Volume 66, issue 5 May 2008

ISSN 1071-5819

ELSEVIER

International Journal of
Human-Computer Studies

Available online at www.sciencedirect.com
ScienceDirect

Int. J. Human-Computer Studies 66 (2008) 318–332

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

A user study of auditory versus visual interfaces for use while driving

Jaka Sodnik[a,b,*], Christina Dicke[a,c], Sašo Tomažič[b], Mark Billinghurst[a]

[a]*Human Interface Technology Laboratory New Zealand, University of Canterbury, Private Bag 4800, Christchurch, New Zealand*
[b]*Faculty of Electrical Engineering, University of Ljubljana, Trzaska 25, Ljubljana, Slovenia*
[c]*Faculty of Mathematics and Natural Sciences II, Department of Psychology, Humboldt University, Chair of Engineering Psychology/Cognitive Ergonomics, Rudower Chaussee 18, Berlin, Germany*

Received 7 February 2007; received in revised form 31 August 2007; accepted 5 November 2007
Communicated by C. Schmandt
Available online 17 November 2007

Abstract

This paper describes a user study on interaction with a mobile device installed in a driving simulator. Two new auditory interfaces were proposed and their effectiveness and efficiency were compared to a standard visual interface. Both auditory interfaces consisted of spatialized auditory cues representing individual items in the hierarchical structure of the menu. In the first auditory interface all items of the current level of the menu were played simultaneously. In the second auditory interface only one item was played at a time. The visual interface was shown on a small in-vehicle LCD screen on the dashboard. In all three cases, a custom-made interaction device (a scrolling wheel and two buttons) attached to the steering wheel was used for controlling the interface. The driving performance, task completion times, perceived workload and overall user satisfaction were evaluated. The experiment proved that both auditory interfaces were effective to use in a mobile environment, but were not faster than the visual interface. In the case of shorter tasks, e.g. changing the active profile or deleting an image, the task completion times were comparable for all interfaces; however, both the driving performance was significantly better and the perceived workload was lower when using the auditory interfaces. The test subjects also reported a high overall satisfaction with the auditory interfaces. The latter were labelled as easier to use, more satisfying and more adequate for performing the required tasks than the visual interface. The results of the survey are not surprising as there is a stronger competition for the visual attention between the visual interface and the primary task (driving the car) than in the case of using the auditory interface. So although both types of interfaces were proven to be effective, the visual interface was less efficient as it strongly distracted the user from performing the primary task.

Keywords: Visual interface; Auditory interface; Spatial sound; Driving simulator; Mobile device; Interaction; Distraction

1. Introduction

Nowadays, mobility is becoming an essential part of our lives. Some of the tasks that we used to perform in the office or at home are now being done on the go. As a reaction to the requirements of a highly mobile and information-dense domain, our handheld communication devices are getting smaller while at the same time their functionality is expanding dramatically. Due to their miniaturization, mobile devices have limited input and output capacities, which makes the traditional mouse and keyboard WIMP-interaction paradigm not applicable in a mobile environment. The shrunken input and output devices like mini qwerty keyboards, joysticks or styli only enable the users to interact on a basic level. Mobility also requires a high degree of visual attention. Visual interfaces are therefore not suitable in that context, as they distract the user's attention from primary tasks such as steering a vehicle (WierWille and Tijerina, 1998; Sodhi et al., 2004). Moreover, mobile devices are often put in pockets, bags or otherwise placed out of sight. As a result, the displayed cues cannot be immediately seen.

[*] Corresponding author at: Faculty of Electrical Engineering, University of Ljubljana, Trzaska 25, 1000 Ljubljana, Slovenia. Tel.: +386 1 4768 494; mobile: +386 41 561 281; fax: +386 1 4768 266.

E-mail addresses: jaka.sodnik@fe.uni-lj.si (J. Sodnik), christina.dicke@hitlabnz.org (C. Dicke), saso.tomazic@fe.uni-lj.si (S. Tomažič), mark.billinghurst@hitlabnz.org (M. Billinghurst).

1071-5819/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2007.11.001

In this paper we explore alternative interfaces for mobile devices. Auditory, tactile or even olfactory capabilities of the human sensual system offer an alternative to the overloaded visual channel. Tactile interfaces such as vibration feedback have the advantage of not drawing the user's visual attention away from their main activity. However, they are mostly useful only for short notifications, not for communicating any complex messages. The user also needs to be within the reach of the device in order to register signals such as vibration.

In contrast, auditory user interfaces are flexible and scalable, ranging from simple non-speech cues, to earcons, auditory icons, hearcons, natural or synthetic speech output, and audio representation of multivariate and multi-dimensional data compounds. Similarly to tactile interfaces, auditory interfaces do not interfere with visual information processing. In addition, audio signals can capture the user's attention even if they come from a certain distance or if their source is hidden from view. This makes auditory interfaces a good alternative for the use in predominantly mobile domains.

In this paper, we explore the use of different types of auditory cues for a mobile phone in a driving environment. Before presenting our interface and experimental results, the next section reviews some previous related work. Section 3 describing our audio interface and the user study is followed by sections on the experimental methods and results. The paper concludes with a discussion section (Section 6), some conclusions and future work.

1.1. Auditory icons, earcons, spearcons, hearcons

Gaver (1986) first developed the concept of using natural everyday sounds to represent events and objects in a computer interface. The so-called "auditory icons" were included for the first time in the Apple SonicFinder (Gaver, 1989) and are still in use today. Auditory cues are usually divided into three categories based on their abstraction level: they can be *iconic, metaphorical* (*indexical*), or *symbolic.* While iconic representations try to acoustically reproduce an event as realistically as possible, the metaphorical or indexical auditory cues establish an analogy between an event and an associated sound. The so-called earcons have the highest abstraction level; they do not allow any semantic relation between an event and a sound, but rather assign an arbitrary audio signal to represent an event. Earcons can be designed not only to represent a single item, but also its position in a hierarchical structure (Brewster et al., 1993), either in audio-only interfaces (such as telephone-based interfaces (Brewster, 1997, 1998; LePlâtre and Brewster, 1998)) or multi-modal interfaces (Brewster et al., 1993; Brewster and Crease, 1999; Vargas and Anderson, 2003). It has been shown that earcons can successfully improve the usability of multi-modal interfaces for mobile use (Pirhonen et al., 2002).

Almost any listener can easily interpret simple auditory icons representing an event or object by playing a typical sound (e.g. deleting or "throwing away" a file being represented by the rattling sound of a trash bin). Developing auditory icons with a high compatibility for more abstract events (e.g. changing the active profile of a mobile phone) can thus be difficult and, in addition, lead to a reduced ability to interpret the auditory icon without training. The meaning of earcons needs to be learned a priori and is not transferable to other earcon "languages". Comparative studies of simple auditory icons and earcons show no significant difference in efficiency between the two (Jones and Furner, 1989; Brewster, 1994; Lucas, 1994), also when used in combination with spoken menu items for locating different items in a hierarchical structure (Walker et al., 2006). As the abstract auditory cues were named "earcons", the above-mentioned study introduces "spearcons". Spearcons are audio cues generated by converting the text of a menu item to speech and then speeding up the resulting audio clip until it is no longer comprehensible as speech. Spearcons in combination with a spoken menu text show a slight advantage over the spoken only menus but a strong advantage over earcons (Walker et al., 2006). The so-called "hearcons" were created to support the navigation of web pages (Bölke and Gorny, 1995) or hierarchical menus (Klante, 2003). Hearcons are three-dimensional (3D) abstract auditory objects positioned in an auditory interaction realm. They constantly emit sound and can be manipulated with the use of a pointing device.

As has already been proven, sound events can be successfully used to represent events within a hierarchical structure. It is possible not only to code information about the meaning of an event but also about its position in a hierarchy. The amount of information that can be represented by the sound event itself is clearly limited, but the way of representing the event can be used to add additional information. One way of representation is using spatial sound. The next section will give a short introduction to spatial auditory interfaces, their potentials and limitations.

