Multimodal Dialog in the Car: Combining Speech and Turn-And-Push Dial to Control Comfort Functions

Sandro Castronovo[1], *Angela Mahr*[1], *Margarita Pentcheva*[2], *Christian Müller*[1]

[1] German Research Center for Artificial Intelligence
{angela.mahr, sandro.castronovo, christian.mueller}@dfki.de
[2] Computational Linguistics, University of the Saarland, Germany

Abstract

In this paper, we address the question how speech and tangible interfaces can be combined in order to provide effective multimodal interaction in vehicles, taking into account the special requirements induced by the circumstances of driving. Speech is used to set the interaction context (determine the object as is to be manipulated) and a turn-and-push dial is used to manipulate/adjust. An experimental study is presented that measures the distraction induced by manual (conventional), speech-only, and multimodal interaction (combination of speech and turn-and-push dial). Results show that while subjects where able to perform more tasks in the manual condition, their driving was significantly safer while using speech-only or multimodal dialog. Supplemental contributions of this paper are descriptions of how a multimodal dialog manager as well as a driving simulation software are connected to the CAN (Controller Area Network) vehicle bus as well as how driver distraction caused by interacting with a system are measured using the standardized lane change task (LCT)

Index Terms: speech recognition, automotive, driver distraction, multimodal interaction, lange-change task, haptic controls

1. Introduction

In recent years, the complexity of on-board and accessory devices, infotainment services, and driver assistance systems in cars has experienced an enormous increase. A current premium-class car already implements hundreds of functions that a user can interact with and these numbers are likely to even grow in the next generation [1]). This development emphasizes the need for new concepts for advanced human-machine interfaces that support the intuitive and efficient use of this large variety of devices and services. To cope with this problem, car manufacturers introduced a rotary device, the so called "turn-and-push dial", located in the center console of the car. The idea was, that a small number of control elements really crucial to driving should be positioned around and on the steering wheel. Most other functions (entertainment, comfort, multimedia and navigation systems) are controlled through the new device. It was designed to enable intuitive operation with a control display in the upper middle section of the instrument panel. The dial can be used with just one hand, and offers haptic feedback so it can be used without looking at it. However, in practice, the turn-and-push dial turned out to be less effective than expected. Motoring organizations complained, the dial would require a lot of practice, especially from people who are less familiar with the interaction with computers. Furthermore, it would be almost impossible to use while driving, because the screen has to be watched constantly [2]. [3] even stated: "it manages to complicate simple functions beyond belief" .

Speech has been proposed as a means of interaction especially suitable for in-car use: 1. it is intuitive and 2. it allows the driver to keep the eyes on the road and the hands on the steering wheel [4, 5, 6]. Consequently, the usage of speech dialog systems of car has evolved from research prototypes into mass production today. However, speech recognition is still error-prone especially in noisy environments. Varying driving noise at different speeds especially affect the recognition performance. Although commercial speech recognition suppliers worked hard on that problem in recent years, this remains to be an issue. Moreover, speech is not the most intuitive means of interaction in every case. When it comes to operations that involve a continuous ranges like opening the window "a little" or adapting the volume of the stereo, manual interaction has obvious benefits.

In this paper, we address the following research question: How can speech and tangible interfaces be combined in order to provide effective multimodal interaction in vehicles taking in to account the special requirements induced by the circumstances of driving? This is done along the lines described in previous research, for example by [7, 8, 9].

2. Multimodal Interaction Based on CAN Bus

CAN (Controller Area Network) is a vehicle bus standard for connecting electronic control units in the car. It is used to connect engine control unit and transmission or, on a different bus, to connect the comfort functions (window lifter, climate control, seat control, etc). CAN is a standard bus in the automotive industry since the late Nineteen Eighties [10]. We connected our multimodal dialog system to bus in order to be able to assess the statuses of the connected electronic control units and manipulate them. It has to be emphasized that we work with a non road-going car. We run the system stationary and usability testing is done using a drive simulation software rear-projected to the windshield (see below).

