Spatial auditory interface for an embedded communication device in a car

Jaka Sodnik, Saso Tomazic University of Ljubljana, Slovenia jaka.sodnik@fe.uni-lj.si Christina Dicke, Mark Billinghurst HIT lab NZ, New Zealand christina.dicke@hitlabnz.org

Abstract

In this paper we evaluate the safety of the driver when using an embedded communication device while driving. As a part of our research, four different tasks were preformed with the device in order to evaluate the efficiency and safety of the drivers under three different conditions: one visual and two different auditory conditions. In the visual condition, various menu items were shown on a small LCD screen attached to the dashboard. In the auditory conditions, the same menu items were presented with spatial sounds distributed on a virtual ring around the user's head. The same custom-made interaction device attached to the steering wheel was used in all three conditions, enabling simple and safe interaction with the device while driving. The auditory interface proved to be as fast as the visual one, while at the same time enabling a significantly safer driving and higher satisfaction of the users. The measured workload also appeared to be lower when using the auditory interfaces.

1. Introduction

A car is no longer used merely for traveling and getting from one place to another, but also more and more as an office-on-the-go. Nowadays, cars are being equipped with new powerful computers functioning as navigation systems, music players, DVD players, communication devices, etc. In order to make use of all that functionality, a great amount of user attention is required. A typical interaction with such a device causes a significant amount of distraction from the driver's primary occupation - driving. Distraction is not only caused by physical stimuli through the sensual apparatus, but also by various cognitive sources, such as thought or emotional arousal [1][2]. Distraction from the primary task, i.e. driving the car, can reduce the driver's safety by degrading the vehicle control (speed maintenance, lane keeping, etc.) and object or event detection [3]. Apart from the visual (eyes-off-the-road), auditory and cognitive distraction (mind-off-the-road), mechanical causes can also lead to distraction. When reaching for objects inside the vehicle or otherwise shifting out of their normal sitting position, drivers can degrade their ability to react to various unexpected anomalies on the road [3][4].

With this in mind, the sound channel could be used as an alternative option for driver-vehicle interaction. Speech synthesis systems are often used with various navigation devices and speech recognition systems with mobile phones in cars. Sometimes they are combined with small screens on the dashboard.

In our study we used two auditory interfaces of different complexity to operate an embedded communication device while attending to a driving task. We reduced the mechanically and visually distracting events, so that we could focus on the influence of the secondary tasks of varying complexity (conducted with an auditory interface) on the primary driving task. We used spoken menu items to build the auditory interface, as they have proven to be very effective [5][6]. We also compared the auditory interface to the classic visual interface comprising of a small screen.

2. Related work

The auditory menu used in our experiment was based on a number of spatial sounds placed on a virtual ring around the user's head. The items on the ring represented all current options at the specific level of the hierarchical menu.

The principle of the hierarchical menu navigation in the use of spatial sound was also used by Crispien et. al. [7]. They designed an interface aligning both non-speech and speech audio cues in a ring rotating around the user's head. The items in the ring were manipulated by using 3D-poiting, hand gestures and speech recognition.

Similar spatialised auditory icons localized in the horizontal plane were also used by Brewster [8]. The user selected an arbitrary auditory icon with a hand gesture which triggered the corresponding event.

The Nomadic Radio was developed as a spatial audio framework for the wearable audio platform [9]. It included a system for notification about the current events: incoming e-mails, messages, calendar entries, etc. The items of the



menu were positioned around the listener's head in this case as well. The input interaction was based on voice command and tactile feedback.

The examples given in this section also use spatial sound for the interaction with various devices. However, so far no such interface has been tested or evaluated in a mobile environment (e. g. while driving a car or a simulator) and compared to a purely visual interface.

3. User study

The main goal of our user study was to evaluate the effectiveness of the acoustic interface in the interaction with a communication device in a car. The communication device had the functionality of a mobile phone (it enabled making phone calls and sending text messages) as well as an entertainment system (it also enabled listening to music, watching pictures, etc.). We were interested in the use of such a device while driving. Due to security reasons a car simulator was used instead of a real vehicle. The interaction with the device was based on a special custom made interaction device attached to the steering wheel in order to be used safely while driving. The car simulator, the device itself and the interaction device are described in detail in the following chapters.