1.2. Spatial auditory interfaces

Spatial audio interfaces represent audio items in different spatial locations and in this way use their position in a 3D space as a way of conveying additional information. As the human ears are located at either side of the head, the so-called binaural effects enable the humans to better determine the location of a sound source in terms of azimuth rather than elevation (Jin et al., 2004). The measurements of the so-called minimum audible angle (MAA) and minimum audible movement angle (MAMA) also report significantly higher horizontal spatial resolution than vertical or diagonal resolution (Grantham et al., 2003; Sodnik et al., 2005).

Spatial sounds can be effectively delivered through headphones in which case the positioning of an audio item in space is done by using head-related transfer functions (HRTFs) (Begault, 1994; Algazi et al., 2001; Cheng and

Wakefield, 2001). HRTFs are frequency responses of an acoustic path from the sound source to the human eardrums. They describe the reflections of the shoulders, head and pinna. HRTFs are usually measured as head-related impulse responses (HRIRs) for each individual listener separately. When generalized HRTFs—measured with a dummy head—are used for creation of virtual sound sources high localization error rate or the so-called localization blur can be perceived (Wenzel et al., 1993; Blauert, 1997).

Spatial sound in audio interfaces can also be delivered through various multiple speaker configurations: 4.1, 5.1, 7.1, etc. In this case multiple speakers are arranged around the listeners and the volume of the individual speaker depends on the position of the sound source. An individual channel is assigned to each speaker and a method of the so-called multi-channel panning is used to generate virtual sound source at any position (Pulkki, 2001; Jakka, 2005; Creative Knowledgebase, 2007). In common 5.1 or 7.1 speaker configurations all speakers are located in the same plane and no vertical positioning of sound sources is possible.

1.3. Attention and distraction

Attention is commonly defined as concentration of awareness to a specific source of information or a phenomenon to the relative neglect of other stimuli (James, 1890; Encyclopaedia Britannica; WordNet). Attention can either be willingly directed to a specific source of information or it can also be instinctive, if the event is a key stimulus, such as an alarming sound or a fast moving object. Attention can also be diverted, e.g. when a person drives a car and talks on a mobile phone at the same time. Distraction caused by a ringing phone, although momentary, diverts the attention from the task at hand. A more specific definition for driver distraction is given by the AAA Foundation for Traffic Safety: "when a driver is delayed in the recognition of information needed to safely accomplish the driving task because some event, activity, object, or person within or outside the vehicle compelled or tended to induce the driver's shifting attention away from the driving task." (Stutts et al., 2001). It becomes clear that the presence of a triggering event distinguishes distraction from inattentiveness.

Distraction leads to a reduced amount of attention on either task, the initially or primary and the new or secondary (Vollrath and Trotzke, 2000). Distraction is not only caused by physical stimuli through the sensual apparatus, but by cognitive sources, such as thought or emotional arousal (Bents, 2000; Pettitt et al., 2005) as well. Distraction from the primary task, i.e. driving the car, can reduce driver safety by degrading the vehicle control (speed maintenance, lane keeping, etc.) and object or event detection (Tijerina, 2000). Apart from the visual (eyes-off-the-road), auditory and cognitive distraction (mind-off-the-road), mechanical causes can also lead to distraction. Drivers either reaching for objects inside the vehicle or otherwise shifting out of their normal sitting position can have a degraded ability to execute manoeuvres (Ranney et al., 2000; Tijerina, 2000).

Handling a mobile phone while driving a vehicle or being otherwise on the move differs fundamentally from using it in the office or at home. Paying attention to operating the vehicle or watching the traffic is very important and being distracted from this is directly relevant to traffic safety. In the case of driving a vehicle, one way of reducing the risk of distraction is to abstain completely from using a mobile phone or other communication devices within the car. But as the car is becoming more and more an "office-on-the-move" (Sodhi et al., 2004), communication devices are indispensable and widely used despite the safety risks. A more appropriate way of reducing distraction from mobile devices is to adjust the user interface to the situation the device is being used in.

According to the multiple resource theory of attention (Navon and Gopher, 1979; Wickens, 1984), humans only have limited amounts of attention available at any given time. Different tasks can use different attention resources or share them. If the performed tasks rely on the same resource, they can interfere with each other and affect the performance. As driving a car is visually demanding, the visual interface of a mobile phone competes for the same resource associated with visual perception and can therefore cause distraction from the primary (driving) task (WierWille and Tijerina, 1998; Green, 2000).

Lee et al. (2000) argue that a speech-based interaction demands resources associated with auditory perception and would therefore be less detrimental. In our study, we tried to address the problem of interfering resources by using an auditory interface to navigate the menu of a mobile phone. To reduce the amount of mechanical distraction, we attached the physical interaction device to the steering wheel, so the driver's hands could remain on the steering wheel at all times while using the interface. Llaneras (2000) points out that although in this way visual and mechanical distraction can be partially reduced, cognitive distraction does not seem to be eliminated. Research (Mcknight and Mcknight, 1991; Lee et al., 2000; Ranney et al., 2000; ICBC, 2001) has shown that the complexity of the competing tasks plays a key role. Young et al. (2003) conclude that physical and cognitive distraction significantly impair the driver's visual search patterns, reaction times, decision-making processes and the ability to maintain speed, throttle control and lateral position on the road. Vollrath and Trotzke (2000) noted that attention can be safely diverted to a secondary cognitive task if the primary task is of low complexity. The perceived complexity of tasks depends amongst other things on age, emotional state and driving experience.

In our study, we used auditory interfaces of different complexity to operate a mobile phone while attending to a driving task. We reduced the mechanically and visually distracting events, so that we could focus on the influence of secondary tasks of varying complexity (conducted with an auditory interface) on the primary driving task. To build the auditory interface, we used spoken menu items as they have proven to be very effective (Lucas, 1994; Walker et al., 2006).

1.4. Related work

In our research we explore the use of a spatial audio interface that distributes the sound sources in a ring around the user's head. Several researchers have used the ring or dial metaphor for designing the auditory interfaces.

An early and influential application has been created by Cohen and Ludwig (1991). Audio Windows is a 3D auditory display which integrates spatial sound, audio highlighting, and gestural input recognition. Users of Audio Windows are wearing headphones and a data glove and can manipulate items by pointing at specific areas which are mapped to corresponding items in a 3D sound space.

Crispien et al. (1996) and Savadis et al. (1996) designed an egocentric spatial interface for navigating in and selecting from a hierarchical menu structure. The interface is designed for aligning both non-speech and speech audio cues in a ring circling around the user's head. These auditory objects can be reviewed and selected by using 3D-pointing, hand gestures or speech recognition input.

Kobayashi and Schmandt (1997) built an egocentric dynamic soundscape, a further development of the Audio-Streamer (Schmandt and Mullins, 1995), to create a browsing environment for audio recordings. In this application a speaker orbits the user's head as it reads out the audio data and hence maps advancing within the audio source to movements on the circular path. Using a touchpad the user can interact with the system to either create a new speaker (to rewind or fast forward) or switch to another already created speaker. There can be up to four speakers simultaneously playing different portions of the same audio stream. One speaker is always focussed, i.e. louder in volume.

Dell (1999) investigated the impact of using spatialized alarms versus stereo and mono alarms in aircrafts. In his study, the users were asked to keep a simulated horizon as level as possible (primary task) and to press a specific button to disable a specific irregularly sounding alarm (secondary task). The users wore headphones and used a mouse to control the simulated flight. One of the conditions was a 3D condition where auditory alarms were arranged in a circle around the user's head, positioned according to their position on the screen. The experiment did not show that 3D audio is more effective than stereo for the given tasks.

Sawhney and Schmandt (2000) created the Nomadic Radio, a spatial audio framework for a wearable audio platform. The framework notifies the user about current events such as incoming e-mails or voicemail, current messages and calendar entries. Confirmation, aborting and status are also represented by sounds. The audio messages are positioned in a circle around the listener's head according to their time of arrival. The user interacts with the nomadic radio by the use of voice commands and tactile input.

