribution of Speakers

To satisfy statistical requirements of the analysis we aim at a total number of 200-250 speakers. At the time of writing a total number of 28 speakers in intoxicated condition and 14 speakers in sober condition have been recorded successfully. We estimate another set of 50-60 recorded speakers until the conference date.

The distribution will be balanced for both genders. We aim at an equal distribution across the four age groups: 22-27, 28-35, 36-50, > 50.

Speakers are recruited and recorded in at least four different locations in southern Germany, namely Munich, Landshut, Augsburg and Traunstein. Unfortunately, locations cannot be changed arbitrarily because the intoxication tests have to be organized by B.A.D.S. at locations where speakers are willing to participate and thus are predestined for the ALC recordings.

7. Meta Data

To achieve a high level of anonymity only the following meta data are associated with each speaker ID: gender, age, BRAC and BAC value, weight, height, region in which the speaker attended elementary school, profession, smoker/non-smoker. Additionally, speaker's drinking habits are classified as *light*, *moderate* or *heavy* drinkers.

To assign a speaker to one of the three classes the speaker is interviewed about the quantity and frequency of consuming alcohol. To keep this as simple as possible, there are two categories for either quantity (*little/much*) and frequency (*infrequently/frequently*). In table 1 you can see all combinations of the categories and their resulting drinking habits.

quantity \ frequency	infrequently	frequently
little	light	moderate
much	moderate	heavy

Table 1: drinking habits in ALC

The following data are associated with each session ID: date and time of recording, speaker ID, age, gender, the region in which the speaker attended elementary school, the automobile in which the recording took place³, BRAC, BAC, general emotional state and current emotional state.

8. Preliminary Results

At the time of writing only 14 speakers (5 female / 9 male) have been recorded in both intoxicated and sober condition. Therefore, the following reported findings should be regarded as preliminary and not (yet) statistically reliable. For this paper we concentrated on two issues, namely the number of classical slips of the tongue and, more specifically, incomplete articulations in the sense of (Künzel/Braun/Eysholdt, 1992).

Regarding the slips of the tongue the error rate was manually determined for each examined recording namely the five tongue twisters and four addresses being considered relevant within the recorded content. The errors were classified in four different groups, *omission*, *insertion*, *substitution* and *repetition*. The error rates reported here are the overall error rates including all error classes. Investigating the five tongue twisters revealed a total error count of 73 (5.21 errors per person) in intoxicated condition and a total of 27 errors (1.93 per person) in sober condition. In the four read addresses we counted 21 errors (1.5 errors per person) in intoxicated and 21 errors (1.5 errors per person) in sober condition likewise.

BAC	sober	intoxicated
0.34 ‰	2 + 2	10 + 5
1.30 ‰	2 + 1	4 + 0

Table 2: total slips of the tongue for *tongue twisters* + *addresses* of the two speakers with extreme BAC values in sober and intoxicated condition

Contrary to expectations the speaker with the lowest measured BAC (0.34 ‰) exhibited more errors in intoxicated condition while the speaker with the highest measured BAC (1.3 ‰) exhibits less errors in intoxicated condition than in

sober condition (see table 2). No correlation between a rising error rate and the BAC across speakers could be found. Since the number of spoken words may differ between the two conditions (e.g. by repetition of words) tables 3 and 4 show the word normalized error rates for the tongue twisters and the addresses of the different groups *omission*, *insertion*, *substitution* and *repetition* for each gender and condition.

error		gender			
		o	i	s	r
f	intoxicated	0.0305	0.0038	0.0611	0.0496
	sober	0.0078	0.0039	0.0194	0.0233
m	intoxicated	0.0281	0.0087	0.0325	0.0065
	sober	0.0043	0.0000	0.0065	0.0173

Table 3: normalized error rates (tongue twisters) of the different groups for each gender and condition

error		gender			
		o	i	s	r
f	intoxicated	0.0345	0.0575	0.0230	0.0345
	sober	0.0330	0.0000	0.0330	0.0330
m	intoxicated	0.0065	0.0129	0.0129	0.0194
	sober	0.0194	0.0258	0.0194	0.0129

Table 4: normalized error rates (addresses) of the different groups for each gender and condition

Based on the same data we didn't find any traces for incomplete articulations in the sense of (Künzel/Braun/Eysholdt, 1992) in both conditions. For example, looking at "... sein Geäst stromwärts ..." the phoneme sequence /stStr/ wasn't reduced by any of our investigated speakers irrespective their BAC. There were no elisions and no incomplete closures of the plosives /t/ or /d/ as far as our recordings are concerned.