Figure 1 depicts a simplified schematic diagram of sensors, actuators, and systems components connected to the CAN bus. Microphone and turn-and-push dial function as sensors for the multimodal dialog manager, which is connected via a CAN interface. The dialog manager interprets user input and triggers the respective actions. Actuators analyzed in this study are the following comfort functions: window lifters, climate control, seat heating, fans, and entertainment system. Other electronic control units can be connected as well (e.g. sensors like thermometer, GPS, accelerometer, etc. as well al actuators like rear view mirrors, windshield wipers, cruise control, and so on). In

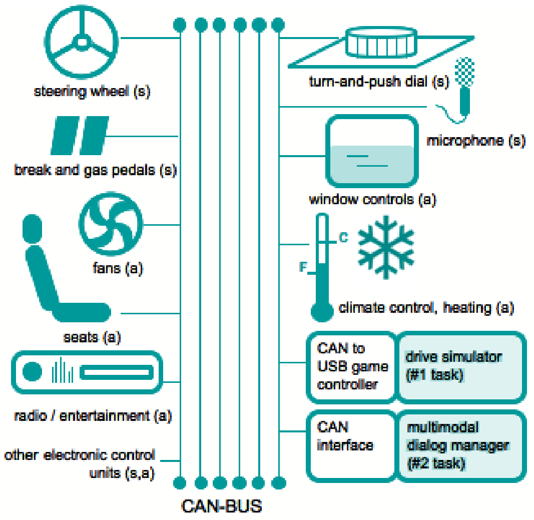
steering wheel (s)
break and gas pedals (s)
fans (a)
seats (a)
radio / entertainment (a)
other electronic control units (s,a)
CAN-BUS
turn-and-push dial (s)
microphone (s)
window controls (a)
C
F
climate control, heating (a)
CAN to USB game controller
drive simulator (#1 task)
CAN interface
multimodal dialog manager (#2 task)

Figure 1: Various sensors (s) and actuators (a) connected to the CAN bus. Main system components are depicted at the lower right. They also represent primary respectively secondary tasks in the study. Actuators analyzed in this study are the following comfort functions: window lifters, climate control, seat heating, fans, and entertainment system.

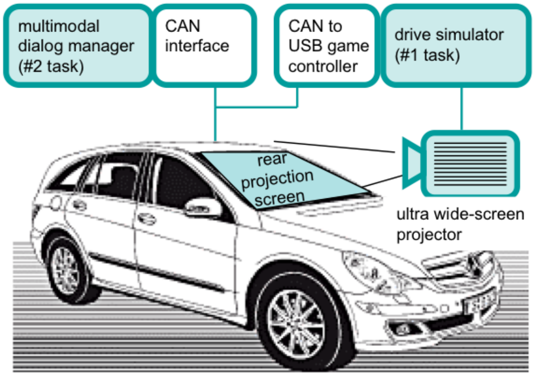
multimodal dialog manager (#2 task)
CAN interface
CAN to USB game controller
drive simulator (#1 task)
rear projection screen
ultra wide-screen projector

Figure 2: The technical setup of the experiment. A driving simulator software is rear-projected onto the windshield of a stationary car.

order to read sensor information and trigger actions, the multimodal dialog manager uses a CAN interface. The current version of this interface works with a Mercedes R-Class vehicle and was developed in collaboration with Daimler. Steering wheel and brake pedal function as sensors for the drive simulation, which is used to measure the distraction induced by interacting with the system. Since the gas pedal has no connection to the CAN bus, a button on the steering wheel was used for accelerating in the study presented here. A standardized drive simulation software was used (see below) that supports USB game controllers like joysticks and steering wheels for games. The software was connected via a CAN-to-USB interface developed at our lab for this purpose. We used ODP, the Ontology-based Dialogue Platform developed at DFKI. OPD is a generic platform for the realisation of multimodal dialogue systems [11].

3. Experimental Study

Apparatus We measured the driver distraction using the standardized ”lane change task” (LCT) [12], a simple laboratory dual-task method that is intended to estimate driver distraction, which results from using in-vehicle devices such as mobile phones, navigation systems, or – like in our case – the car comfort functions. LCT was developed by Daimler and BMW and is currently in the process of becoming an ISO standardized tool (ISO Draft International Standard 26022). As stated in [12], it has already been successfully used in a large number of relevant studies. LCT can be regarded as a simple driving simulation: the simulated driving task (driving activity on a roadway with three lanes) resembles the visual, cognitive and motor demands of driving. The test involves visual stimuli (pop up signs) and responses (initiating a lane change maneuver). More precisely, the maneuver represents a lateral displacement from the current lane to a parallel lane, while it is possible that one lane is crossed. Signs that instruct the subject to change into a specific lane appear at regular intervals on both sides of the road. The speed is kept constant at 60 km/h (37 mph). As a performance measure, LCT calculates the mean deviation of the lane between a normative model and the actual driving along the track. The performance of the baseline (only driving) is then compared with driving with a secondary task in order to objectively assess the level of distraction induced by that activity. In our experiment, the LCT screen was rear-projected to the windshield of the car using an ultra-wide-screen projector that was mounted on top of the hood (see Figure 2). The software was connected via a CAN-to-USB interface as described above.

