The entropy of intoxicated speech – lexical creativity and heavy tongues

Uwe D. Reichel, Thomas Kisler

Institute of Phonetics and Speech Processing, University of Munich, Germany
{reichelu|kisler}@phonetik.uni-muenchen.de

Abstract

Spontaneous speech produced in sober and intoxicated conditions has been compared in information theoretic terms on the phoneme and word level to examine phonological and lexical aspects of intoxication. Word level entropy has been calculated to capture roughly the effect of alcohol on cognitive lexical creativity. Phoneme level entropy is intended to reflect heavy tongue influences on phoneme combinations. Moreover, mispronunciations have been investigated by relating canonical to realised pronunciation by means of mutual information and the Levenshtein distance. To account for the gradual nature of intoxication, examinations have been carried out regarding the offsets and slopes of linear functions mapping the blood alcohol concentration (BAC) to the information theoretic variables. It turned out that male speakers compensate less for the alcohol-induced degradations with regard to lexical creativity and articulatory precision than female speakers. Furthermore, the pronunciation of male speakers generally deviates more from canonical forms.

Index Terms: intoxication, entropy, transcription similarity

1. Introduction

Intoxication generally causes degradations of cognitive functions and of motor control. With respect to speech the impact of intoxication can be subdivided into gross, segmental and suprasegmental effects [1]. Gross effects are related to the stage of cognitive speech planning and consist in word and morphology level alternations [2]. Segmental effects as a result of articulation errors mainly consist in phone substitutions and omissions [3]. Suprasegmental effects affect speaking rate (generally reduced) [4], rhythmic variability (sometimes increased) [4], as well as changes in pitch range and variability [3]. Phonetic differences between sober and intoxicated speech often turned out to be non-uniform and speaker-dependent [3]. Generally, only minor gender differences were to be reported [5]. The examined factors also have been used to train classifiers for automatic intoxication detection [6].

Common practice of intoxication studies is a categorical division of the examined subjects into a sober and one or more intoxicated groups accounting only approxi-
crease of assimilations and thus higher segment predictabilities, which is measurable in terms of lower phoneme entropy values in the spontaneous transcriptions $T_s$.

3. provokes a higher amount of mispronunciations so that $T_s$ deviates more from the canonical forms $T_c$ which will be reflected in terms of lower mutual information and higher Levenshtein edit distances between $T_c$ and the $T_s$.

3.1. General procedure

To account for the gradual nature of intoxication we did not compare the examined variables word/phoneme entropy, mutual information and Levenshtein distance in a binary sober vs. intoxicated paired test setting, but compared offsets and slopes of linear trajectories keeping track of the continuous BAC domain. For each variable in question and for each speaker two values are given: $y_0$ for the sober condition ($BAC=0$), and $y_{bac}$ for a BAC value measured in the intoxicated condition. We fitted a trajectory through these two values as a linear function of the BAC. In the upper left plot of Figure 1 the centroid word entropy trajectories derived from the mean offsets and slopes are shown for female, male, and all speakers. We further accounted for the relative change of a variable value $y$ by fitting trajectories through normalised values defined by the quotient $y_{bac}/y_0$. In the upper right plot of Figure 1 these trajectories are shown in gray color for female and male speakers. Again centroid trajectories are derived from the (constant) offset and the mean slopes and are plotted in black.

We then compared across the female and male subject group (1) the offsets $y_0$ of the $y$-trajectories, (2) the slopes $b$ of these trajectories $y_{bac} = y_0 + b \cdot bac$, and (3) the slopes $c$ of the normalised trajectories $y_{bac}/y_0 = 1 + c \cdot bac$.

Since no significance differences between (2) and (3) have been observed, we only present (1) and (3) in this paper.

3.2. Lexical creativity: Word entropy

**Method** For each session the word cross entropy with respect to the probability model $P$ was calculated for the word sequence $W = w_1 \ldots w_n$:

$$H(W) \approx -\frac{1}{n} \log P(w_1 \ldots w_n) \quad (1)$$

$P$ is given by a linear interpolated word bigram probability model which was calculated over all sessions. Good Turing smoothing was applied to the n-gram counts, and the interpolation weights were calculated by means of an Expectation-Maximisation algorithm.

Generally, high entropy values correspond to low word predictabilities. Since a higher degree of lexical creativity is assumed to be expressed in lower word predictabilities it should be marked by higher entropy values.

As described in 3.1, $H(W)$ is modelled as a linear function of BAC, and the offsets $H(W)_0$ (see Figure 1, left half) and the slopes $c$ of the the normalised $H(W)$ trajectories (right half) are compared across male and female subjects.

