Dialogue systems when the dialogue is just a secondary task – some preliminaries to the development of in-car dialogue systems

Nils Dahlbäck, Arne Jönsson
Department of Computer and Information Science
Linköpings universitet, SE-5818 83 Linköping, SWEDEN
nilda@ida.liu.se, arnjo@ida.liu.se

Introduction

Work on dialogue systems has gone a long way, since the first attempts in the 70'ies. Today we have a variety of commercial dialogue systems utilising simple, but robust, dialogue and a large number of research systems demonstrating sophisticated human-computer dialogue; CoreSong (Wärenstål, *et.al*. 2007), Higgins (Skantze *et.al*., 2006) and GoDiS/DICO (Larsson & Villing, 2007) are three examples of (Swedish) dialogue research systems.

But what is common to basically all current systems, both those mentioned above and others, is that they are based on the assumption that the dialogue is the primary or only task. This is true for Q/A systems, information retrieval, advisory systems, tutoring systems etc. Not in the sense that the ongoing dialogue is the only task at hand, but at the time of engaging in the dialogue there is normally no other task that competes for the attention.

Recently, however, exceptions to this rule are emerging, when the speech interaction is subservient to another concurrently on-going task. An example of this, and the one we will discuss below, is spoken dialogue interaction with

modern IT-based systems for drivers of cars, both so called ADAS (Adaptive Driver Assistance Systems) and navigation and entertainment systems. We know of no existing such systems today that are dialogue systems in the strict sense of the word. Today speech inputs from drivers to cars are basically systems for using speech to operate the equipment, e.g. turning on the radio or making a phone call without taking the hands off the steering wheel (Becker *et.al*. 2006b). And also current research systems under development that are aiming for in-car dialogue systems in our view do not to a sufficient extent consider the specific characteristics of the dialogue and task when one of the interlocutors is the driver of a car. For instance, based the lessons learned from the work on the SAMMIE system (*ibid*.) the following six issues are considered the biggest challenges to natural, intuitive multimodal interaction: "Reliable and robust ASR", "Natural and understandable speech synthesis", "System responsiveness", "Close to real time system feedback", Speech-centered multimodality", "Adaptive, context-sensitive presentation" (Becker *et.al*. 2006). There is no denying that these are important aspects of dialogue systems. However, note that none of the six issues mentioned consider aspects related to the fact that the driver might suddenly have more important things at hand than which music to play from a list of 150.000 albums available.

To exemplify with a dialogue system from our own lab developed for a similar task, consider the dialogue in Figure 1 presenting a conversational recommender system that helps users set up a play list for various situations, e.g. for composing playlists for music to listen to when driving (Wärenstål, *et.al*. 2007). It is, of course, also possible to set up play lists for various driving situations, such as, driving on a German Autobahn compared to driving on Highway One in California.

(1) U: Create a new drive playlist please.
(2) S: Ok. The play list Drive is active.
(3) Would you like to provide one of your favourite genres? You can add more genres later.
(4) U: Yes, I like techno.
(5) S: Ok. Would you like to provide one of your favorite artists?
(6) U: The artist does not matter.
(7) S: Would you like to provide one of your favorite albums?
(8) U: I like the album BodyCombat27.
(9) S: To refine the current playlist, please give more genre, artist or album preferences. Say "done" when you are ready.
(10) U: Start player
(11) S: Ok. *(Starts playing the song with the highest score)*
(12) U: What genres are on the album Odyssey?
(13) S: The genres on album Odyssey are: Electronic.
(14) Tell me more, or say "done" if you like the playlist.
(15) U: I think I like the genre electronic as well.
(16) S: Ok. Say "done" when you are ready.
(17) U: This song is good when I work.
(18) S: Ok. Created the playlist Work.
(19) Do you want to switch to playlist Work?

Figure 1 Example of dialogue with a conversational recommender system

Figure 1 displays a variety of expressions typical for dialogue in recommender systems, such as interviewing the user before delivering a recommendation (Wärenstål, *et.al*. 2007), the use of preferences *I like the album BodyCombat27,* hedging *I think I like the genre electronica as well,* set shifts (Allen *et.al*. 2000) (17)-(19) and switching between task, in this case between asking for information and discussing the playlist(s).