Goose and Djennane (2002) developed WIRE, the Web-based Interactive Radio Environment voice browser for providing drivers with access to WWW services whilst driving. WIRE analyses HTML documents and positions extracted elements according to type and location on to an arc expanding in front of the user. Distinct synthesized voices speak headers, content, and hypermedia link anchors. Earcons help supporting the differentiation of internal and external links and aid orientation. By turning a knob and giving voice commands the user can interact with the system.

Brewster et al. (2003) created a mobile system based on Audio Windows by Cohen and Ludwig (1991). They used spatialized auditory icons localized in the horizontal plain either around or in front of the user's head. By using head or hand gestures the user can select an auditory icon from the menu to trigger the corresponding event, such as, e.g. checking for news on traffic or the weather. Brewster and colleagues found that their auditory interface improves the usability of the wearable device.

Frauenberger and Stockman (2006) positioned the user in the middle of a virtual room with a large horizontal dial in front of him or her. Menu items were presented on the edge of the dial facing the user while most of the dial disappeared behind the wall. The user could turn the dial in either direction by using a gamepad controller. Only the item in front of the user could be selected or activated. All items were synthesized speech.

As can be seen from these projects, spatial audio using a ring metaphor has been applied in a number of interfaces. However, there have been fewer examples of this being applied in a mobile phone setting, and no previous work compared audio interfaces to purely visual conditions in a realistic mobile phone task. This is the area that we are addressing in this paper.

In the next section we describe our interface in more detail and present the user study we conducted to test the interface conditions.

2. User study

In our study we tried to find out whether an auditory user interface for a mobile phone is less distracting and more efficient than a purely visual interface. To simulate a real task, we created a typical "mobile" context by using a driving simulator. We observed drivers performing different tasks with an in-vehicle mobile phone or hands-free mobile device while they were driving the simulated vehicle. The driving simulator consisted of a large projection screen, steering wheel, accelerator, brake and mobile communication device which could be controlled with a custom-made interaction device attached to the steering wheel (see Fig. 1—the figure showing the driving simulator). In addition, an external keyboard was attached next to the steering wheel, which could be used for entering letters or text if necessary. A more detailed description of the simulator is given in Section 3.

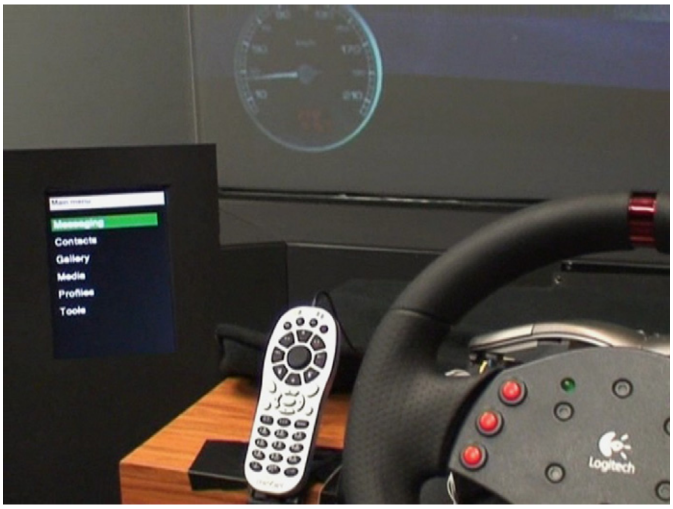

Fig. 1. The visual interaction based on a small screen and phone-like keyboard. The items in the visual menu were displayed in large white fonts and the selected item was highlighted with a green bar.

2.1. Conditions

Three different experiment conditions were created by the using three different user interfaces. In this section we describe these conditions in more detail.

The same menu structure was used with all three interfaces. The items and the levels of the menu were based upon a Nokia 60-series mobile phone menu but were reduced to a set of items most likely to be accessed in a mobile situation. There were up to six items on each level with the top level containing the following items:

- Messaging
- Contacts
- Gallery
- Media
- Profiles
- Tools

2.1.1. Condition 1: visual interface (V)

The first interface was a visual interface with the menu shown on the small screen. The screen was positioned at about 40° to the lower left of the dashboard where it could easily be seen while driving. The items of the menu were displayed in large white fonts and the selected item was highlighted with a green bar. In the tasks where a text message had to be entered, a small phone-like keyboard was used for entering individual letters (see Fig. 1).

2.1.2. Conditions 2 and 3: auditory interface with multiple (AM) or just one simultaneous sound (AS)

The second condition (AM) was an auditory interface, in which all the items of the menu and commands were presented with spatial sounds and played to the driver via the speakers installed in the simulator. In general, each item in the menu was assigned a corresponding virtual sound source. The sound sources were spoken words—readings of the menu items. At each level of the menu hierarchy the items—the sound sources were placed on a virtual circle which could be rotated around the user's head. Due to poor elevation localization all audio items were positioned in the horizontal plane.

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 was 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.

In the second condition (AM), all sound sources (1–6) of the current level of the menu were played simultaneously. The selected item was the loudest one and was positioned directly in front of the user at 0° azimuth (see Fig. 4). This makes use of the "Cocktail Party Effect", which is the human ability to filter several simultaneous sounds and to concentrate on only one (Arons, 1992; Cohen, 1992; Stifelman, 1994). We believe that the users are able to perceive the location of other sound sources (menu items) as well and use the additional information for better orientation within the menu (Hawley et al., 1999; Drullman and Bronkhorst, 2000).

In the third condition (AS), the positions of the menu items—the sound sources—are the same as in the AM condition, but only the front sound source is played.

In the auditory conditions (AM and AS), textual input was also realized with an acoustic interface. Two major letter groups (vowels and consonants) were represented on one level of the menu together with "Space", "Erase letter" and some other commands (see Fig. 2). On the next level, the consonants were further divided into six smaller groups of letters ("b, c, d", "f, g, h", etc.). On the next level each single letter is represented with the corresponding sound source. After each selection of an individual letter the user was automatically moved back to the first level of text input menu.

For example, to compose a text message "HI", the user would first need to select the group "Consonant", then the group "f, g, h" and then the letter "h". After this selection, the user would be automatically moved back to initial position of the text input and would again need to select between "Consonant", "Vowel", etc. This time the user would select the group "Vowel" and then further the letter "i". The input of the message "Hi" would thus be completed (see Fig. 2).

2.2. Interaction technique

In all three cases the interaction in the experiment is conducted through a custom made interaction device

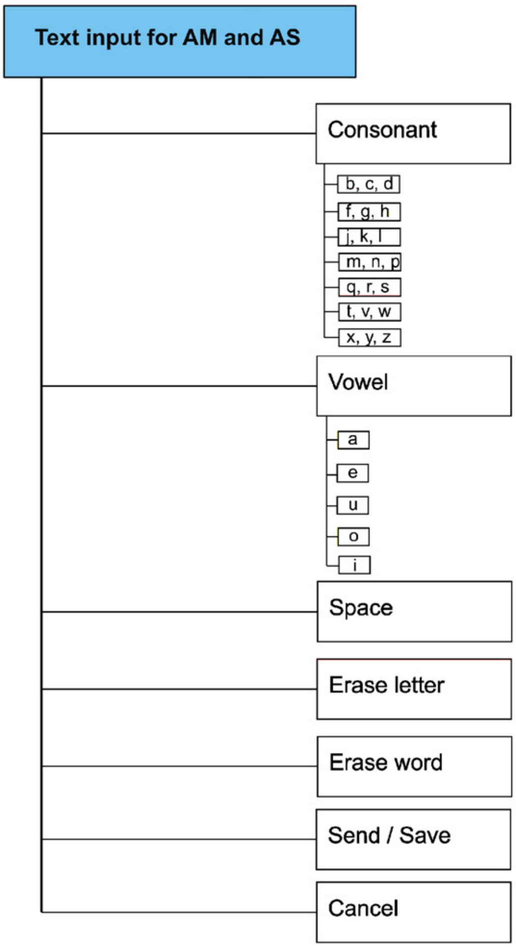

Fig. 2. A schematic diagram of the auditory menu used to compose text messages in the AM and AS conditions.

consisting of a small scrolling wheel and two buttons (left and right) attached to the steering wheel (see Fig. 5).

