Automotive Speech-Recognition - Success Conditions Beyond Recognition Rates

K. Bengler

BMW AG Knorrstr. 147 80788 Munich Germany

klaus.bengler@bmw.de

Abstract

From a car-manufacturer's point of view it is very important to integrate evaluation procedures into the MMI development process. Focusing the usability evaluation of speech-input and speech-output systems aspects beyond recognition rates must be fulfilled. Two of these conditions will be discussed based upon user studies conducted in 1999:

Mental-workload and distraction

Learnability

Introduction

In general we can observe an increase of speech technology in everyday live. Especially telephone services and office-functionality get more and more supported by automatic speech recognition (ASR). Also - but time-lagged - the number of voice controlled in-car devices increases rapidly. Existing examples are BMW's speech operated telephone, navigation and dictation system and the DaimlerChrysler Command-System. Both systems are speaker independent, controlled via simple command words but only partially robust against ambient noise.

Compared to the technical status quo the users' expectations towards the future of speech products are huge. They expect that the availability of a speech HMI will lead to a perceivable reduction of distraction from the driving task and an increase of usability. This premise can be hold if we know more about the interaction between driving and speaking and the possibilities to design in-car dialogs for speech driven application; especially for the correction of recognition errors dealing with an error prone input modality.

But databases and HMI approaches done for office and phone applications have only limited benefit for the automotive scenario. Projects like SpeechDat-Car "Speech Databases recorded in Vehicles" (EU-Project LE4-8334) contribute to enhance recognition rate of in-car ASR via enhanced training databases. The project will deliver a set of uniform, coherent databases for nine European languages. The project frame is described by v.d. Heuvel et al. (1999).

What about the HMI aspects ? Noyes & Frankish (1994) observe that rate of errors and subsequent error correction is - of course - an important item influencing the efficiency of a HMI and must be regarded from the beginning of the system design process. This fact is no longer trivial as the correlation between error frequency and efficiency is non-linear as error correction involves new error possibilities

In other scenarios we are able to identify and transfer basic dialogue requirements for qualification, navigation/control and general dialogue structure.

Proposed time-outs for dialogues in a phone session environment or even office applications - like 4-10 sec for choosing a menu item - might not hold for automotive applications.

Because of the coexistence of a driving (primary) task and a speech production task (secondary) we have to regard further items. Therefore specific automotive requirements must be added like facilities for user motivated interrupts or multimodal error recovery. Error recovery plays a specific part in this context and drags drivers' attention by increased workload.

Mental-workload and distraction while driving

In the case of automotive applications mentalworkload and distraction in different traffic situations plays an important role. In a field trial experiment conducted in 1999 at the Technical University of Munich Schweigert (1999) conducted an experiment on driver behaviour while working on different secondary task. The study examined the validity of eyemovement behaviour, driving behaviour and further data for the measurement of compatibility of seondary tasks with the driving task.

In the study subjects had to solve a listening task while managing several traffic situations. The situations differed in their distraction potential and were distinguished in "Highway", "Suburban" and "inner city". Single words ("Horse", "Tree" ...) that were presented acoustically during different driving situations (motorway, suburban road, inner city) had to be classified ("Is this an animal name,,) within one second by pressing a button at the steering wheel ("Listening task"). In the control condition the subjects had no additional task. In both conditions they were instructed to regard traffic safety at any time. The measurement of eyemovement behaviour was done with the JANUS system developed by the Technical University of Munich.

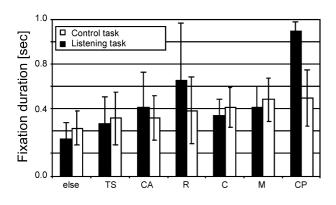
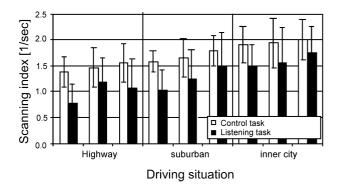
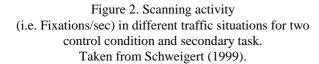


Figure 1. Mean duration of fixations on different objects for two experimental conditions.

Schweigert (1999). (TS=traffic signs; CA=approaching car; R=road; C=cockpit; M=mirror; CP=preceding car)

Figure 1 shows that during the listening task mean duration of fixations is significantly increased for specific objects ("Preceding car", "Road"). Other objects are not influenced or fixated for slightly shorter periods. This might indicate that mental processing has an impact on the processing of surrounding information.





An analysis of the effect across different driving situations shows that scanning activity is decreased but in different situation for a different percentage. Depending on the demands of the actual driving situation drivers use specific strategies to shift attention between both tasks and to compensate increased workload in oder to guarantee traffic safety. In condition "inner city" the reduction is smaller than on the "highway" that allows better prediction. In case of mental operation on the listening task reduced visual activity and scanning can be observed; in case of a shift to the driving task the error frequency in the listening task is increased.