Two different interfaces were compared in the user study, both of which represented the same hierarchical menu structure of the device. In the acoustic interface, all menu items were presented with spatial sounds coming from different pre-fixed positions in the simulator. Other sounds, such as the car engine, environment noise, etc., were non-spatial and were played through all speakers as a background noise. In the visual interface, all items of the menu were shown on a small LCD screen attached to the dashboard of the car. The evaluation of the two interfaces was made by observing the drivers while they were driving and performing different tasks with the communications device. The main parameters of the evaluation were:

- efficiency of the individual interface (the time required to finish an indiviual task)
- safety of the driving (penalty points were given for unsafe driving)
- perceived workload (reported by drivers)
- overall satisfaction of the test subjects (expressed through the modified Questionnaire for User Interface Satisfaction QUIS)

We expected the acoustic interface to be much safer than the visual one, since all interaction was based only on the acoustic channel. The visual channel could therefore be used for driving only, enabling a much lesser distraction of the drivers. On the other hand, the time required to finish different tasks was expected to be shorter when using visual interaction, since the visual communication channel offers a much greater bandwidth, therefore more information can be perceived at a certain time.

4. Experiment design

4.1. Car simulator

The experiment took place in a visualization room equipped with a large projection screen (2.4m x 1.8m) and 7.1 surround sound system (Creative GigaWorks S750). All sounds used in the experiment were played with Creative Sound Blaster X-Fi ExtremeMusic sound card and Creative OpeanAL sound library was used for spatial sound positioning [10]. OpenAL enables easy positioning of virtual sound sources in 3D space using the CMSS-3D surround sound technology on the X-Fi Creative sound card [11].



Figure 1. The car simulator consisting of a big projection screen, a steering wheel and a small LCD screen.

CMSS-3D creates eight individual sound channels using a multi-channel upmix process. Multiple-speaker configuration (7.1) was used instead of the headphones in order to enable drivers to also perceive the co-occurring auditory events (car engine, braking, environment noise, etc.).

The speakers in the simulator were positioned according to Dolby recommendations for 7.1 systems. The listener was positioned in the sweet spot in order to ensure accurate sound localization.

The "Swiss-Stroll" track of the RACER car simulation software version 2.1 [12] was projected on the screen. The simulator was controlled with the Logitech MOMO Racing steering wheel and automatic gear changing was applied.

The same type of car (Peugeot 307) was used throughout the entire experiment. The experiment was performed in New Zealand and therefore the car was equipped for driving on the left hand-side of the road. Although the validity of the car simulator was not preformed we believe a very good approximation of a real driving task was achieved by using big screen projection, surround sound and steering wheel with force feedback.

The communication device used in the experiment was operated through a hierarchical multi-level menu. A simplified version of a NOKIA series 60 mobile phone menu was modified in order to have a maximum of six items at each menu level. The reason for this was our assumption that more than six items presented with simultaneous spatial sounds could not be perceived clearly.

4.2. Visual interface

The visual interaction was based on a small LCD screen (12cm x 15 cm) attached to the dashboard where it could be seen easily while driving. Different items of the menu were presented with large white fonts on a black background. The selected item was highlighted with a light green bar. When a specific item was selected, new submenu items were shown or, in the case of moving back in the menu structure, the previous items were loaded again.

4.3. Acoustic interface

In the two acoustic interfaces, different items of the menu were presented with spatial sounds and played to the driver through the speakers in the simulator. Spatial sounds were placed on a virtual ring around the driver's head. Each individual item was therefore represented with the sound at a certain position. The driver could navigate the menu by rotating the virtual ring with the sounds in any direction (i.e. left or right). The sound source located directly in front of the user represented the selected item (equivalent to the highlighted row in a visual menu). All the sound sources in the ring were always distributed equally in order to achieve the maximum possible spatial angle between them. For example, if there were three items in the current menu, the spatial angle between the individual items was 1200; if there were 6 items in the menu, the angle was 600, etc. The listener or the driver was positioned slightly to the front of the centre of the ring (closer to the front items). Due to this fact, the central front source, the one representing the selected menu item, was perceived as the loudest one.

