Collaborative Human-Machine Communication: User-Centered Design of In-Vehicle Speech Dialog Systems

Linn Hackenberg

Collaborative Human-Machine Communication: User-centered Design of in-vehicle Speech Dialog Systems

Linn Hackenberg

VOLKSWAGEN GROUP
38436 Wolfsburg
Germany

Department of Psychology, University of Brunswick, Brunswick, D
linn.hackenberg@volkswagen.de

Dissertation advisor: Prof. Dr. Mark Vollrath, University of Brunswick
mark.vollrath@tu-braunschweig.de

Research-area: User-centered Design, speech dialog systems, in-vehicle HMI

Description of the research topic: Evaluation of Speech Dialog Systems that make use of collaborative strategies from human conversations by providing continuous and appropriate feedback whilst showing dynamic interaction-structures.

1 Research Problem

One of the main challenges of today's speech dialog systems (SDS) is that they fail to provide appropriate feedback about status and capabilities of the system. Until now they lack any representation of comprehension. Immediate acknowledgements or back-channels, known from everyday conversation with people [1] are missing. This leads to high uncertainty of users about what the system is up to. Operation errors like talking too early or entering input again while the system is still processing are due to state confusion which demonstrates the users' need for process indicators. These operating errors consequently lead to recognition errors followed by a low user satisfaction. In addition, there is substantial evidence that counter-intuitive system interaction leads to high cognitive workload [2], which is a severe problem especially in the in-vehicle dual-task context.

Another challenging issue is the perceived length of dialogs. Inflexible interaction structures of current systems lead to low user satisfaction with the efficiency of in-vehicle SDS. Analogous to Grice's maxim of manner [3] the least long dialog is preferred. Rather than the number of words used per speech output, length here refers to situation-appropriate system responses. In particular, dispensable dialog loops demanding the user to confirm his input are often misinterpreted as recognition errors

and eventually lead to poor ratings of the system. Moreover, these redundant confirmations provoke specific operating errors (talking to early), hyperarticulation and frustration on the part of the users. Latter goes back to a mismatch with existing mental models of interpersonal communication. Here the amount of effort that both partners dedicate to a dialog step before moving on is determined by their grounding criteria [4] and is highly context-sensitive. The higher the grounding criterion is set, the more evidence conversational partners will require before concluding that an utterance is accepted [5]. The problem postulated here is that the systems grounding criterion is inflexible and so at most of the turn-takes too high.

Until now, the process of grounding, which entails people systematically seeking and providing evidence about what has been said and understood does not sufficiently take place with SDS. It can be shown that this lack of grounding reduces the intention to use and the joy-of-use experience. Therefore, transparency is an extremely important precondition in order to create user acceptance for new speech interfaces.
Traditional approaches mainly focus on improving the correctness of the underlying speech recognition. The presented experiments employed here establish dialog strategies from human conversation to investigate whether these strategies can reduce the amount of misrecognition and have an impact on system usability ratings through reducing frustration and operating errors.

2 Research Hypotheses

The overall hypothesis is that, adapting the system to existing communication strategies and facilitating grounding processes should lead to improved user satisfaction. By testing systems, which use collaborative strategies known from human conversations against systems which do not, differences in usability ratings (efficiency, effectiveness, satisfaction), frustration indicators (hyperarticulation) and objective data (operation errors) are expected to be found. It is not expected that the on-demand visualization or the flexible grounding criterion affects neither gaze nor driving behavior. In particular, the first study was set up to evaluate additional graphical visualizations, to test design alternatives and to discuss distraction potential in the dual-task context.
Transparency of the speech interaction process through state feedback should help the user avoid seeking for evidence and thus reducing the operating errors. Providing visual acknowledgment about what has been said and understood should give the user a better sense of system capabilities.
Regarding the dynamic dialogs, the purpose of the second study was to see whether the flexible grounding criterion can enhance the efficiency and effectiveness of the interaction by reducing the amount of turn-takes. Moreover, the reduction of redundant confirmation questions should reduce the operating errors and misrecognitions. Both is expected to lead to better ratings of system usability.