24 subjects (11 men and 13 women), were paid to participate in the study. The age range was between 21 and 60 with an average age of 35 years. None of the subjects ever drove a Mercedes R-Class before. Hence, subjects were not familiar with this particular layout of standard controls or the central multifunctional device. The range in driving experience was between 3 and 40 years (average 15,54) measured in years since when the participants possess their drivers licenses. The mileage was: 58 % of the participants drive up to 10 000 kilometers (6 200 miles) a year; 38 % drive between 10 000 and 25 000 kilometers (6 200 and 15 500 miles) a year; and 4 % drive more than 25 000 kilometers (15 500 miles) a year. Before the experiment, each participant was advised that their information would be confidential.

Procedure The entire experiment took about one hour to complete. During the test, the experimenter was sitting in the passenger seat, controlling the equipment and reading the instructions aloud from a manuscript. A second experimenter was sitting in the back prompting commands (see below). The experiment started with a brief explanation of the entire procedure followed by a warm-up with the driving simulator. The motivating scenario was a car rental at the airport. Hence, the functions were introduced briefly but no training was allowed. The baseline (driving performance without secondary task) was measured afterwards, followed by the main part of the experiment. Here, the conditions manual, speech-only, and multimodal were presented in a balanced order (between subjects), accompanied by a respective questionnaire. Nine simulated routes (tracks) were completed, each track being 3 km (1.8 Miles) long, which corresponded to a duration of approximately 3 minutes. Except for the sequence of signs, the tracks had no visual differences in order to minimize learning effects. Type and number of the signs on each route were kept constant (three changes from left lane to right lane, three from center lane to left lane, etc.). In total, 18 lane changes per track were performed, which were triggered every 150 m. The participants were instructed to change

modality	explanation	example
manual	using the standard knobs, levers, and switches	subjects opens the rear left window by pressing the respective button integrated into the casing of the driver's door; subject increases the volume of the car stereo by turning the respective knob at the very device.
speech-only	using only speech	subject says ”Open the rear left window”, ”Increase the volume”.
multimodal	using speech to specify the context (object) and the turn-and-push dial to manipulate/adjust it	subject says ”Rear left window” – system confirms ”The rear left window can now be controlled” – subject pulls down the dial; subject says ”Volume” – system confirms ”The volume can now be controlled” – subject turns the dial clockwise.

Table 1: Overview over the interaction modality variants (experimental conditions). Besides the example in the right-most column, the following function were **also used in the study**: other window lifters, climate control, seat heating, fans, and entertainment system.

lanes fast but controlled while at the same time solve as many of the secondary tasks as possible (neither of them should be preferred). The prompts to perform the secondary task were designed to provide the wording of the corresponding commands to the least extend possible. For example, instead of prompting ”Open the rear left window” the experimenter (sitting in the rear) would say ”Could you please let in some fresh air here?”. The prompts were the same in all conditions only the order was random.

Immediately after completion of the driving trial in the respective condition, the drivers were given a Driver Activity Load Index (DALI) questionnaire [13], derived from NASA TLX, assessing the subjective demands in the following standardized categories:1) global attention demand: mental, visual and auditory demand required to complete the task; 2) visual demand only; 3) auditory demand only; 4) tactile demands: originally related to vibrations but here adapted to manual handling (there were no vibrations of any sort); 5) stress: fatigue, insecure feeling, irritation, discouragement, etc.; 6) temporal demand: pressure and specific stress felt due to timing; 7) interference: distraction of the driver induced by the secondary task. For each factor, the participants were asked to rate the level of demand felt during the session on a scale from 0 (low) to 5 (high) with regard to their usual driving.

After the experiment, the subjects were presented an additional (non-standardized) form with questions of general nature, such as regarding the difficulties, experiences with the driving simulator. Additionally, the participants had the opportunity to make free comments on the experiment.