Ideally, such systems should be used before starting to drive, but will, of course, be used while driving as well. This change in the environments and situations in which spoken dialogue systems will be used in the future brings about at least two important changes. First, the dialogue manager will need to be able to accommodate to interruptions of new kinds, and also in other ways adapt the interaction to the traffic situation and the driver's occupation with the primary task of driving the car. This includes, but also extends beyond advanced personalisation models (Thompson *et.al*. 2004). Second, we need new methods and metrics for analyzing and evaluating the systems, since we, in addition to considering how well they perform the dialogue as such, also need to consider how they perform with respect to the driver's entire operating environment and on-going tasks.

Below we will discuss briefly these two issues, and give some suggestions for aspects that we believe will need to be considered in future work in this area. What we are providing here is only some preliminary first thoughts on some of the unique features of these systems which we believe make them an interesting challenge for future work in the field.

Dialogue situation characteristics

Informal observations suggest that passengers adapt the content and the timing of their dialogue contributions and turn taking to the traffic situation and the driver's varying need to concentrate on the driving of the car. Or, perhaps more correct, this is something that passengers that also are drivers do, whereas at least some passengers that do not have any own experience of driving a car will sometimes engage in a conversation at for the driver extremely ill-chosen moments. This means that in-car dialogue systems ideally need to be able to adapt to the current traffic situation and adapt the dialogue to it. While this

might seem like an almost science fiction inspired vision, given the well known difficulties with adaptive and so-called intelligent interfaces and systems (Höök, 2000), we believe that the increasing use of sensor information for assessing the car's situation for the use of modern safety features like preparing the safety belt and other similar systems for a possible oncoming collision, might possible also be used for this purpose, though in all likelihood this will require the development of systems beyond current state of the art to be used in adapting the car's dialogue system to the present traffic situation. This is, however, something that will depend more on the development of in-car information technologies not directly related to dialogue systems, and therefore something that we will not address further here.

But even if the car cannot adapt to the current driving situation and the driver's current tasks, the driver/speaker most surely will do so. Here we need to collect dialogue corpora on exactly how speakers that also are drivers actually manage this situation from a dialogue management point of view. That we will have new forms of interruptions etc. is not exactly a daring hypothesis. But exactly how should they be diagnosed and treated by the system? And perhaps even more interesting from a communication research perspective, how will the dialogue be re-activated and anaphoric and other dialogue relations be re-established? Will we be able to distinguish between interruptions caused by the speaker losing interest and simply dropping the subject, from interruptions caused by the driver needing to focus completely on the driving of the car because of an imminent danger?

This far we have treated the dialogue between the driver and the car as if it were an either or situation; either the conversation flows naturally or it is interrupted. But in all probability there will be intermediate situations, where for instance the speech output from the system needs to be slowed down. But also we believe

that in some situations which require some extra attention from the driver without making him/her completely drop the on-going dialogue, both the speech output and the dialogue structure might be affected. Previous research on hesitations, false starts and filled pauses in dialogue have related theses phenomena to the speech production process, and how listeners make use of these signals in the comprehension process (e.g. Clark & Fox Tree 2002, Fox Tree 2001). But we are not aware of any research that has analyzed the changes in speaker output caused by him/her simultaneously attending to another task. Similarly, the phenomenon of hypercorrections of speech causing problems for automatic speech analysis are well known. But what happens to the speech output when the speaker is speaking slower because of a split attention between the task being performed in the dialogue and another concurrent task?

The use of Wizard of Oz-experiments (Dahlbäck *et.al*. 1993), often used for dialogue system development, is today, also used for research on multimodal dialogue in cars like e.g. (Becker *et.al*. 2006b). In this project experiments were conducted in a one user simulator where the Wizard simulated the entertainment system. We believe that such experiments are important, but see a need for much more complex simulations as well as other developments in the Wizard of Oz studies in this domain.