3 Methods

Using a deductive research approach two empirical user studies were conducted to examine the impact of visual feedback and of the flexible system grounding criterion. A control-group design was chosen in order to evaluate the effects of the innovations. While driving the simulator the probands where using the SDS for severals task concerning the adressbook (e.g. making a call, starting navigation to a contact). The tested SDS differed in visualizations (none, states, content) and dialog behavior (static, dynamic). Each usability construct was measured by at least one objective and one subjective variable. In both studies gaze behaviour and driving dynamics were recorded to analyse the impact on distraction. The system's responses were logged and the dialog behavior, such as amount of turn-takes, hyperarticulation and misrecognitions were coded by two qualified raters. In order to examine the mental models, methods like free recall and recognition tests were employed. Subjective data were collected through standardized usability questionnaires (SUMI, SUS) and interviews.

4 Solutions

The grounding process requires partners to be able to find incremental evidence of each other´s understanding. There is substantial evidence that "feedback from a spoken language system need not be in the form of speech […] it can be graphics."[5]. Accordingly, an obvious first step was to implement a system that provides a supplementary visualization of the system status and dialog results. In addition to the latter, the following system states are shown; Ready, Receiving, Processing and Speaking.
These visual indicators should spare users the effort of guessing about system status and dialog intentions [6]. Given the fact that this work is set in an in-vehicle context, the visual applications for SDS have been adjusted to demands of dual-task contexts.
The results of the first simulator study clearly indicate the positive effect of visualizations, which are showing processing information and also representing the recognized slots by the system. Furthermore it could be shown that visual feedback can trigger turn-takes and increase the perceived system transparency. It was found that an abstract presentation of the system state does not cause significant increase in the head-down time when compared to a control group without visualization. In addition to these and other findings about usability and distraction, extensive design recommendations can be stated.

Another area of action concerns the adaptation of the system behavior to the current dialog situation. By implementing a flexible grounding criterion the system will only ask for confirmation if it is insecure, similar to what humans do. In consequence, the system demands confirmation only when an increased user effort is justified. This is the case, if the previous dialog turn was difficult (low accuracy), ambient noise is present or if a misunderstanding would have serious consequences. For each dialog step, a predefined function computes, whether a confirmation request is necessary at this point of the dialog. Trimming the dialogs in this specific way can lead to efficient and satisfactory speech interaction despite the persisting shortcomings of the speech rec-

ognizers. By reducing the amount of turn-takes one automatically reduces the possibility of recognition errors. Also, it can be shown that avoiding redundant confirmations also reduces the occurrence of operating errors and hyperarticulation, which has a positive impact on system evaluation.

5 Contributions

Primarily a feedback model and a grounding criterion function for human-computer speech interaction is presented, which is based on a collaborative theory of human communication [7]. The model is used to systematically provide additional visual feedback from a SDS, thus improving and expanding the cognitive model of the user with respect to system functionality and capabilities. Additionally the grounding criterion function realizes a system that shows confirmation requests only, when necessary and reduces thereby operating errors and misrecognitions.

Results from extensive system evaluation show that it reduces both, the uncertainty of the user about the skills of the system and, in consequence, the reservations towards the voice control interface.

References

1. Yngve, V. H.: On getting a word in edgewise. In: Papers from the sixth regional meeting of Chicago Linguistic Society, pp. 567--578. Chicago Linguistics Society, Chicago (1970)

2. Burnett, G. E., Joyner, S.M.: An assessment of moving map and symbol-based route guidance systems. In: Ian Noy, Y. (eds.) Ergonomics and safety of intelligent driver interfaces, pp. 115--136. Lawrence Erlbaum Associates, Mahwah (1997)

3. Grice. H. P.: Logic and conversation. In: Cole, P., Morgan, J. L. (eds.) Syntax and Semantics III: Speech Acts, vol. 58, pp. 41. Academic Press, New York (1975)

4. Clark, H. H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition. 22, 1--39 (1986)

5. Brennan, S. E., Hulteen, E.: Interaction and feedback in a spoken language system: A theoretical framework. Knowledge-Based Systems. 8, 143--151 (1995)

6. Brennan, S. E.: The grounding problem in conversation with and through computers. In: Fussell, S. R., Kreuz, R. J. (eds.) Social and cognitive psychological approaches to interpersonal communication, pp. 201--225. Lawrence Erlbaum, Hillsdale (1998)

7. Clark, H. H., Schaefer, E. F.: Contributing to discourse. Cognitive Science. 13, 259--294 (1989)