In the first condition (V), the scrolling wheel was used to move the selection bar up and down in the menu and the two buttons were used to confirm or cancel the selection. In the second (AM) and third (AS) conditions, the scrolling wheel turned the virtual circle with the sound sources, thus changing the position of the items in the menu. As already mentioned, the item in front of the user was always the selected one. The two buttons again enabled the confirmation or cancellation of the selected menu item.

2.3. Tasks

We were interested in observing the users operating the car (primary task) and at the same time performing different (secondary) tasks with the in-build mobile device. In analysing the driver performance, we focused on possible anomalies and unsafe reactions in driving. By unsafe reactions we refer to winding on the road, reducing the speed significantly when this is not required, driving off the side of the road or even crashing the car.

The users were asked to perform five different tasks that are typical for using a mobile phone in a car:

1. Writing a text message to a specific person—MSG.
2. Changing the active profile of the device—PRF.
3. Making a call to a specific person—CAL.
4. Deleting a specific image from the device—IMG.
5. Playing a specific song—SNG.

Three different interfaces with the same interaction equipment were compared. Our main research questions were:

1. Which interface will distract the user less from the primary task?
2. Which interface will cause the user to make more errors?
3. Which interface will have the shortest task completion times?
4. Will the audio interface with multiple simultaneous sounds (AM) be more distracting than the audio interface with just one sound (AS)?

Our main expectation was that the use of the auditory interfaces (AM and AS) would distract the users less from the primary task (driving) than the visual interface (V). Consequently, the driving performance should therefore be significantly better in conditions AM and AS. We also expected shorter task completion times when using AM or AS, especially with simple tasks, such as changing the user profile or calling someone.

In comparing AM and AS, we expected the AM condition to be more efficient due to a larger information flow (in AM many sounds are played simultaneously). In the AM condition, the users should therefore have a better awareness of their current location in the menu and the positions of individual items in the level of the menu. On the other hand, we expected some users to find the AM condition quite noisy or confusing and therefore prefer the AS case where only one sound is played at a time.

3. Methods

3.1. Test subjects

A total of 18 test subjects (8 female and 10 male) participated in our experiment of performing the tasks with the three different interfaces. The average age of the test subjects was 27.7 years with an average of 8.7 years of driving experiences. Half of the test subjects were more experienced with driving on the left-hand side of the road and half of them on the right-hand side. They all reported normal sight and hearing. An additional group of 5 test subjects (1 female and 4 male) participated as a control group for driving performance.

3.2. Experiment procedure

All test subjects were first asked to fill out a questionnaire on their age, sex, driving experiences, hearing and sight disabilities. After sitting in the simulator and getting acquainted with the interface control (scrolling wheel, buttons, screen), the test subjects were asked to drive for approximately 5 min to get used to the driving simulator and the road conditions. The warm-up drive was followed by performing 5 tasks with the use of the first interface. After a 15-min break, the test subjects were asked to repeat the tasks with the second interface and, after an additional 15-min break, with the third interface. No warm-up drive was preformed before the second and the third tasks. After each segment, the users were asked to fill in the NASA TLX workload test (The Task Load Index) (Hart and Wickens, 1990) and a slightly modified Questionnaire for User Interface Satisfaction (QUIS) (QUIS, 2006). They were interviewed in order to collect their personal evaluation of the experiment.

In order to eliminate the learning effects between the different interfaces, three groups of six participants were formed. Each group performed the tasks with the conditions in a counterbalanced sequential order:

1. group: V, AM, AS.
2. group: AS, AM, V.
3. group: AM, V, AS.

In all conditions the test subjects were asked to drive the car safely and perform the tasks as quickly as possible. Each task was read to the test subjects loudly and clearly. The tasks were ordered randomly for each interface. The successful completion of the individual tasks was signalled 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 used to evaluate the correctness of the individual user's driving.

The control group of five people was asked to just drive on the same road for approximately 5 min without performing any tasks.

The experimental measures collected were the following:

1. Task completion time.
2. Driving performance.
3. NASA TLX workload test.
4. QUIS test.
5. Personal comments of the test subjects.
6. Digital camera recording of the entire experiment.

3.3. Design

3.3.1. Car simulation

The experiment took place in a visualization room equipped with a large projection screen (2.4 m × 1.8 m) and 7.1 surround sound system. The car simulation software RACER version 2.1 (RACER, 2006) with the "Swiss-Stroll" track was projected on the screen. The simulator was controlled with the Logitech MOMO Racing steering wheel and automatic gear changing was applied (see Fig. 3). 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.

3.3.2. Sound reproduction

The Creative Sound Blaster X-Fi ExtremeMusic sound card and the Creative GigaWorks S750 speaker configuration system were used for sound reproduction. We decided not to use headphones, as blocking the auditory sense would keep the user from parsing other co-occurring auditory events. We wanted the users to also hear other co-occurring auditory events in the simulator (the sound of the car engine, braking, environment sounds, etc.). Spatial sound generation was driven by the Creative OpenAL sound library (OpenAL, 2006) which enabled access to all X-Fi hardware accelerated 3D sound features. OpenAL enables the simple positioning of virtual sound sources in 3D space using the CMSS-3D surround sound technology on the Creative sound card. CMSS-3D creates eight individual sound channels using multi-channel upmix process (Creative Knowledgebase, 2007). Each sound channel drives individual speaker.

The sound sources were the spoken words ("Messages", "Pictures", "Contacts", etc.) of the items in the menu, recorded by a female native English speaker. The signal-to-noise ratio of the signals was approximately 50 dB.

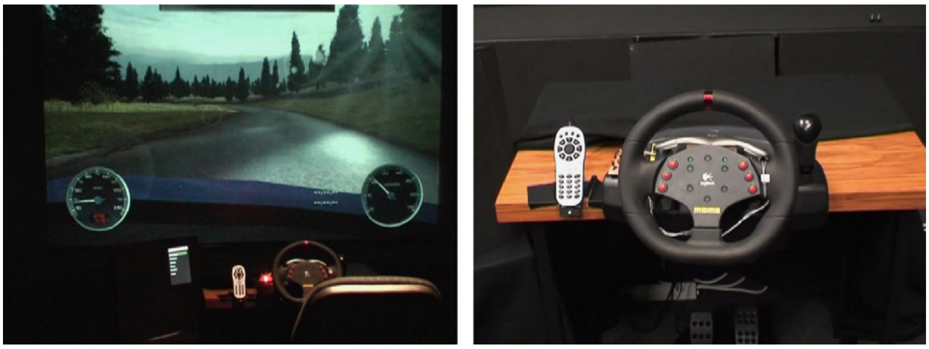

Fig. 3. Car simulator with a large projection screen, steering wheel, and interaction device.

All eight speakers in the driving simulator were positioned according to Dolby recommendations for 7.1 speaker systems (Dolby, 2007). The driver was poisoned in the sweet spot in order to ensure accurate sound localization.

3.3.3. *Visual menu*

A hierarchical multi-level menu was used. The visual menu was presented to the users on a 12 cm × 15 cm LCD screen with a large white text on a black background, similar in style to the one used in Blaskó and Feiner (2002). The current selection in the menu was highlighted with a light green bar. The application with the visual menu was developed using the .NET programming environment (see Fig. 1).

