FUNDAMENTAL FREQUENCY AND HUMAN PERCEPTION OF ALCOHOLIC INTOXICATION IN SPEECH

Barbara Baumeister, Florian Schiel

Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich
bbalschiel@phonetik.uni-muenchen.de

ABSTRACT

The majority of speakers raise their fundamental frequency when speaking while intoxicated. In this study we describe three perception experiments based on manipulated sober and intoxicated speech from the German Alcohol Language Corpus to answer the question whether human listeners use fundamental frequency as a cue to recognize an intoxicated person. Our results show that although fundamental frequency is a good indicator for intoxication, listeners do not predominantly use this feature. A possible explanation is that fundamental frequency is also influenced by other speaker states such as fatigue, mood etc.

Keywords: alcohol, perception, intoxication, fundamental frequency

1. INTRODUCTION

Previous findings concerning the effect of intoxication on fundamental frequency (f0) vary as much as the experimental setups do. A significant increase in average f0 with intoxication is reported in [8], [7] and [3]. On the other hand in [15] and [1] a decrease was found. In some studies (e.g. [14], [11], [6]), no significant change in average f0 could be found, while in [9] f0 varies non-linear with breath alcohol concentration. Regarding the range of f0 most earlier studies are consistent in suggesting an overall increase of f0 range in the intoxicated condition. However, all of these studies dealt with a relatively low number (ranging from 4 to 35) of mostly male participants.

Regarding the human ability to perceive alcoholic intoxication of a speaker solely by hearing a speech sample, again previous findings vary considerably. In a forced choice identification task [10], where speech samples of 8 speakers were judged by 44 listeners, the accuracy rate was 61.5%. In [8] listeners discriminated between two stimuli (which is a much easier task), and reached a relatively high rate of 82%. However, this rate only holds for speakers whose blood alcohol concentration (BAC) was above 0.1%. For less intoxicated speakers, listeners recognized only 54.2% of the intoxicated stimuli. Based on a small part of the German Alcohol Language Corpus (ALC) a perception test with speech samples of 16 speakers (8f, 8m, BAC 0.05% - 0.142%) was conducted with 47 listeners [12], who reached an average discrimination rate of 71.7%. The relatively low number of speakers may have made the discrimination task easier for listeners.

In this study we describe three perception experiments to test f0 as a perceptual cue when discriminating intoxicated from sober speech.

The first perception experiment tests the general ability of listeners to discriminate between sober and intoxicated stimuli pairs, and how the performance of the listeners correlates with f0 features.

The second experiment involves the same sober and intoxicated stimuli as before, but we adjusted the intoxicated stimulus so that f0 had the same level and range as that of the sober stimulus. The hypothesis is that discrimination performance will drop on these data.

In the third experiment we use control group recordings where the speaker was sober in both stimuli, but in one of them f0 was up-shifted and stretched by a fixed proportion. If f0 is a cue for intoxication, discrimination performance on these manipulated data should be above chance.

2. SPEECH DATA

The ALC contains recordings of 142 German male and female speakers, in sober and intoxicated condition, and of 20 speakers in two sober and one intoxicated condition. A detailed description of the ALC and the recordings can be found in [13]. The ALC may be downloaded for free by academic users via the CLARIN repository [5] at the Bavarian Archive for Speech Signals (BAS); commercial licenses may be obtained via the ELRA or BAS.

3. EXPERIMENT DESIGN

3.1. Experiment 1 - General ability

To test the general ability of listeners to discriminate between intoxicated and sober speech, 8 stimuli pairs of each speech style, read speech, command and control speech and spontaneous speech, spoken by 132 speakers are selected. One speaker of ALC had to be excluded due to too much laughter in her spontaneous speech. 29 speakers showed audible artefacts after f0 manipultation (in experiment 2 and 3); to avoid that listeners judge stimuli as intoxicated because of these artefacts, we excluded these speakers from all perception experiments. The spontaneous speech stimuli were checked for speech errors or laughter, which should occur in either none or both conditions. The mean duration of a single stimulus varies from 0.8s to 15.8s (median is 3.9s) but is similar within one pair of stimuli. As a control condition additional stimuli pairs of 20 speakers, where both stimuli involved sober speech from the same speaker, were added to the set.

3.2. Experiment 2 - Compensation of f0 effects

To keep the results comparable to experiment 1, the same 24 discrimination pairs of stimuli spoken by 132 speakers were used.

