

Recognition of Psychologically Relevant Aspects of Context on the Basis of Features of Speech

Anthony Jameson¹, Barbara Großmann-Hutter², Christian Müller², Frank Wittig²,
Juergen Kiefer³, and Ralf Rummer³ *

¹ DFKI, German Research Center for Artificial Intelligence

² Department of Computer Science, Saarland University

³ Department of Psychology, Saarland University

Abstract. Especially in mobile systems, one important part of the context of use involves psychological variables like cognitive load and time pressure. This abstract looks at one possible way of capturing such aspects of context: the analysis of features of the users' speech. In a replication and extension of an earlier study of our group, we created four experimental conditions that varied in terms of whether the user was (a) navigating within a simulated airport terminal or standing still; and (b) subject to time pressure or not. The speech produced by these subjects was coded in terms of 7 variables. We trained dynamic Bayesian networks on the resulting data in order to see how well the information in the users' speech could serve as evidence as to which condition the user had been in. The results give information about the accuracy that can be attained in this way, the methods that can be used to implement the classifiers, and the diagnostic value of some specific features of speech.

1 Background and Motivation

When we think about modeling and representing context, we may think first in terms of sensors that directly detect features of the environment, such as the location of the user, the presence of other persons, physical features like temperature and noise level, or activities that the user is engaged in. But the importance of these features of context is often due to the psychological effects that they have on the user. For example, the fact that a user is engaged in communication with other persons may be important mainly because it implies that the user has little time and attention left over for interacting with a system.

It is therefore natural to view the contextually influenced psychological states of the user as constituting an important part of the context. But how can these psychological states be detected by a system?

* The research summarized here was supported by the German Science Foundation (DFG) in its Collaborative Research Center on Resource-Adaptive Cognitive Processes, SFB 378, Projects B2 (READY) and A2 (VEVIAG). We thank one of the anonymous reviewers for perceptive comments on the submitted version of the manuscript.

Two strategies, which are not mutually exclusive, can be distinguished:

1. The system can detect objective features of the context and make inferences about the psychological states that they are likely to induce; and
2. The system can detect behavioral or other responses of the user that can be treated as symptoms of the psychological states in question. In this abstract, we focus on the second approach, though we believe that in general a combination of the two approaches should be considered.

One class of symptoms of psychological states that have often been considered comprises physiological responses that can be detected by sensors attached to the user's body. In this abstract, we discuss another sort of symptom, which may be especially useful when a system is involved that requires the user to produce a good deal of speech (e.g., when giving commands via speech or creating voice recordings). An obvious first question is: Is there a useful amount of information available in a user's speech that can enhance the recognition of contextually determined psychological states? A partial answer to this question was given in an earlier publication from our group (Müller et al. [1]) that studied the recognition of two states: cognitive load and time pressure. In the present abstract, we summarize a replication and extension of the experiment reported on in the earlier paper—a second experiment that both corroborates the initial results and adds some new ones.

Because of the space limitation, for details of the methods and results the reader must be referred to the poster presentation at the workshop, the earlier paper by Müller et al. [1], and/or the originally submitted longer version of the present abstract.¹

2 Summary of Methods and Results

The basic design of the two experiments is illustrated in Figure 1. Each subject took the role of a traveler in an airport terminal, which was simulated on a PC screen. The subject's main task was to ask questions of two fictitious airport helpers by speaking into a microphone. In Experiment 1, which is illustrated in the left-hand side of Figure 1, two variables were manipulated experimentally:

- Navigation: Whether or not subjects were required to navigate within the simulated airport terminal while speaking.
- Short-term time pressure: Whether or not the subjects were motivated to formulate their questions quickly.

On the basis of the data acquired from 32 subjects, we experimented with the learning of dynamic Bayesian networks that were designed to recognize what condition a subject was in on the basis of several features of the subject's speech, such as the length of utterances, articulation rate, frequency and duration of pauses, and several types of disfluency.

¹ By the time of the workshop, these materials will be available via the web page <http://dfki.de/~jameson/mrc05/>.

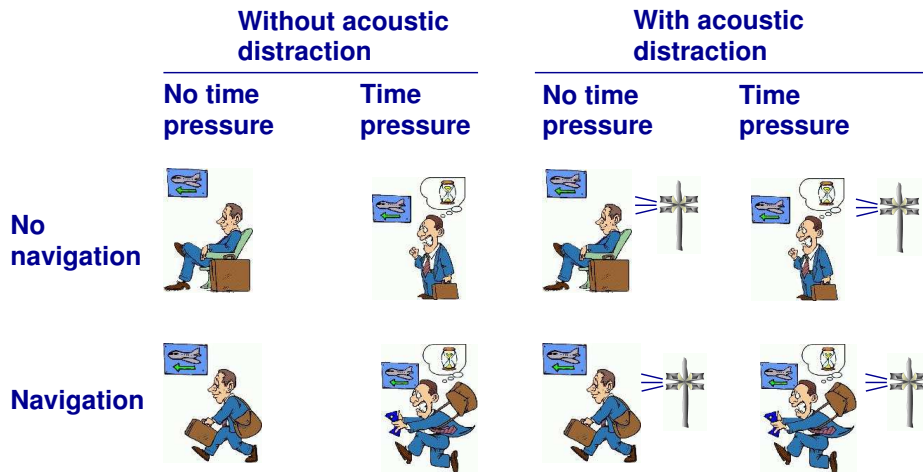


Fig. 1. Illustration of the eight conditions realized in Experiments 1 (left) and 2 (right).

The results were moderately encouraging, and they shed some light on the diagnostic value of the various features of speech. But it seemed that many subjects were able to handle the navigation task so easily that it induced too little cognitive load to affect their speech. One motivation for Experiment 2 was the desire to see if the navigation task would have more noticeable effects in a situation where the subject was already distracted by another contextual factor. We therefore replicated the experiment in the way illustrated in the right-hand side of Figure 1: While the subjects performed their tasks, typical airport announcements (which had been recorded at Frankfurt Airport) were played back to them.

Even though the subjects were not required to pay attention to the content of the announcements, they did report that the announcements made it more difficult for them to generate appropriate questions. Consistent with this result, the difference between the navigation and the no-navigation conditions was easier to detect in this experiment. Evidently, because of the increased distraction, the subjects more often showed speech symptoms of cognitive overload while navigating.

In other respects, the results of Experiment 2 corroborated those of Experiment 1.

We also repeated the learning experiments while systematically leaving out one feature of speech at a time, so as to determine which ones might be dispensable. This analysis revealed that the features that were most difficult to detect automatically could be omitted with little loss in accuracy.

In sum, the two experiments showed in a consistent way that contextually induced cognitive load and time pressure can (at least in some situations) have effects on features of the person's speech that are strong enough to permit significantly above-chance discrimination; and they yield information about the diagnostic value of particular features.

Any attempt to apply the ideas and results of these experiments in a particular application scenario will necessarily involve considerable further work and creativity. But we believe that the results of these experiments will be helpful as a starting point.

References

1. Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R., Wittig, F.: Recognizing time pressure and cognitive load on the basis of speech: An experimental study. In Bauer, M., Gmytrasiewicz, P., Vassileva, J., eds.: UM2001, User Modeling: Proceedings of the Eighth International Conference. Springer, Berlin (2001) 24–33