3.3.4. *Acoustic menu*

Both acoustic menus were developed in .NET programming environment using the OpenAL library. At each level in the menu, one to six sounds were generated and positioned at an equal distance around the user on a virtual circle (see Fig. 4). For example, if there were, for example, three items in the current menu, the spatial angle between the individual items was 120°, while if there were 6 items in the menu the angle was 60°, etc. The centre of the virtual acoustic circle was positioned slightly to the back (see Fig. 4) in order to put the listener closer to the front items of the circle. The sound source positioned directly in front of the user (at azimuth 0°) was the selected one and therefore the loudest.

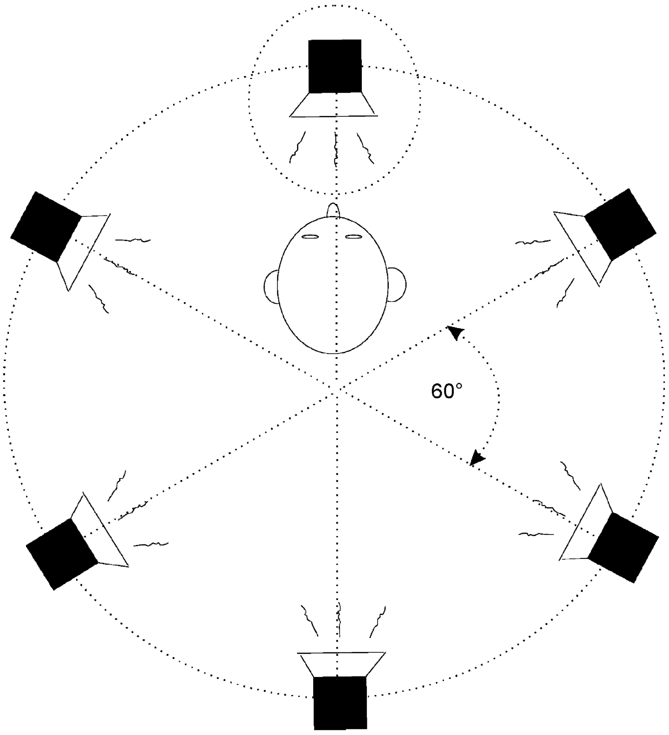

Fig. 4. The virtual sound sources were distributed equally around the user's head. The virtual circle could rotate in any direction. The sound source located in front of the user was always the active one or the selected one.

3.3.5. *Menu control*

All three menus could be controlled with a custom made navigation device attached to the steering wheel. The navigation device consisted of a scroll wheel and two buttons. The device was designed to be easy to operate by the test subjects' when driving. The menu could be rotated in any way with the use of the scroll wheel: in the visual menu condition the selection bar moved up or down and in the auditory menu conditions the virtual circle with sound sources was rotated around the user's head. The angle of the turn was always the angle between two neighbouring items in the menu—so that one item was always selected. The left button confirmed the selected item and loaded the items of the following level in the menu. The right button enabled a step back in the menu or cancelled the selected option. A small phone-like keyboard enabled text input when using the visual interface (see Fig. 3).

4. Results and interpretation

4.1. *Task completion times*

The task completion time was measured between the initial command "Please start now." and the final

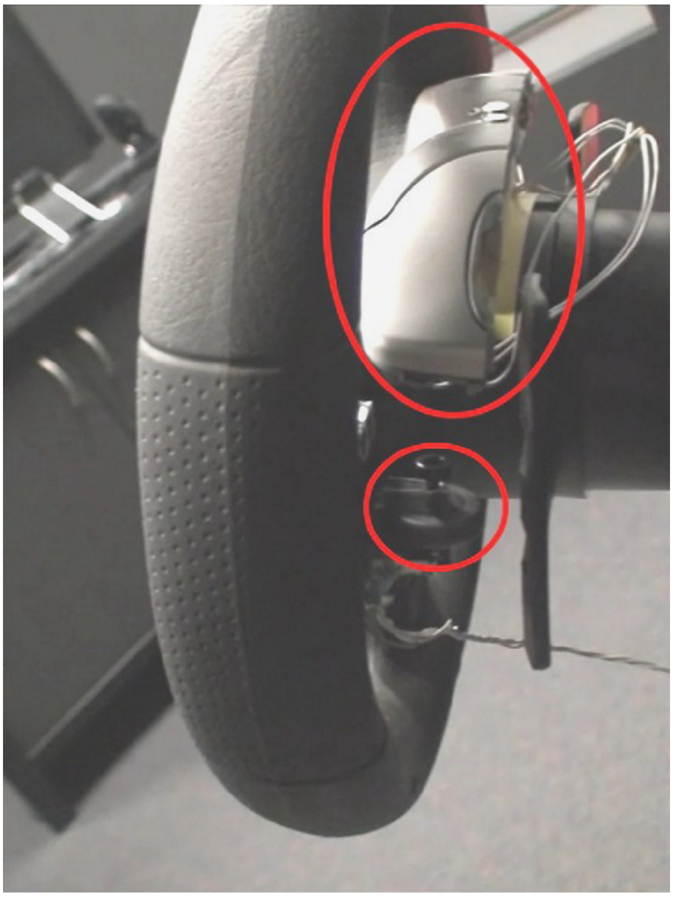

Fig. 5. The interaction device consisted of a scrolling wheel and two mouse buttons. The scrolling wheel enabled the selection of the item in the menu and the two buttons enabled the confirmation (left button) or the cancellation (right button) of the selection.

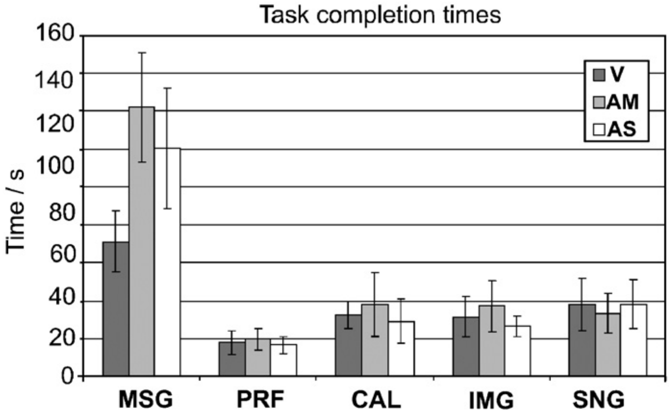

Fig. 6. Mean task completion times for all tasks with the three interfaces employed in the experiment (V—visual menu; AM—auditory menu with multiple simultaneous sounds; AS—auditory menu with single sound). The tasks preformed: MSG—composing and sending a message; PRF—changing the active profile; CAL—making a call to a specific person; IMG—deleting a specific image; SNG—playing a specific song.

notification "Task completed". The command was read to the users after the instruction on the individual task and the final notification was shown or played automatically. Fig. 6 shows the average task completion times for the five tasks in the three different interface conditions.

There was a significant difference in task completion times for the message composition task (MSG). The visual menu with a mobile phone keyboard proved to be the fastest way to write a text message. The within subject ANOVA test for MSG task with three different conditions gave the following result: $F_{\mathrm{MSG}}(2,51)=8.52$, $p=0.001$. A post-hoc Bonferroni test with a 0.05 limit on familywise error rate confirmed a significant difference between the visual (V) and auditory menus (AM and AS), but no significant difference between the AM and AS. The mean task completion times (in seconds) of MSG tasks are presented in Table 1.

We believe that the reason that the visual interface was significantly faster lies in the fact that most test subjects were already skilled in writing text messages with mobile phone keyboards. The audio interface for entering text messages turned was too slow and quite inappropriate for this environment.

The ANOVA tests for the other four tasks show no significant difference between the individual interface conditions:

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

So there was no difference in task performance time for the other four tasks. These results did not confirm our expectations that the auditory menus would support faster task completion times.