The sound sources were spoken words - the menu items recorded by a female native English speaker. The signal-to-noise ratio of the signals was approximately 50 dB. A gentle background melody was assigned to each individual branch of the menu. The melody started as soon as the user left the

main menu and entered one of the submenus. The central pitch of the melody changed according to the current depth of the user in the submenu. Each time the user moved to a lower level of the menu, the pitch was lowered, and vice versa. The background melody helped the users to be aware of their absolute position in the menu.

4.4. Interaction device

With both types of interfaces the interaction with the communication device was performed with the help of a custom-made device consisting of a small scrolling wheel and two buttons.

All three parts of the device were attached to the steering wheel in order to be used safely while driving. The scrolling wheel was used to navigate between all available items at a certain level of the menu.



Figure 2. The interaction device consisting of a scrolling wheel and two buttons (left and right).

When used with the visual interface, the scrolling wheel would move the selection bar up and down in the menu. In the case of the acoustic menu, the wheel would turn the virtual ring with the sound sources in one of the two possible directions (i.e. left or right). In this case, the angle of each individual turn was always the angle between two neighboring items in the acoustic menu, so that one item was always selected.

The two buttons were used to either confirm the selection or move back (i.e. upwards) within the hierarchy.

4.5. Experiment conditions

Three different experiment conditions were created. The first two conditions were based on the two interfaces described in the previous section:

- condition V: the interaction was based on the visual interface
- condition A: the interaction was based on the acoustic interface with multiple simultaneous sounds

The third condition (A1) was also based on the acoustic interface. In this case, however, just one sound was played at a time. In condition A, up to six sound sources were played at different spatial positions and one of the sources represented the selected menu item. In condition A1, just one sound source was played at a time. Also in this case the sound source was spatially positioned in order to be easily separated from all other sounds (engine noise, traffic, environment noise, etc.).

We expected the interface with multiple simultaneous sounds to be more efficient and faster than the one with just one sound played at a time. By comparing A and A1 conditions, we wanted to check whether the capacity of the acoustic channel could be increased and the selection or the search time could be shortened with the use of multiple sounds. 4.6. Experiment procedure

A total of 23 test subjects participated in the experiment. Approximately half of them were more experienced with driving on the left hand side and half of them on the right hand side. They all reported normal sight and hearing. Before performing the experiment, all test subjects were asked to fill out a questionnaire on their age, sex, driving experiences, and hearing and sight disabilities. After a short demo of both interfaces and the interaction device, the test subjects were allowed a 5 minute test drive in the simulator in order to get familiar with the steering wheel, pedals, road conditions, etc.

After the demo, 18 test subjects were asked to perform four different tasks while driving:

- 1. Changing the active profile of the device PRF
- 2. Making a call to a specific person CAL
- 3. Deleting a specific image from the device IMG
- 4. Playing a specific song SNG

The tasks were performed three times (i. e. for each experiment condition). A 15-minute break was assigned after each condition and the test subjects were also asked to fill out the NASA TLX workload questionnaire and the QUIS test. In order to eliminate the learning effects between the different interfaces, three groups of six participants were formed. Each group performed the tasks with a different order of the conditions:

- 1. group: V, A, A1
- 2. group: A1, A, V

3. group: A, V, A1

In all three conditions, the test subjects were asked to drive the car safely and perform the tasks as fast as possible. Each task was read to the test subjects loudly and clearly. For each interface, the tasks were given to the test subjects in a random order. A successful completion of the individual task was signaled with the message "Task completed" (a sign on the screen in the visual menu and a recorded spoken message in the auditory menu). The duration times of the tasks and average speeds of the drivers were logged automatically. The entire experiment was recorded with a digital video camera and a post-analysis of the driving was performed in order to evaluate the safety of an individual test subject's driving.

The remaining 5 test subjects served as a control group and were asked to just drive the car without performing any tasks.

5. Results

In the tasks performed by 18 test subjects, four parameters or variables were evaluated:

- task completion times
- driving anomalies
- NASA TLX workload questionnaire [13]
- QUIS test [14]

The main results and interpretations are summarized in the following four subchapters.

5.1. Task completion times

The time required to finish each individual task was measured and logged automatically. The timer started when the initial command "Please start now!" was read to the test subject, and turned off automatically when the task was concluded successfully.