To minimize the influence of f0 on listener decisions, within each pair we manipulated the intoxicated stimulus so that both the mean f0 and the range of f0 resembled that of the sober stimulus. The f0 contour of the intoxicated stimulus was multiplied with the ratio of the median of the sober contour to the median of the intoxicated contour, and then stretched or compressed around the (new) median, depending on the ratio of the interquartile ranges of the two stimuli. The following formula was applied to the original f0 contour of the intoxicated speech stimulus:

$$(1) \quad f0_{new} = k_2 k_1 f0_{intox} + (1 - k_2) k_1 median_{f0_{intox}}$$

where

$$(2) \quad k_1 = \frac{median_{sober}}{median_{intox}}$$

and

$$(3) \quad k_2 = \frac{iqrange_{sober}}{iqrange_{intox}}$$

and *iqrange* is the interquartile distance 25 - 75.

Figure 1: *F0 contours of one pair of stimuli, intoxicated on the left and sober on the right*

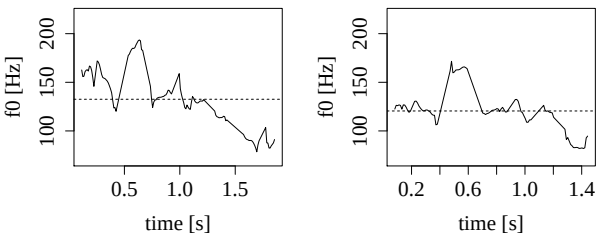


Figure 2: *Original (grey) and manipulated intoxicated f0 contour (black)*

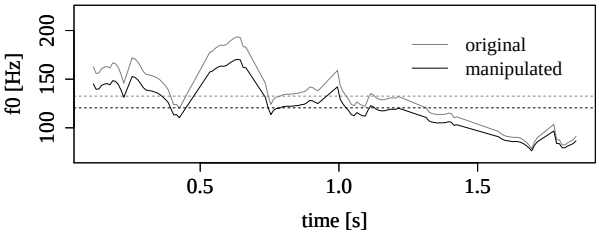


Figure 1 shows two f0 contours from the same speaker, articulating the German phrase "Temperatur 23°C", sober on the right and intoxicated on the left. The dashed lines show the f0 median for each stimulus. Figure 2 displays the intoxicated contour after manipulation (black) compared to the original contour (grey).

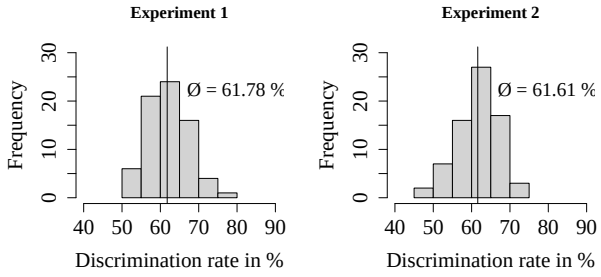
F0 manipulation using PSOLA works very well for small changes (less than 10%), but introduces audible artefacts for larger f0 changes. To avoid such artefacts distracting listeners or even inducing them to choose the "weird" stimulus as the intoxicated one, we limited the maximum f0 change to 10%.

3.3. Experiment 3 - Simulation of f0 effects

The same listeners who participated in experiment 2 were asked to discriminate the 480 sober stimuli pairs of the control set of experiment 1. For this the contour of one sober stimulus of each pair was upshifted by a fixed value of 5% and also stretched by 5%.

In all three experiments f0 contours were extracted using Praat [4], manipulated with an algorithm in R and re-synthesized using the standard PSOLA settings in Praat. To avoid the influence of the audibility of the re-synthesis of the stimuli on the listener's decision, both stimuli of one pair were re-sythesized, one of them without previous manipulation of the f0 contour in experiment 2 and 3.

Figure 3: *Histogram of discrimination rates of 72 listeners in experiment 1 (left) and experiment 2 (right)*



3.4. Perception test

The stimuli for all three experiments were presented pairwise in random order in a forced-choice discrimination test. One listener group completed experiment 1 and a second listener group experiments 2 and 3. Listeners were asked to choose (after a maximum of five repetitions) the stimulus where the speaker sounded intoxicated. Because of the large amount of stimuli pairs per listener group, each listener judged only a subset (161 + 20) with exactly one stimuli pair per speaker. This resulted in 24 subsets of stimuli with balanced speech styles. Three different listeners evaluated each subset of stimuli, i.e. each stimuli pair was judged three times by three different listeners, which leads to a total number of 72 (36 female and 36 male) listeners per group. Listeners were native German speakers and aged between 19 and 36 (median 23.5) in experiment 1 and aged between 20 and 36 (median 23) in experiments 2 and 3.