Table 1
Mean task completion times (M) and standard deviations (S.D.) for MSG task

Interface	M_{MSG}	S.D.$_{\mathrm{MSG}}$
V	71.22	32.24
AM	120.50	63.54
AS	142.22	57.55

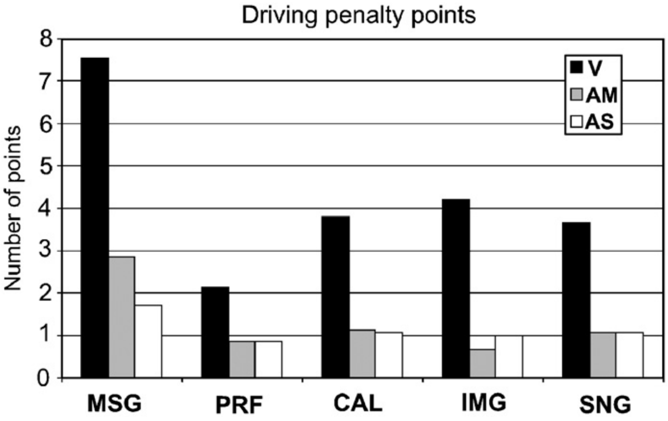

Fig. 7. Mean driving penalty points for all tasks with the three interfaces used in the experiment (V—visual menu; AM—auditory menu with multiple simultaneous sounds; AS—auditory menu with single sound). The tasks preformed: MSG—composing and sending the message; PRF—changing the active profile; CAL—making a call to a specific person; IMG—deleting a specific image; SNG—playing a specific song.

4.2. Driving performance

The driving performance was evaluated using video recordings of the subjects driving. The user's driving during each individual task was observed and penalty points were assigned according to the following criterion:

- *1 penalty point*: unsafe driving such as slight winding on the road and 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 and crashing the car.

The penalty points for each driver were summed up and the average penalty points for all users were calculated for each task (see Fig. 7).

The number of penalty points was significantly greater in the case of visual menu condition for four tasks: MSG, CAL, IMG and SNG. In addition, the PRF results are also nearly significantly different across the three conditions. The ANOVA test yielded the following results:

$$F_{\mathrm{MSG}}(2,41)=10.075,\ p<0.001;$$
$$F_{\mathrm{PRF}}(2,41)=2.795,\ p=0.073;$$
$$F_{\mathrm{CAL}}(2,41)=6.493,\ p=0.004;$$

Table 2
Mean driving penalty points (M) and standard deviations (S.D.) for the tasks: MSG—composing and sending the message; CAL—making a call to a specific person; IMG—deleting a specific image; SNG—playing a specific song

Interface	M_{MSG}	S.D.$_{\text{MSG}}$	M_{CAL}	S.D.$_{\text{CAL}}$	M_{IMG}	S.D.$_{\text{IMG}}$	M_{SNG}	S.D.$_{\text{SNG}}$
V	7.53	5.11	3.80	3.32	4.20	5.22	3.67	4.30
AM	2.86	3.58	1.13	1.59	0.67	0.62	1.07	1.33
AS	1.71	1.32	1.07	1.68	1.00	1.66	1.07	1.43

Table 3
The relative improvement of the driving performance comparing the AM (auditory menu with multiple sound sources) condition and the AS (auditory menu with a single sound source) condition to the V (visual menu) condition

Comparison	MSG (%)	PRF (%)	CAL (%)	IMG (%)	SNG (%)
AM to V	71	33	78	50	66
AS to V	55	57	64	77	59

$$F_{\text{IMG}}(2,41) = 5.479,\ p = 0.008;$$
$$F_{\text{SNG}}(2,41) = 4.395,\ p = 0.019.$$

A post-hoc Bonferroni test with a 0.05 limit on familywise error rate confirmed a significant difference between the results of the visual interface and auditory interfaces, but no difference between the individual auditory interfaces (AS and AM). The mean values of the four tasks are presented in Table 2.

The average number of penalty points for the control group (test subjects who just drove the car) is 0.8. That means that there is almost no difference in the mean values of the two auditory conditions (see Table 2) and the control group (except perhaps for MSG task). The ANOVA test cannot be performed in this case since only five test subjects participated as test drivers.

We can also corroborate the statement about driving improvement I in auditory conditions by calculating the relative change of the driving penalty points, comparing the AM or AS condition to V condition. The relative change I_{AM} and I_{AS} per user per task can be defined as

$$I_{\text{AM}} = \frac{D_{\text{AM}} - D_{\text{V}}}{D_{\text{V}}} \quad I_{\text{AS}} = \frac{D_{\text{AS}} - D_{\text{V}}}{D_{\text{V}}},$$

where D_{AM} and D_{AS} are the number of penalty points when using each of the auditory interfaces and D_{V} is the number of penalty points when using the visual interface. The mean values of driving improvement of all users are presented in Table 3:

The average improvement (in all tasks) of driving performance in AM condition compared to V condition is 62% and 60% in AS condition compared to V condition.

4.3. NASA TLX workload test

The NASA TLX is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales (NASA TLX for Windows):

1. *Mental demand*: How much mental and perceptual activity was required (thinking, deciding, calculating, remembering, etc.)?
2. *Physical demand*: How much physical activity was required (pushing, pulling, turning, controlling, etc.)?
3. *Temporal demand*: How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?
4. *Performance*: How successful do you think you were in accomplishing the goals of the tasks set by the experimenter?
5. *Effort level*: How hard did you have to work (mentally and physically) to accomplish your level of performance?
6. *Frustration level*: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content and relaxed did you feel during the task?

Fig. 8 shows the final overall workloads with standard deviations for all three interfaces:

The ANOVA test found a significant difference in the overall workload between the three conditions: $F(2,321) = 15.386$, $p < 0.001$. The post-hoc Bonferroni test with a 0.05 limit on familywise error rate showed that the workload reported in the visual condition (V) was significantly higher from the workload in the AM condition ($p = 0.001$) and from the workload in the AS condition ($p < 0.001$). No significant difference between the two auditory conditions could be established ($p = 0.053$).

Further examination of the results of the individual subscales of TLX workload test revealed some interesting

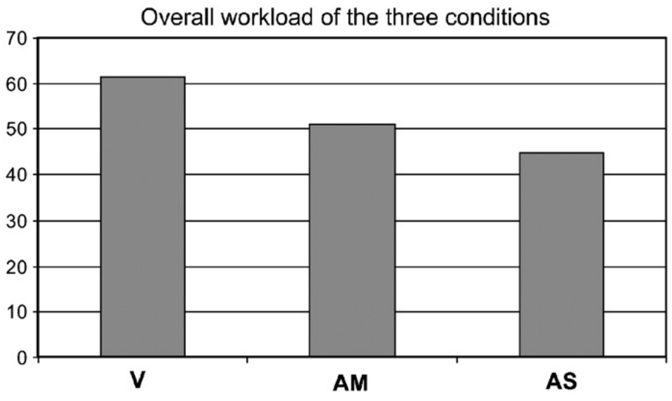

Fig. 8. Mean values and standard deviation of the final TLX workload test (V—visual menu; AM—auditory menu with multiple simultaneous sounds; AS—auditory menu with single sound).

Table 4
Mean values of the first part of QUIS test (whether the specific interface was more wonderful than terrible)

Interface	M_1	S.D.$_1$
V	3.06	2.07
AM	5.11	2.22
AS	6.00	1.97

results. There is a significant difference between the conditions in the following four subscales:

- physical demand: $F(2, 51) = 4.090$, $p = 0.023$;
- temporal demand: $F(2, 51) = 4.648$, $p = 0.014$;
- performance: $F(2, 51) = 4.237$, $p = 0.020$;
- frustration: $F(2, 51) = 3.188$, $p = 0.049$.

The post-hoc analysis showed that, in all four cases, the visual (V) condition differed significantly from the other two (AM and AS), but the auditory menus did not differ from one another.

We find the significant difference in the reported workload very encouraging. The auditory interfaces did not result in a significantly lower workload in only two subscales: mental demand and effort. We believe that both subscales indicate a relatively high cognitive workload due to the novelty of the auditory interfaces, which demanded some learning and adaptation. The very positive results of all other subscales indicate a high overall satisfaction of the test subjects with the auditory interfaces.