The analysis of variance (ANOVA) test compared the results of the tasks and showed no significant difference between the three conditions:

- F_{PRF} (2, 51) = 0.358, p = 0.701;
- $F_{CAL}(2, 50) = 0.550$, p = 0.581;
- F_{IMG} (2, 51) = 1.213, p = 0.306;
- F_{SNG} (2, 50) = 0.211, p = 0.811;

The mean values of task completion times are shown in table 1:

Table 1. Mean task completion times (M) and standard deviations (SD) in seconds

Condition	M_{PRF}	SD_{PRF}	M_{CAL}	SD_{CAL}
V	17.83	12.83	32.39	14.38
A	19.83	11.47	37.94	33.78
A1	16.72	8.87	29.12	23.63
Condition	M_{IMG}	SD_{IMG}	M_{SNG}	SD_{SNG}
V	31.50	21.36	37.90	27.31
A	37.17	27.18	33.38	20.81
A1	26.33	10.58	38.17	25.61

Table 2 shows the average task completion times of all tasks under individual conditions:

Table 2. Average task completion times of all tasks

Condition	Time / s		
V	17.83		
A	19.83		
A1	16.72		

We believe that the reason for non-significantly different results in all three conditions lies in the fact that the same interaction device was used in all cases. The test subjects were already used to watching the screen while driving. On the other hand, we expected the task completion times in condition A to be shorter that those in condition A1. In condition A, multiple simultaneous sounds were used and the information flow should therefore have been greater. However, the majority of the test subjects reported that condition A was too complicated due to the fact that it contained too many sounds for them to be able to perceive all of them at a certain moment. They reported condition A1 with just one sound played at a time to be more effective and easier to follow while driving.

5.2. Driving anomalies

The entire experiment was recorded with a digital video camera and the recordings were used for evaluating the driving performance. The car simulation program also enabled automatic logging of the driving speeds, crashes, etc. All drivers (the 18 drivers performing different tasks + the control group consisting of 5 test subjects) were evaluated for

each individual task. They were given the following penalty points for anomalies in driving:

- 1 penalty point: unsafe driving (slight winding on the road or slowing down unexpectedly and unnecessarily),
- 2 penalty points: extreme winding on the road and driving on the road shoulders,
- 5 penalty points: causing an accident or crashing the car.

The penalty points for each task were then summed up and the three conditions were compared again. The mean driving penalty points are shown in table 3:

Table 3. Mean driving penalty points (M) and standard deviations (SD) for the tasks

Condition	M_{PRF}	SD_{PRF}	M_{CAL}	SD_{CAL}
V	2.13	2.58	3.80	3.32
A	0.86	0.86	1.13	1.59
A1	0.87	0.99	1.07	1.68
Condition	M_{IMG}	SD_{IMG}	M_{SNG}	SD_{SNG}
V	4.20	5.22	3.67	4.30
A	0.67	0.62	1.07	1.33
A1	1.00	1.66	1.07	1.43

Figure 3 shows the average penalty points for all three conditions and the control group.

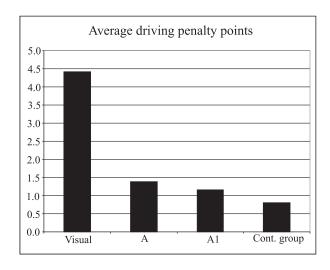


Figure 3. The average number of penalty points for all four conditions

The ANOVA test showed significantly different results for the tasks CAL, IMG and SNG and non-significantly different results for PRF task:

- FPRF(2, 41) = 2.795, p = 0.073;
- FCAL(2, 41) = 6.493, p = 0.004;
- FIMG(2, 41) = 5.479, p = 0.008;
- FSNG(2, 41) = 4.395, p = 0.019;

The control group consisting of five test subjects who were asked to just drive the car as safely as possible scored an average of 0.8 penalty points.

The results presented above show significantly fewer driving anomalies and a much greater safety when using the auditory interfaces. The two auditory interfaces were compared with a post-hoc T-test (0.5 limit on familywise error rate) and no significant difference in the results could be reported. Again, no advantage of condition A compared to condition A1 could be found.

The average driving speed was logged automatically by the driving simulator. Only the average speed of each individual test subject and each individual condition was recorded, not the speed for each task separately. The average speeds of the three conditions were:

- V: 32 km/h
- A: 59 km/h
- A1: 55 km/h
- Control group: 60 km/h

There is almost no difference in the average speed when using the two auditory conditions (A and A1); however, the speed of the test subjects using the visual condition (V) is approximately 25 km/h lower. We believe the difference reflects a great amount of cognitive workload in the visual condition, since the drivers had to concentrate on the road and on the screen simultaneously.