4. RESULTS

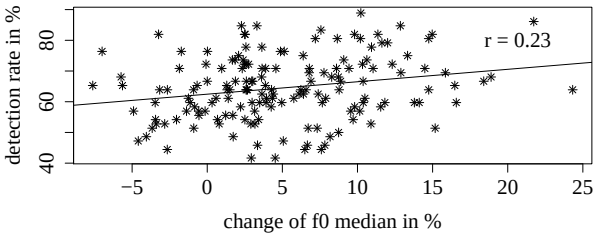
4.1. Experiment 1 - General ability

Results showed that the average overall performance of the listeners (mean percentage of correct answers given by each listener) was 61.8% which is above chance. Discrimination rates varied among listeners from 51.5% to 76.5% (see figure 3 on the left).

In the stimuli pairs of the control set (two sober stimuli compared), listeners chose randomly between the two recordings. It follows that there are no hidden factors in the different recording setups that bias listener judgements.

A logistic mixed effect model analysis [2] with the listener's answers as dependent variable and listeners and speakers treated as random factors showed a significant influence of f0 median changes in the stimuli on the listener's decision ($p < 0.001$). The larger the difference in the f0 medians, the more likely the stimulus with higher f0 median was chosen to be the intoxicated one.

Figure 4: *Correlation between difference in f0 and detection rates per speaker*



A weak correlation between the relative change of f0 per speaker and the speaker's detection rate was found ($r = 0.23$). Speakers with a larger mean f0 difference between stimuli tend to be recognized better than speakers with smaller differences. The scatter plot with one dot per speaker is shown in figure 4.

91,7% of the listeners chose the stimulus with higher f0 as the intoxicated one in more than 50% of the cases. This shows a general listener preference for the stimulus with higher f0. The listener with the strongest preference chose the stimulus with higher f0 in 68% of the pairs.

4.2. Experiment 2 - Compensation of f0 effects

The performance of the listeners in experiment 2 was on average 61.6% and varied among listeners from 48.5% to 74.2%. Figure 3 shows the histogram of listener discrimination rates for manipulated f0 (right) in comparison to the test with the original stimuli (experiment 1, left). The average performance of listeners is not visibly worse than in the test with the original stimuli.

A logistic mixed effects model test with listeners and speakers treated as random factors showed no significant difference between the results of experiment 1 and 2.

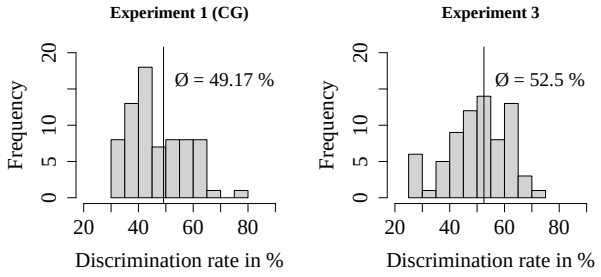
Since we obtained the same results for two listener groups, we conclude that this (non)effect is robust across different listener populations.

4.3. Experiment 3 - Simulation of f0 effects

In experiment 3 the mean discrimination rate is 52.5%. The listeners showed a tendency to choose the stimulus with the up-shifted and stretched f0 contour to be intoxicated. In figure 5 the discrimination rates across listeners for experiment 3 compared to the test with the unaltered stimuli (experiment 1) are shown.

Although there seems to be a visible differ-

Figure 5: *Histogram of discrimination rates of 72 listeners for the 5% up-shifted (right) and the unaltered stimuli (left) on the control group data*



ence, a mixed effects model analysis only reveals a marginally significant difference between both experiments ($p < 0.1$).

5. DISCUSSION AND CONCLUSION

In experiment 1 a (weak) correlation between change in f0 and the listener's ability to discriminate between sober and intoxicated speech was found. This suggests that listeners use f0 as a cue for intoxication. But in the second perception experiment listeners perform the same, even when differences in f0 were eliminated from the stimuli pairs. This seems to indicate that f0 does not function as a cue for listeners in discriminating between sober and intoxicated speech. One possible explanation is that the change in f0 is loosely correlated with other features that are exploited by listener to discriminate between intoxicated and sober speech. Our conclusion from the outcome of experiment 2 is therefore that in human perception features other than f0 play the major role in distinguishing intoxicated vs. sober speech. These features may be other acoustic features, linguistic or para-linguistic features (which are not easily measured automatically from the speech signal).