4.4. QUIS test

The QUIS test was designed to assess the users' subjective satisfaction with specific aspects of the human–computer interface. With our questionnaire we intended to measure the overall system satisfaction (the reaction to the software used in the experiment). We added some additional questions on learning abilities with the individual interfaces (questions 7–9). The average scores of all three interfaces were compared with the ANOVA tests and post-hoc Bonferroni test with a 0.05 limit on familywise error rate. The auditory interfaces resulted in significantly higher mean values than the visual interface when we asked the users if they found the individual interface more (on the scale 0–9) (Tables 4–7):

1. wonderful than terrible: ($F(2, 51) = 9.401$, $p < 0.001$);
2. easier than difficult: ($F(2, 51) = 14.171$, $p < 0.001$);
3. satisfying than frustrating: ($F(2, 51) = 7.413$, $p = 0.001$);
4. adequate than inadequate: ($F(2, 51) = 11.814$, $p < 0.001$).

The scores were not significantly different when the users were asked if they found the interface more (Table 8):

5. stimulating than dull: ($F(2, 51) = 3.143$, $p = 0.052$);
6. flexible than rigid: ($F(2, 51) = 2.495$, $p = 0.093$).

Table 5
Mean values of the second part of QUIS test (whether the specific interface was more easy than difficult)

Interface	M_2	S.D.$_2$
V	2.22	2.42
AM	5.11	2.65
AS	6.39	2.43

Table 6
Mean values of the third part of QUIS test (whether the specific interface was more satisfying than frustrating)

Interface	M_3	S.D.$_3$
V	3.50	2.30
AM	4.89	2.02
AS	6.00	2.30

Table 7
Mean values of the fourth part of QUIS test (whether the specific interface was more adequate than inadequate)

Interface	M_4	S.D.$_4$
V	2.83	2.31
AM	5.39	2.48
AS	6.33	1.88

Table 8
Mean values of the fifth and sixth parts of QUIS test (whether the specific interface was more stimulating than dull or more flexible than rigid)

Interface	M_5	M_6	S.D.$_5$	S.D.$_6$
V	4.33	4.44	2.27	1.76
AM	5.94	5.17	1.82	2.23
AS	5.61	6.00	1.98	2.25

In addition there was no significant difference between the interface conditions in the users' ability to (Table 9):

7. learn to operate the system: ($F(2, 51) = 1.073$, $p = 0.350$);
8. explore new features by trial and error: ($F(2, 51) = 2.146$, $p = 0.127$);
9. remember names and use commands: ($F(2, 51) = 1.529$, $p = 0.226$).

The results show that the users had a high overall satisfaction of the users with the new visual and auditory interfaces. The users found the auditory interfaces wonderful, 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

Table 9
Mean values of the seventh, eighth and ninth parts of QUIS test (whether the specific interface was easy to learn to operate, easy to explore new features by trial and error and easy to remember names and use commands)

Interface	M_7	M_8	M_9	S.D.$_7$	S.D.$_8$	S.D.$_9$
V	7.33	7.17	6.83	2.00	2.06	1.89
AM	6.39	5.67	5.67	2.23	2.45	2.17
AS	7.00	6.39	6.56	1.61	1.98	2.20

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. This could reflect the fact that most of the users had experience with and were used to visual interfaces in everyday use.

4.5. User comments

After each experiment we interviewed the subjects about their experience. In this section we list the most frequent positive and negative comments made by the subjects for each interface.

The visual interface (V):

Positive

- The interface was very simple to use.
- There was better information on the current position in the menu.
- It was faster than the auditory interfaces.

Negative

- It demanded full attention for operation.
- The users had to wait for an "easy" segment of the road to complete the tasks.
- It was very distracting and dangerous.

The auditory interface with multiple simultaneous sounds (AM):

Positive

- It was very easy to drive and complete the tasks simultaneously.
- The drivers could keep their hands on the wheel.
- It was very useful, especially for short tasks.

Negative

- It was hard to listen to, especially when the engine noise was loud.
- There were too many different sounds at the same time.
- There was no overview of the entire menu structure.

The auditory interface with just one sound (AS):

Positive

- It was less distracting than the visual interface.
- It was easy to understand and adapt to.
- It was less confusing than the interface with more sounds.

Negative

- Writing a message with the acoustic menus was too complicated and took too long.
- There was no good feedback on the entered words.
- There was no information on the current position in the menu and the users sometimes had to scroll through all the items.

5. Discussion

In our study, we observed users driving a car and performing different tasks with a communication device in the car. Five tasks with different difficulty levels were used to distract the users from their primary task—driving. The four main variables measured in the experiment were task completion time, driving performance, NASA workload test and the QUIS test.

The main goal of this study was the evaluation of an acoustic interface as a substitute for a traditional visual interface (V) on an in-vehicle display. Two acoustic interfaces were compared with either just one (AS) or up to six (AM) sounds played simultaneously. All three interfaces consisted of the same hierarchical menu structure simulating the common mobile phone interface structure and were controlled with the same custom-made interaction device, consisting of two buttons and a scroll wheel.

We did not find any significant difference in the task completion times, except for the MSG task, whereby the users had to enter and send a text message to a specific person. The longer task completion time in this case is a consequence of the use of different and unequally efficient interaction devices with visual and acoustic interfaces (a mobile phone keyboard and an auditory menu for writing messages). We believe that the same interaction procedure should be used in all cases—some test subjects suggested using a speech recognition system. We believe the similar task completion times in the other three cases are encouraging, since entirely new acoustic interfaces were compared to a well-known and widely used visual interface.

Our high expectations about significant improvement of the driving performance were justified. The users drove the car more safely when operating the auditory interfaces, since the average number of penalty points dropped by 60% in the audio conditions when compared to the visual interface condition. Comparison of the driving performance of the control group—just driving a car—to the test subjects performing tasks with auditory interfaces shows no degradation of driving performance. On the other hand the driving speed of the control group was in average a bit higher since the drivers were not asked to do any peripheral activity while driving. The users found performing the tasks with the visual menu very difficult, dangerous and unpleasant. In some cases, they had to slow down or even stop the car to perform the task safely, which caused large variations in the speed. In the case of auditory interfaces, these variations were not noticeable.

The users reported a significant difference in the perceived workload between the three conditions. In general, 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. We expected the users to find the new auditory interfaces harder to use and to adapt their behaviour to them. Therefore we estimate the similar amount of perceived workload, in terms of mental demand and effort, very encouraging. The results of the QUIS test also showed a high satisfaction of the users with the auditory interfaces. Subjects reported that the auditory interfaces were more wonderful, easier to use, more satisfying and more adequate than the visual interface.

Most of the test subjects commented on the importance of learning effects in the experiment, especially with the auditory interfaces. The visual interface was more effective and easier to use at the beginning, but the auditory interfaces became as effective after a few uses. The users reported that the auditory interfaces could be quite confusing when performing longer tasks that required a lot of movement through the hierarchical menu structures. The users reported having difficulty orientating themselves within the menu structure, which was not the case in the visual interface. The latter was confirmed with the last three results of the QUIS test where the users reported the visual interface to be easier to learn to operate and to explore new features. Also the general orientation within the menu was easier in the visual condition.

In the experiment we also studied the use of more simultaneous sounds in the interface. In the AM interface, all items of the current menu level were played simultaneously in an attempt to present as much information as possible at a given time. While in the AS case, the sounds were played one at a time. Although all simultaneous sounds could be perceived clearly in the stationary situation (when using the AM interface without driving), the users reported the AS option to be more effective, since it made it easier for them to concentrate on the driving. When driving they reported all additional (non-selected) sounds at different virtual positions around their head to be perceived more or less just as background noise and not as additional information which would enlarge the information flow. Therefore also the time required to find an individual item in a certain level of a menu seemed to be almost the same in both auditory conditions.

5.1. Design recommendations

The auditory interface could represent a great benefit of the in-vehicle information systems or mobile devices. The safety of driving could increase significantly since the users' eyesight would not be disturbed by watching the screen or the in-vehicle control panel.