Realistic situations are necessary in order to acquire ecologically valid data. However, simulations will play an important role and the simulation setup must consequently be as realistic as possible. Ideally the traffic situation should allow more than one driver, i.e. many subjects driving in the same simulated world at the same time. This can then be complemented with artificial cars, i.e. cars having a ”normal” driving behavior, to further distract the subjects.

Results from simulator studies

We will here present some results from a series of as yet unpublished driver simulator studies of voice interaction in cars by Ing-Marie Jonsson, associated with our group in Linköping. All these are based on a similar methodology. Participants are driving in a simulator while at the same time receiving spoken information on the traffic situation etc. from the car's speech system, in some cases also engaging in a spoken dialogue with the system while driving.

In one study the linguistic complexity of the utterances and the complexity of the driving task were varied. The task was to book flights using an "in-car information system". Not surprisingly, when the driving situation is simple, drivers can handle complex utterances, but not so when the driving situation is hard. But the solution is not here to always use simple sentence structures, since the results also show that in easy driving situations complex sentences are better understood and remembered, and also the driving is safer. This suggests to us that it is important for future in-car dialogue systems to be able to adapt its linguistic output to variations in the traffic situation.

Another study compared interrupting or non-interrupting speech information systems for drivers in a situation creating a high cognitive load. The results showed that interrupts affects both driving performance and attitude towards the system negatively. It was also shown that interrupts had more profound negative impacts on young drivers (18-25) than on older drivers (55+). These results point in the same direction as the previous study, i.e. underscoring the importance of the car's dialogue system being able to adapt to the drivers on-going task, in this case by not engaging in a dialogue when the driver is occupied with the driving task.

It is now a well established finding that users of speech interfaces exhibit the same range of social responses to these artificial interlocutors as they do in conversations with other people (Nass & Brave, 2005, Dahlbäck *et.al*. 2007). That this will be the case also for in-car systems has been shown in a number of studies. One example is (Nass *et.al*. 2005), where it was shown that when user emotion matched car voice emotion (happy/energetic and upset/subdued) drivers had fewer accidents, attended more to the road, and spoke more to the car. But much more work is required here to guide the development of useful and safe in-car speech dialogue systems.

Methods for analysis and evaluation of in-car systems

One of the most prominent evaluation methods of dialogue systems today is PARADISE (Walker *et.al*. 1997). PARADISE uses a combination of performance efficiency and user satisfaction metrics to evaluate (spoken) dialogue systems. As most evaluation methods, PARADISE evaluates the human-artefact interaction, which is sufficient for most dialogue systems' applications and situations. It is however not sufficient when evaluating dialogue systems in cars and other situations where the user of the system is also attending other tasks, especially when these have a higher priority such as is the case with in-car dialogue systems. Then the entire system, comprising of the driver and the car with all its equipment needs to be evaluated too.

While human factors aspects of driving is not exactly a new field (for an early example see e.g. (Gibson & Crooks, 1938), the recent development of advanced driver support systems and active safety functions have significantly changed the nature of driving (Hollnagel *et.al*. 2003). Our suggestion is that we here need to combine two approaches. First, analytic tools from research which views the operator and the equipment used not as two separate entities, but as a so-called

Joint Cognitive System (JCS) (Hollnagel & Woods, 2005), and expand this to not only include aspects directly related to the driving and the driving situation, but also to the interaction with the new support and entertainment systems in the cars of today and even more so in the cars of tomorrow. Second, we need to expand current models of dialogue management (e.g. Allwood, 1995) to also include the interaction and possible interference with other concurrent tasks in e.g. turn management. It is our belief that we here need to study especially management of interruptions and how interlocutors re-connect and re-establish both dialogue and dialogue task structure using for instance further developed approaches like those we have previously described in (Dahlbäck & Jönsson, 1999). We probably need both to develop new task models and to gain a deeper understanding of how the interlocutors re-establish the interrupted task.