Hypotheses The Hypotheses with respect to the **performance of driving (primary task)** compared to the baseline where: MANUAL << MULTIMODAL < SPEECH-ONLY,where $x < y$ means ”y allows a better performance than x” (”y induces less distraction than x”). Hence, it was hypothesized that 1. speech-only and multimodal interaction would induce less distraction than the ”conventional” manual interaction as the former two are believed to be more intuitive to the drivers (especially since they are unfamiliar with the car); 2. that speech-only interaction would outperform multimodal. We believed that this would be the case because speech-only allows complete hands-free interaction, which is not the case for multimodal. However, as indicated by the double <<, we also assumed that the difference between manual and multimodal would be greater than the difference between multimodal and speech-only. With respect to the **performance of the secondary task**, we hypothesized MANUAL << SPEECH-ONLY < MULTIMODAL. As stated in the introduction, we believe that speech interaction has clear advantage in the car. However, it is not always intuitive and it does not always allow precise settings (try adjusting your review mirror by using speech). Therefore we hypothesized that the multimodal interaction as we introduced it for this study outperforms speech-only in terms of task completion in the secondary task.

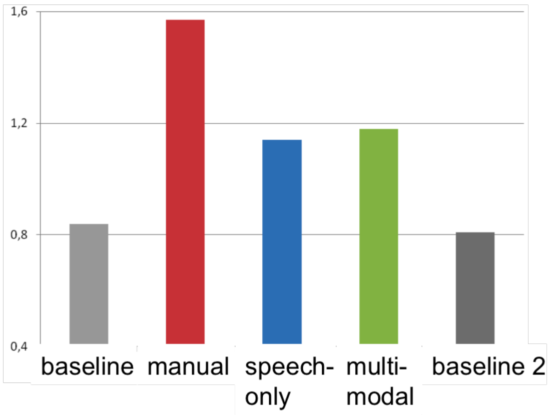

Figure 3: Primary task performance: mean deviation in meters between a normative model and the actual driving. Distraction in the manual condition was significantly higher than both in speech-only and multimodal. The difference between speech-only and multimodal is not statistically significant. However, the tendency is according to our hypothesis given above. The order of the colored bars does not correspond to the flow of the experiment as the order of the conditions was randomized.

4. Results

Figure 3 shows the mean deviation in meters between a normative model and the actual driving without secondary task (baseline 1, baseline 2), and with manual, speech-only, and multimodal interaction. The results confirm our hypotheses: the distraction in the manual condition was significantly higher than both in speech-only and multimodal ($p < .001$). We believe that this can be attributed to the fact that the interaction is eyes-free (for both speech-only and multimodal) and hands-free (for speech-only). Moreover, the turn-and-push dial has a clear advantage over the standard knobs. The difference between speech-only and multimodal is not statistically significant. However, the tendency is according to our hypothesis given above. The graph further shows that distraction without a secondary task (baselines) is lower than with a secondary task ($p < .001$). The difference between baseline 1 and baseline 2 is not significant which indicates a moderate learning effect. Note, that the order of the colored bars does not correspond to the flow of the experiment as the order of the conditions was

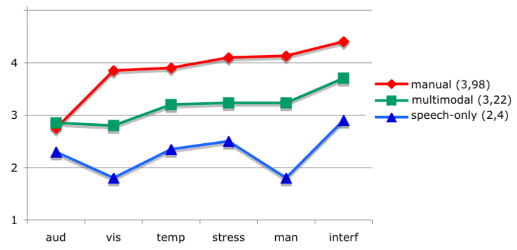

Figure 4: Subjective DALI rating by the subjects on the demand of the interaction depending on the modality. Manual is rated with the highest score, the least demanding is speech-only; multimodal lies in-between.

randomized.

Figure 4 illustrates the results of the subjective DALI ratings. Generally, the manual condition is rated with the highest score, indicating that it is perceived as most demanding F(1, 23)= 15,20, p<.01. The least demanding was the speech-only condition – multimodal is in-between. The categories were sorted according to the rating in the manual condition. It can be seen that the interference factor is the one which was rated the highest in all three conditions. Compared to the other two conditions, the visual demand in manual tasks received high ratings, which appeals to our intuition. The comparably low score of manual demand with speech-only is notable as well. Hypotheses are confirmed that speech is especially beneficial for hands-free interaction.