Summing up - the analysis of eye movement data shows that the operation of the listening task has some influence but no severe impact on driving behaviour. The results are in line with the well known fact that visual behaviour is strongly related to mental load as described in Recarte et al (1999): Mental load results in reduced variability and longer fixation times.

The data illustrate that it is absolutely necessary that in-car speech-recognition has to deal with compensation effects in demanding traffic situations. Transferred to higher level dialogues this means that the HMI has to manage:

- extended driver reaction time on system prompts
- increase of syntax and vocabulary neglections
- disorientation about system state in dialogue sequences

Learning & Errors

The learnability of speech systems and their dialogue concepts is a remarkable indicator for suitability for in-car use while driving and important for customer-acceptance. Different measures help to improve the learnability of a speech driven HMI:

- vocabulary and command-syntax are intuitive
- the negative consequences of beginner-errors are limited
- special support is given to novice users
- behaviour. of novice users can be used as an indicator

For error handling and support actions a scheme for error classification is helpful. This structure can be used to activate system behaviour in relation to the dialogue state if man-machine interaction was erroneous (e.g. for tailored help prompts).

In an evaluation sense the structure can be used to calculate the benefits of dialogue improvements and the overall quality of an HMI concept.

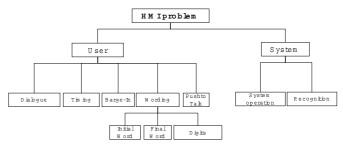


Figure 3. Error taxonomy. For reasons of readability the image is reduced to the first three layers of description

Roughly spoken this taxonomy distinguishes between "system based" and "user based" error situations. The value of this classification shall be shown referencing the data of a user study on a prototypical system for voice dialling done by Niedermaier (1999).

A remarkable frequency of errors of type Timing: "spoken too late,, and Wording: "wrong initial command word,, led to misrecognitions. The users were unfamiliar and unsure about what they had to do next. To risen system transparency and enhance learning on the system the dialogue was modified as follows:

"If user	input is	delayed	for more	than x
seconds,	the user	shall be	prompted	by the
system	with	а	help	text."

The modified dialogue was evaluated again using the same taxonomy. The frequency of errors type Timing: "spoken too late" and Wording: "wrong initial command word" could be reduced by 75%. Whereby the users responded quicker on the help texts than on other dialogue prompts acts and overrided the system. This resulted in an increase of error type Timing: "spoken too early"

This example shows that errors in speech driven human machine interaction represent an informative resource for optimisation. But a structured approach is needed to be able to exploit this resource and reach real improvements instead of a shift of error frequencies.

Conclusion

Two studies show - as an example - that not only recognition rate but especially dialogue design is necessary to rise the positive contribution of in-car ASR. The goal is to reach commandments for in-car dialogues like proposed by Krahmer et al. (1997) for natural language dialogues. In-car speech recognition - in comparison to other environments - has to be understood as a scenario in which the priority of the dialogue task is reduced by the driving task. New features of recognition engines could help to bridge dialogue situations in which the speech dialogue is neglected.

It has been shown that a structured error-driven approach is suitable to enhance the design and evaluation of usable HMI-concepts. For sure the given taxonomy is not complete and still implies - unproved - independence of disjunct error types.

In the text just eye-movement behaviour was mentioned as an indicator of mental workload. Within the studies other parameters like steering behaviour and driving speed were evaluated as further measures. The goal must be to reach a set of well-known valid parameters for the evaluation of compatibility of dialogue with different driving situations.

References

- v.d. Heuvel, H., Boudy, J., Comeyne, J., Euler, S, Moreno, A., Richard, G. (1999). The Speechdat-car Multilingual Speech Databases for In-Car Applications: Some First Validation Results. Eurospeech Budapest.
- Krahmer, E., Landsbergen, J, Pouteau, X. (1997): How to Obey the 7 Commandments for Natural Language Dialogue ?. Manuscript no.1190. Conf. "Applied Natural Language Processing", 1997, Washington, USA.
- Niedermaier, B. (1999). Entwicklung und Bewertung eines nutzerorientierten Dialogkonzepts zur Sprachbedienung eines Autotelefons. Diploma. Technical University of Munich.
- Noyes, J.M. & Frankish, C.R. (1994). Errors and Error Correction in Automatic Speech Recognition Systems. In *Ergonomics* 37,11 (pp. 1943-1957).
- Schweigert, M. (1999). Fahrerverhalten beim Führen eines Kraftfahrzeuges unter gleichzeitiger Bearbeitung von Zusatzaufgaben. MOTIV MMI AP8: Project Deliverable. Technical University of Munich.