Conclusions

In this paper we have presented some of the interesting research challenges which we believe will emerge when trying to develop speech dialogue systems for applications where the spoken dialogue is not the primary task. We have taken in-car dialogue systems as an example and presented some early results from work in this area. If nothing else, we hope we have been able to show that much further work is needed before we fully understand the properties of dialogue systems when dialogue is not the primary task. To conclude, we want to stress our firm belief that progress in the development of dialogue systems for in-car systems need to proceed along the same path that have been used in other application domains, i.e. through a combined effort of theoretical analysis, empirical work on understanding the language used in the specific situation, system development and evaluation of these systems.

References

Allen, J. Byron, M., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2000) An Architecture for a Generic Dialogue Shell, Natural *Language Engineering*.

Allwood, J. (1995) Reasons for Management in Spoken Dialogue. In Beun, R.J., Baker, M. & Reineer, M. (eds)., *Dialogue and Instruction*, Springer Verlag, pp. 251-260.

Becker, T., Blaylock, N., Gerstenberger, C., Kruijff-Korbayová, I., Krothaur, A., Pinkal, M., Pitz, M., Poller, P. & Schel, J. (2006) Natural and Intuitive Multimodal Dialogue for In-Car Applications: The SAMMIE System. In In Proceedings of the ECAI Sub-Conference on Prestigious Applications of Intelligent Systems (PAIS 2006), Riva del Garda, Italy.

Becker, T., Poller, P., Schehl, J., Blaylock, N., Gerstenberger, C., & Kruijff-Korbayová, I. (2006b) The SAMMIE System: Multimodal In-Car Dialogue, *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, pages 57–60.

Clark, H.H. & Fox Tree, J.E. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition*, 22, 1-39

Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993) *Knowledge-Based Systems,* Vol. 6, No. 4, pp. 258-266.
Reprinted in Mark T. Maybury and Wolfgang Wahlster (1998) *Readings in Intelligent User Interfaces* Morgan Kaufmann Publishers.

Dahlbäck, N. & Jönsson, A. (1999) Knowledge Sources In Spoken Dialogue Systems, in *Proceedings of Eurospeech'99*, Budapest, Hungary.

Dahlbäck, N, Wang, Q., Nass, C & Alwin, J (2007) Similarity is more important than expertise: Accent effects in speech interfaces. In *Proceedings of CHI*.

Fox Tree, J.E. (2001) Listeners' use of um and uh in speech comprehension. *Memory and Cognition*, 29, 320-326.

Gibson, J.J. & Crooks, L.E. (1938) A theoretical field-analysis of automobile-driving. *The American Journal of Psychology*, LI, 453-471.

Hollnagel, E., Nåbo, A. & Lau, I. (2003) A Systemic Model for Driver-In-Control. In *Porceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 86 – 91.

Hollnagel, E. & Woods, D. (2005) *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, London, Taylor & Francis

Höök, K. (2000) Steps to Take Before Intelligent User Interfaces Become Real. *Interacting with Computers* (4) 409 – 426.

Larsson, S. & Jessica Villing, J. (2007) The DICO project: A Multimodal Menu-based In-vehicle Dialogue System. In Bunt, H.C., and Thijsse, E. C. G. (eds): *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)*.

Nass, C., Jonsson, I-M., Harris, H., Reaves, B., Brave, S & Takayama, L. (2005) Improving automotive safety by pairing driver emotion and car voice emotion. In *Proceedings of CHI 2005*.

Nass, C., & Brave, S. (2005) *Wired for Speech* the MIT Press, Cambridge, Mass.

Skantze, G., Edlund, J., & Carlson, R. (2006) Talking with Higgins: Research challenges in a spoken dialogue system. In André, E., Dybkjaer, L., Minker, W., Neumann, H., & Weber, M. (Eds.), *Proceedings of Perception and Interactive Technologies* (pp. 193-196). Springer Verlag.

Thompson, C. A., Göker, M., & Langley, P. (2004) A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research*, 21, 393-428.

Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1997) PARADISE: A Framework for Evaluating Spoken Dialogue Agents, *Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics , ACL 97*.

Wärnestål, P., Degerstedt, L., & Jönsson, A. (2007) Interview and Delivery: Dialogue Strategies for Conversational Recommender Systems, Proceedings of 16th Nordic Conference of Computational Linguistics (Nodalida). Tartu, Estonia.