In terms of the relative number of completed tasks, the conditions rank $multimodal < speech - only < manual$. A task was marked as completed if the subject was able to perform the instruction. For example if the instruction was opening the rear left window a little, a "completed" mark was given if the window was opened eventually, even if the subject needed more than one trial or the window was opened more than a little. Hence, we look at a dissociation effect between the number of tasks completed and the driving performance. Novice drivers (remember that none of our subjects ever drove a Mercedes R-Class before) are obviously able to perform more tasks when using the standard knobs, levers, and switches (compared to speech-only or multimodal dialog). However, when they do this, their driving is less safe. A study investigating the learning effects would therefore be of much interest.

5. Conclusions

We proposed a specific approach on how to combine speech and tangible interaction in a car. Speech is used to set the interaction context (determine the object as is to be manipulated) and a turn-and-push dial is used to manipulate/adjust. An experimental study was presented that measures the distraction induced by manual (conventional), speech-only, and multimodal interaction (combination of speech and turn-and-push dial). Driver distraction in the manual condition was significantly higher than both in speech-only and multimodal. With respect to subjective ratings, manual is rated with the highest score, followed by multimodal followed by speech-only. Generally, results show that while subjects where able to perform more tasks in the manual condition, their driving was significantly safer with using speech-only or multimodal dialog.

6. References

[1] Christian Müller-Bagehl and Peter Endt, Eds., *Infotainment/Telematik im Fahrzeug – Trends für die Serienentwicklung*., Haus der Technik Fachbuch 38. Expert, Renningen, 2004.

[2] Michael Burmester, Ralf Graf, Jürgen Hellbrück, and Meroth Ansgar, "Usability — Der Mensch im Fahrzeug," in *Infotainmentsysteme im Kraftfahrzeug*. Springer, 2008.

[3] James R. Healey, "Bmw 7 series loses gold on technical merit," *USA Today*, vol. 7, no. 2, 2002.

[4] R. Graham and C. Carter, "Comparison of speech input and manual control of in-car devices while on-the-move," *Personal and Ubiquitous Computing*, , no. 4, pp. 155–164, 2000.

[5] D. L. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway," *Human Factors*, vol. 43, no. 4, pp. 631–640, 2001.

[6] Omer Tsimhoni, Daniel Smith, and Paul Green, "Address entry while driving: Speech recognition versus a touch-screen keyboard," *Human Factors*, vol. 46, no. 4, pp. 600–610, 2004.

[7] R. Vilimek, T. Hempel, and B. Otto, "Roman vilimek, thomas hempel, birgit otto: Multimodal interfaces for in-vehicle applications," *Human Computer Interaction*, vol. 3, no. 216–224, 2007.

[8] Katharina Bachfischer, Moritz Neugebauer, Niels Pinkwart, and Tobias Brandt, "Modality management for multimodal human-machine interfaces," in *Reports on Distributed Measurement Systems*, Fernando Puente, Ed., pp. 155–170. Shaker Verlag, 2008.

[9] M Ablaßmeier, G. McGlaun, J. Gast, T. Poitschke, and G. Rigoll, "A robust, context-adaptive and multimodal search engine for efficient information retrieval in car environment," in *Proc. of HCI 2005, 11th Intern. Conf. on Human-Computer Interaction*, G. Salvendy, Ed., Las Vegas, NV, 2005, pp. 22–27.

[10] M. Farsi, K. Ratcliff, and M. Barbosa, "An overview of controller area network," *Computing and Control Engineering Journal*, vol. 10, no. 3, pp. 113–120, August 1999.

[11] Norbert Pfleger and Jan Schehl, "Development of advanced dialog systems with pate," in *Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech 2006 – ICSLP)*, 2006, pp. 1778 – 1781.

[12] Stefan Mattes, "The lane change task as a tool for driver distraction evaluation," in *Quality of Work and Products in Enterprises of the Future*, H. Strasser, H. Rausch, and H. Bubb, Eds. Ergonomia, 2003.

[13] A. Pauzié and G. Pachiaudi, "Subjective evaluation of the mental workload in the driving context," *Traffic and Transport Psychology*, pp. 173–182, 1997.

[14] S. R. Dixon, C. D. Wickens, and J. S. McCarley, "On the independence of compliance and reliance: Are automation false alarms worse than misses?," *Human Factors*, vol. 49, no. 4, pp. 564–572, 2007.

[15] J. Bliss, "An investigation of alarm related accidents and incidents in aviation," *Intern. Journal of Aviation Psychology*, vol. 13, pp. 249–268, 2003.