5.3. NASA TLX workload test

TLX workload test reports on the overall workload perceived by the test subjects under different conditions. It is based on a subjective questionnaire divided into six different subscales: mental demand, physical demand, temporal demand, performance, effort level and frustration level. The final score for each condition is a weighed average of all the ratings of the six subscales. The results of the test subjects reported a significant difference between the three conditions: F(2, 321) = 15.386, $p \mid 0.001$. The post-hoc T-test showed a significant difference in the workload between conditions V and A (p = 0.001), between conditions V and

A1 (p; 0.001), but no significant difference between the two auditory conditions (p = 0.053).

The reported results of the test subjects also reflect a high level of cognitive workload when operating a visual menu, since there is a lack of concentration which is mandatory for safe driving. The test subjects found the use of the auditory menus while driving easier and safer, and they also reported a lower perceived workload.

5.4. QUIS test

The QUIS test was designed to assess the users' subjective satisfaction with specific aspects of the human-computer interface. We intended to measure the reaction of the users to the software used in the experiment. We asked the users to rank each of the interfaces on a scale from 0 to 9 (0 being entirely false and 9 being entirely true), based on the following statements about each individual interface:

- 1. the interface was more wonderful than terrible (W&T)
- 2. the interface was more easy than difficult (E&D)
- 3. the interface was more satisfying than frustrating (S&F)
- 4. the interface was more adequate than inadequate (A&I)
- 5. the interface was more stimulating than dull (S&D)
- 6. the interface was more flexible than rigid (F&R)
- 7. it was easy to learn how to operate the system (O)
- 8. it was easy to explore new features by trial and error (E)
- 9. it was easy to remember names and use commands (R)

The ANOVA test showed a significant difference in the scores for the statements 1 to 4:

- W&T: F(2,51) = 9.401, p; 0.001;
- E&D: F(2,51) = 14.171, p; 0.001;
- S&F: F(2,51) = 7.413, p = 0.001;
- A&I: F(2,51) = 11.814, p; 0.001;

No significant difference in the scores could be found for the statements 5 to 9:

- S&D: F(2,51) = 3.143, p = 0.052;
- F&R: F(2,51) = 2.495, p = 0.093;
- O: F(2,51) = 1.073, p = 0.350;

- E: F(2,51) = 2.146, p = 0.127;
- R: F(2,51) = 1.529, p = 0.226;

Figure 4 shows the average scores of individual interfaces:

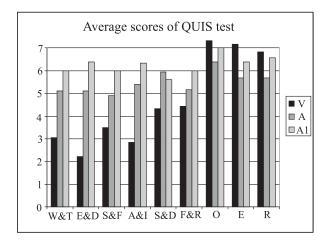


Figure 4. The average scores of individual QUIS factors

The results show that, in general, the users were satisfied with the auditory interfaces. The users found the auditory interfaces more wonderful than terrible, easy to use, satisfying and adequate. On the other hand, the users did not find them significantly more stimulating or flexible than the visual interface. As regards the learning required to use the interfaces, the users reported all interfaces to be equally difficult to learn to operate, to explore new features by trial and error, and also to remember names and commands.

6. Discussion

The main goal of this study was the evaluation of an acoustic interface as a substitute for the traditional visual interface (V) of an in-vehicle display. The four main variables measured in the experiment were task completion time, driving performance, workload and user satisfaction.

We did not find any significant difference in the task completion times. We believe the reason for this lies in the fact that the same interaction device was used in all three conditions. We find the result that prove the auditory and visual interfaces were equally fast very encouraging, since an entirely new interface was compared to a well-know and widely used visual interface. On the other hand, we expected condition A to be faster than condition A1 due to multiple simultaneous sounds and a larger information flow. That was not the case, since the majority of the test subjects found condition A too difficult to understand while driving.

The driving performance evaluation showed increased safety and a significant reduction in the distraction of the driver when the auditory interfaces were used. There was approximately a 60% difference in the penalty points between the visual and the auditory conditions. The average speed in the auditory conditions was approximately 25 km/h higher and therefore almost the same as the average speed of the control group. This most probably reflects that fact that the drivers felt more confident because they were not distracted by the information on the screen and were thus capable to pay attention to the road. The variations in the driving speed were also significantly smaller in the auditory conditions.