Another interesting result is the difference in performance variation across listeners and speakers: While the distribution of discrimination performance of the listeners is quite narrow (Fig. 3), the same distribution across speakers is much wider (not shown). This means that listeners in average (i.e. over a large number of different speakers) do not differ widely in their ability to detect intoxication from speech. On the other hand speakers differ considerably in their expression of intoxication: some are easily spotted, others camouflage their condition perfectly. Some even show an 'inverse' behavior: they appear to be intoxicated, when in fact they are sober.

In experiment 3 listeners showed a tendency to choose the stimulus with the altered f0 as intoxicated. At first glance this seems to contradict the outcome of experiment 2. But if we assume that our explanation for experiment 2 is true, it might be that listeners use f0 as a 'fall-back' feature, if no other features for intoxication can be detected in the speech signal (as in experiment 3 where both stimuli were extracted from sober recordings). A simple 5% increase in the f0 median and f0 range seems (in a few cases) to cause the listener to choose the manipulated stimulus as the intoxicated one.

To summarize, the results of this and previous studies suggest that f0 functions as a promising feature in (speaker-dependent) automatic detection of intoxication based on the speech signal, but that it is not a major cue in human perception of alcoholic intoxication. The latter aspect might be important in forensic cases where witnesses claim to recognize intoxication from the speech signal alone (e.g. via a phone connection). A possible reason why listeners do not exploit f0 might be that it is prone to changes caused by many other user states, such as stress or positive emotions, and could therefore in many real life situations lead to misleading classifications.

6. ACKNOWLEDGMENTS

This work was partly supported by the Deutsche Forschungsgemeinschaft, funding number SCHI1117/1-1, and the Bavarian Archive of Speech Signals (BAS), Ludwig-Maximilians-Universität München, Germany. We also would like to thank the ALC team at BAS for producing the speech data and the orthographic transcription, and the European CLARIN consortium for providing the ALC speech data for the scientific community [5].

7. REFERENCES

[1] Aldermann, G. A., Hollien, H., Martin, C., DeJong, G. 1995. Shifts in fundamental frequency and articulation resulting from intoxication. *Journal of the Acoustical Society of America* 97, 3363–3364.

[2] Baayen, R. H. 2008. *Analysing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

[3] Baumeister, B., Heinrich, C., Schiel, F. 2012. The influence of alcoholic intoxication on the fundamental frequency of female and male speakers. *Journal of the Acoustical Society of America* 132, 442–451.

[4] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.3.66, retrieved 9 March 2014 from http://www.praat.org/.

[5] CLARIN Repository Bavarian Archive of Speech Signals, Ludwig-Maximilians-Universität, München, accessed January 2015, http://hdl.handle.net/11858/00-1779-0000-0006-BF00-E. .

[6] Cooney, O. 1998. Acoustic analysis of the effects of alcohol on the human voice. Master's thesis Dublin City University.

[7] Hollien, H., DeJong, G., Martin, C. A., Schwartz, R., Liljegren, K. 2001. Effects of ethanol intoxication on speech suprasegmentals. *Journal of the Acoustical Society of America* 110(6), 3198–3206.

[8] Klingholz, F., Penning, R., Liebhardt, E. 1988. Recognition of low-lewel alcohol intoxication from speech signal. *Journal of the Acoustical Society of America* 84(3), 929–935.

[9] Künzel, H., Braun, A. 2003. The effect of alcohol on speech prosody. *Proceedings of the ICPhS2003, Barcelona, Spain* 2645–2648.

[10] Martin, C. S., Yuchtman, M. 1986. Using speech as an index of alcohol-intoxication. *Research on Speech Perception* 12, 413–426.

[11] Pisoni, D. B., Hathaway, S. N., Yuchtman, M. 1985. Effects of alcohol on the acoustic-phonetic properties of speech: Final report to GM Research Laboratories. *Research on Speech Perception Progress Report* 11, 109–171.

[12] Schiel, F. 2011. Perception of Alcoholic Intoxication in Speech. *Proc. of the Interspeech 2011, Florence, Italy,* 3281–3284.

[13] Schiel, F., Heinrich, C., Barfüsser, S. 2012. Alcohol language corpus: The first public corpus of alcoholized German speech. *Language resources and evaluation* 46(3), 503–521.

[14] Sobell, L. C., Sobell, M. B., Coleman, R. F. 1982. Alcohol-induced dysfluency in nonalcoholics. *Folia Phoniatrica* 34, 316–323.

[15] Watanabe, H., Shin, T., Matsuo, H., Okuno, F., Tsuji, T., Matsuoka, M., Fakaura, J., Matsunaga, H. 1994. Studies on vocal fold injection and changes in pitch associated with alcohol intake. *Journal of Voice* 8(4), 340–346.