**Results** As can be seen in Figure 1 mean male trajectories show a negative slope starting from a higher offset $H(W)_0$, while female trajectories have a positive slope starting from a lower offset. $H(W)_0$ differed significantly (two-tailed t-test, $t_{160} = -3.55, p < 0.001$) while the slope $c$ shows a weakly significant difference (two-tailed t-test $t_{160} = 1.66, p < 0.1$).

The predication of a degradating effect of alcohol on lexical creativity was therefore only supported for the male group, while for women rather the opposite pattern emerged.

3.3. Phonotactics: Phoneme entropy

**Method** Analogously to the word level examinations, for each session the phoneme entropy was calculated with respect to a phoneme bigram model trained on the spontaneous speech transcriptions $T_s$ of all sessions.

**Results** The differences between male and female speakers are presented in Figure 2, on the left for the trajectory offsets $H(T_s)_0$, and on the right for the normalised transition slopes $c$. The results turned out to mimic those found for the word level. Again mean male trajectories show a higher offset and a negative slope, while female trajectories show the opposite pattern.

![Figure 1: Word bigram entropy $H(W)$. Top-left: Mean entropy trajectories for female (solid), male (dashed) and all (dotted) speakers as linear functions of BAC. Top-right: Mean normalised entropy trajectories for female (solid) and male (dashed) speakers derived from the speaker-dependent observations (gray). Bottom-left: Entropy in sober condition. Bottom-right: The slope of the normalised entropy trajectories for female and male speakers.](image-url)

As described in 3.1, $H(W)$ is modelled as a linear function of BAC, and the offsets $H(W)_0$ (see Figure 1, left half) and the slopes $c$ of the the normalised $H(W)$ trajectories (right half) are compared across male and female subjects.
The mutual information between the canonic Method information

3.4. Deviation from canonic pronunciation: Mutual information about the other, and therefore, how close the spontaneous speech is related to the canonical forms.

Method The mutual information between the canonic $T_c$ and the spontaneous speech transcriptions $T_s$ has been calculated for each session and is derived from the speaker-dependent observations (gray).

Results While the general MI tendencies correspond to those of $H(W)$ and $H(T_s)$ which is shown in Figure 3, this time none of the differences between male and female subjects turned out to be significant (two-tailed t-test, $p > 0.3$). Only male speech shows the expected tendency to diverge more from canonical forms under the influence of alcohol.

3.5. Deviation from canonic pronunciation: Edit distance

Method Generally, the Levenshtein distance between to sequences $T_c$ and $T_s$ is given as the minimum number of edit (i.e. substitution, insertion, and deletion) operations to convert $T_c$ to $T_s$. This distance directly emerged from the string alignment carried out by the preceding statistical PermA alignment. Editing is weighted by 1 minus the co-occurrence probabilities of $T_c$ and $T_s$ symbols. Furthermore, distances are normalised with respect to sequence length by dividing them by the length of $T_c$. For each session the mean normalised weighted Levenshtein distance has been calculated this way. Like the Mutual information measure it gives a notion about how close $T_c$ and $T_s$ are related, this time focusing rather on similarity and not on systematic correspondencies.

Results As can be seen in Figure 4 for both genders Levenshtein distances rise with increasing intoxication supporting our expectation about the impact of alcohol on increasing deviation from canonical speech. The male trajectory starts significantly higher than the female one indicating an overall higher deviance level for male speakers (two-tailed t-test, $t_{160} = -2.26, p < 0.05$). The slopes do not differ significantly (two-tailed t-test, $p > 0.65$).

4. Discussion and conclusion

In this study we have modelled cognitive and motor aspects of intoxication in speech production in information theoretic and edit distance terms. To account for the gradual nature of intoxication we treated the chosen measures as linear functions of the BAC and compared the offsets and slopes of these functions across female and male speakers.
alternative to the linear trajectories since only two data points were available for each speaker not allowing for fitting more complex functions.

4.3. Interpretation of the results

For our data we concluded from entropy comparisons, that male speakers show a higher lexical variability and richer phoneme combinatorics in sober condition than female speakers, which both decline under the influence of alcohol, whereas for female speakers they rise. The impact of alcohol therefore seems to be rather restraining for men while women seem to spend more effort in compensating for the alcohol-induced degradations.

At the segmental level potential gender- and intoxication-related differences might be obscured by the used automatic segmentation method which being trained on sober speakers’ data might simply not track phoneme sequences only occurring in intoxication due to their low probability. Nevertheless, for all speakers the deviation of the utterance from the canonical forms slightly increases, starting at a higher level for male speakers indicating that they generally stick to a lesser extent to canonical pronunciation.

5. Acknowledgments

The work of the authors was carried out within the CLARIN-D project [11] (BMBF-funded).

6. References