9. Availability

The ALC corpus will be made available for unrestricted scientific and commercial usage (first edition scheduled for Aug 2008). Interested parties may obtain copies of the corpus at BAS. Please contact Florian Schiel schiel@bas.uni-muenchen.de or refer directly to the BAS catalogue at www.bas.uni-muenchen.de/Bas.

10. Acknowledgments

The work presented in this paper as well as the resulting speech corpus ALC was funded by the Bavarian Archive for Speech Signals (BAS) at the Ludwig-Maximilians-Universität München, Germany as well as by the 'Bund gegen Alkohol und Drogen im Straßenverkehr e.V.' (B.A.D.S., League Against Drug and Alcohol Abuse in Traffic), Hamburg, Germany and the BAS Services Schiel, Munich, Germany. The authors thank all colleagues within the BAS and the Institute of Legal Medicine for their help and support.

³including the interview partner

11. References

- O.M. Cooney, K. Mc Guigan, P. Murphy, R. Conroy. 1998. Acoustic analysis of the effects of alcohol on the human voice. *Journal of the Acoustical Society of America*, p. 2895.
- D.M. Behne, S.M. Rivera, D.B. Pisoni. 1991. Effects of Alcohol on Speech: Durations of Isolated Words, Sentences and Passages. *Research on Speech Perception*, No. 17, pp. 285-301.
- D.M. Behne, S.M. Rivera. 1990. Effects of alcohol on speech: Acoustic analysis of spondees. *Research on Speech Perception*, No. 16, pp. 263-291.
- A. Braun. Speaking while intoxicated: Phonetic and forensic aspects. 1991. *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence*, pp. 146-149.
- Chr. Draxler, K. Jänsch. 2004. SpeechRecorder – a Universal Platform Independent Multi-Channel Audio Recording Software. *Proc. of the IV. International Conference on Language Resources and Evaluation, Lisbon, Portugal*.
- K. Johnson, D.B. Pisoni, R.H. Bernacki. 1990. Do voice Recordings Reveal whether a Person is Intoxicated? A Case Study. *Phonetica*, vol. 41, pp. 215-237.
- H. Hollien, G. De Jong, C.A. Martin, R. Schwartz, K. Liljégren. 2001. Effects of ethanol Intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America*, pp. 3198-3206.
- F. Klingholz, R. Penning, E. Liebhardt. 1988. Recognition of low-level— alcohol intoxication from speech signal. *Journal of the Acoustical Society of America*, vol. 84, pp. 929-935.
- H.J. Künzel, A. Braun. 2003. The effect of Alcohol on Speech Prosody. *Proceedings of the International Congress of Phonetic Sciences, Barcelona*, pp. 2645-2648.
- H.J. Künzel, A. Braun, U. Eysholdt. 1992. Einfluß von Alkohol auf Sprache und Stimme. *Kriminalistik Verlag Heidelberg*.
- C.S. Martin, M. Yuchtman. 1986. Using speech as an Index of Alcohol-Intoxication. *Research on Speech Perception*, No. 12, pp. 413-426.
- H. Mögele, M. Kaiser, F.Schiel. 2006. SmartWeb UMTS Speech Data Collection. *The SmartWeb Handheld Corpus, Proc. of the LREC 2006, Genova, Italy*.
- L.C. Sobell, M.B. Sobell, R.F. Coleman. 1982. Alcohol-Induced Dysfluency in Nonalcoholics. *Folia Phoniatrica*, No. 34, pp. 316-323.
- L.C. Sobell, M.B. Sobell. 1972. Effects of alcohol on the speech of alcoholics. *Journal of Speech and Hearing Research*, No. 15, pp. 861-868.
- SpeechDat 1997. <http://www.speechdat.org/speechdat/deliverables/public/SD132V24.PDF>
- F. Trojan, K. Kryspin-Exner. 1968. The Decay of Articulation under the Influence of Alcohol and Paraldehyde. *Folia Phoniatrica*, No. 20, pp. 217-238.
- Verbmobil 1997. <http://www.phonetik.uni-muenchen.de/Forschung/Verbmobil/VMtrlex2d.html>
- H. Watanabe, T. Shin, H. Matsuo, F. Okuno, T. Tsuji, M. Matsuoka, J. Fukaura, H. Matsunaga. 1994. Studies on Vocal Fold Injection and Changes in Pitch Associated with Alcohol Intake. *Journal of Voice*, pp. 340-346.