The results of the TLX workload test indicate that the users felt less physical and temporal demand when interacting with the auditory interfaces. They felt a high level of satisfaction and were confident about their performance. The use of the auditory interfaces made them feel more secure and less stressed than the use of the visual interface.

7. Design recommendations and conclusions

Our experiment offers some useful design recommendations for embedded communication systems in cars. The auditory interface with spoken commands proved to be very effective and as fast as the visual interface. Our test subjects reported the lack of feedback on the current location in the acoustic menu. They complained about occasionally getting lost and having to move back to the main menu to restart the task. The background music with a changing central pitch turned out to be a good solution as it helped the user to identify the individual submenus at any given time; however, it should perhaps be upgraded with a few spoken feedback options. For example, the option "current location" could read all the previously selected commands and inform the user on his or her current location.

Multiple simultaneous sounds did not prove to have any advantages when compared to a single sound source or menu item played at a time. The perception of multiple sounds while driving seems to be almost impossible and disturbing. The best results in the experiment were achieved in the auditory condition with just one sound source played at a time.

The visual interface turned out to be very unsafe and disturbing for the drivers. Although the LCD screen was attached to the dashboard where it could be seen easily when driving, a high number of driving penalty points still calls for a better solution. A head-up display developed by the BMW might turn out to be a better option for the visual interface; however, some further evaluations are still necessary [15]

The interaction device is also very important for the safety of the driver. Our solution with the scroll wheel and

two buttons turned out to be very practical and easy to use while driving a car. The test subjects found it safe to use since they could maintain both hands on the steering wheel at all times.

As this was only a pilot study, further research has to be done on comparing the auditory interfaces to novel visual interfaces, for example a head-up display or a speech interface. In addition, a more realistic and demanding driving scenario should be tested, such as a major street in an urban environment or driving under different weather conditions.

References

- [1] F. Bents, "Driver Distraction Internet Forum," From: http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/AskTheExperts.htm#CurrentExpertQuestions, 2000.
- [2] M.A. Pettitt, G.E. Burnett, "Defining driver distraction," *Proc. of World Congress on Intelligent Transport Systems*, San Francisco, USA, 2005.
- [3] L. Tijerina, "Issues in the Evaluation of Driver Distraction Associated with In-Vehicle Information and Telecommunications Systems," From: http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/3.PDF, 2000.
- [4] T.A. Ranney, E. Mazzae, E. Garrot, R. Goodman, "NHTSA Driver Distraction Research: Past, Present, and Future," From: http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/233.PDF, 2000.
- [5] B. N. Walker, A. Nance and J. Lindsay, "Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus," *Proc. of the International Conference on Auditory Display (ICAD 2006)*, London, England, 2006, pp. 63-68.
- [6] P. Lucas, "An evaluation of the communicative ability of auditory icons and earcons," *Proc. of the Second International Conference on Auditory Display, Santa Fe, USA, 1994, pp. 121-128.*
- [7] K. Crispien, K. Fellbaum, A. Savidis, C. Stephanidis, "A 3D-Auditory Environment for Hierarchical Navigation in Non-visual Interaction," *Proc. of the 3rd International Conference on Audio Display (ICAD '96)*, Palo Alto, USA, 1996, pp. 18-21.
- [8] S. Brewster, J. Lumsden, M. Bell, M. Hall, M. Tasker, "Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices," SIGCHI conference on Human

- Factors in Computing Systems, vol. 5, no. 1, 2003, pp. 473-480.
- [9] N. Sawhney and C. Schmandt, "Nomadic radio: speech & audio interaction for contextual messaging in nomadic environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 3, 2000, pp. 353-383.
- [10] Openal, From: http://www.openal.org/, 2007.
- [11] Creative Knowledgebase, From: http://us.creative.com/support/kb/, 2007.
- [12] RACER, From: http://www.racer.nl/, 2006.
- [13] NASA TLX for Windows, From: http://www.nrl.navy.mil/aic/ide/NASATLX.php, 2006.
- [14] QUIS, About the QUIS version 7.0. From: http://www.lap.umd.edu/quis/, 2006.
- [15] BMW, From: http://www.bmw.com, 2007.