Based on the results of our experiment, we can give some recommendations for the design and development of in-vehicle information systems or mobile devices. The auditory interface with spoken command proved to be very effective for shorter tasks such as changing the settings, playing songs or making a call to someone. However, good feedback on the current location of the user in the menu should be given in order to avoid confusing situations where the user gets completely lost and needs to restart the task from the beginning. The background music with a changing central pitch turned out to be a good solution to help the user identify the individual submenus at any time, but it should be upgraded with some spoken feedback options. For example, an option "the current location" could read all previously selected commands and inform the user on his or her current location.

Our interface with an auditory text input system proved to be too slow and therefore inappropriate for composing messages or performing longer tasks that demand the input of text. An effective voice recognition system would be worth testing, but we expect various problems due to the noisy in-vehicle environment which consists of the noise of the engine, passengers, traffic, etc. We also believe that an effective interaction device is very important. Our solution with the scroll wheel and two buttons turned out to be very appropriate and easy to use while driving a car. The users found it safe to use, since they could have both hands on the steering wheel at all times.

Finally, it is important to emphasize that multiple simultaneous sounds might not be the best option for use while driving since high cognitive workload and reduced concentration prevents the drivers from perceiving so much information at a time.

6. Conclusion

We believe that one of the most common distractions in a car is the use of mobile phones or other communication devices. The acoustic interfaces described in this paper offer an alternative to the traditional visual interfaces currently used in cars and other vehicles. Our simple prototype of a hierarchical acoustic menu seems to enable effective interaction with a mobile device used in a car and presents a low-level distraction to the driver. The safety of the driver while performing different tasks with the communication device can therefore be increased significantly. The proposed custom made interaction device also proved to be very effective and safe to use, since the users could have both hands on the steering wheel at all times.

Although the acoustic menu could sometimes be difficult and confusing to use, it offered significant improvement in the driver behaviour. It proved to be especially effective for short but commonly performed tasks like playing a song, calling someone or changing a profile. The perceived workload when operating the device with an acoustic interface proved to be lower to the one perceived when operating the device with a visual interface. The possibly complicated menu structure could be learned quickly and then be as effective as the traditional visual menu.

In the future, acoustic interfaces with various numbers of simultaneous sound sources should be evaluated. The users' statements and the workload measured in our experiment indicate that menus with a large number of simultaneous sounds could be more confusing and less efficient than menus with just one sound played at a time. Perhaps up to three sounds played at once could have the advantage of enabling a larger information flow than just one sound and also of not being hard to perceive or understand. In such as case, only the information on the current, previous and next item in the menu level would be played to a user.

The awareness of the current position within the menu hierarchy proved to be the biggest disadvantage of the auditory interface compared to the visual interface. Therefore additional acoustic cues and frequent feedback messages should be added to the interface.

The ineffective text input mechanism made the acoustic interface much slower for longer tasks where the users had to compose a message or enter some commands manually. In the future we would like to implement a simple voice recognition system with a very limited functionality (merely the commands for text input) in order to achieve a high recognition in a noisy environment.

We would also like to test different road conditions. The car simulator in our study consisted of a countryside road without other cars or more dynamic obstacles on the road. A city centre road with dense traffic conditions would demand an even higher degree of user concentration and would give us some useful additional results due to the more realistic driving conditions.

References

Algazi, V.R., Duda, R.O., Thompson D.M., Avendano C., 2001. The CIPIC HRTF Database. In: Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics. Mohonk Mountain House, New Paltz, pp. 99–102.

Arons, B., 1992. A review of the cocktail party effect. Journal of the American Voice I/O Society 12 (July), 35–50.

Begault, D.R., 1994. 3-D Sound For Virtual Reality and Multimedia. Academic Press, Cambridge.

Bents, F., 2000. Driver Distraction Internet Forum. From: 〈http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/AskTheExperts.htm#CurrentExpertQuestions〉.

Blaskó, G., Feiner, S., 2002. A menu interface for wearable computing. In: Sixth IEEE International Symposium on Wearable Computers (ISWC 2002), pp. 164–165.

Blauert, J., 1997. Spatial Hearing: The Psychophysics of Human Sound Localization, revised edition. MIT Press, Cambridge, USA.

Bölke, L., Gorny, P., 1995. Direkte Manipulation akustischer Objekte. In: Proceedings of the Software-Ergonomie' 95. Fachtagung des German Chapter ACM und der GI. Teubner, Darmstadt, Stuttgart, Germany, pp. 93–106.

Brewster, S.A., 1994. Providing a structured method for integrating non-speech audio into human-computer interfaces, Ph.D. Thesis, University of York, UK.

Brewster, S.A., Wright, P.C., Edwards, A.D., 1993. An evaluation of earcons for use in auditory human–computer interfaces. In: Proceedings of the INTERCHI, 93 Conference on Human Factors in Computing Systems, Amsterdam, pp. 222–227.

Brewster, S., 1997. Navigating telephone-based interfaces with earcons. In: Proceedings of the BCS HCI'97. Springer, Bristol, UK, pp. 39–56.

Brewster, S., 1998. Using non-speech sounds to provide navigation cues. ACM Transactions on Computer–Human Interaction (TOCHI) 5, 224–259.

Brewster, S., Crease, M.G., 1999. Correcting menu usability problems with sound. Behaviour and Information Technology 18, 165–177.

Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S., 2003. Multimodal 'Eyes-Free' interaction techniques for wearable devices. SIGCHI Conference on Human Factors in Computing Systems 5(1), 473–480.

Cheng, C.I., Wakefield, G.H., 2001. Introduction to head-related transfer functions (HRTF's): representations of HRTF's in time frequency and space (invited tutorial). Journal of the Audio Engineering Society 49 (4), 231–249.

Cohen, J., 1992. Monitoring background activities. In: Proceedings of the First International Conference on Auditory Display. Addison-Wesley, Santa Fé, USA, pp. 499–532.

Cohen, M., Ludwig, L.F., 1991. Multidimensional audio window management. International Journal of Man–Machine Studies 34 (3), 319–336.

Creative Knowledgebase, 2007. From: 〈http://us.creative.com/support/kb/〉

Crispien, K., Fellbaum, K., Savidis, A., Stephanidis, C., 1996. A 3D-auditory environment for hierarchical navigation in non-visual interaction. In: Proceedings of the Third International Conference on Audio Display (ICAD '96), Palo Alto, USA, pp. 18–21.

Dell, W., 1999. The use of 3D audio to improve auditory cues in aircraft. Report, Department of Computing Science, University of Glasgow.

Dolby, 2007. From: 〈http://www.dolby.com/〉.

Drullman, R., Bronkhorst, A.W., 2000. Multichannel speech intelligibility and talker recognition using monaural, binaural and three-dimensional auditory perception. Journal of Acoustic Society of America 107 (4), 2224–2235.

Encyclopaedia Britannica: From: 〈http://www.britannica.com/eb/article-9109387/attention〉.

Frauenberger, C., Stockman. T., 2006. Patterns in auditory menu design. In: Proceedings of the 12th International Conference on Auditory Display (ICAD06), London, UK, pp. 141–147.

Gaver, W., 1986. Auditory icons: using sound in computer interfaces. Human–Computer Interaction 2 (2), 167–177.

Gaver, W.W., 1989. The SonicFinder: an interface that uses auditory icons. Human–Computer Interaction 4 (1), 67–94.

Grantham, D.W., Hornsby, B.W.Y., Erpenbeck, E.A., 2003. Auditory spatial resolution in horizontal, vertical, and diagonal planes. Journal of Acoustic Society of America 114 (2), 1009–1022.

Green, P., 2000. Crashes induced by driver information systems and what can be done to reduce them (SAE Paper 2000-01-C008). In: Proceedings of Convergence 2000 (SAE Publication P-360). Society of Automotive Engineers, Warrendale, PA, pp. 26–36.

Goose, S., Djennane, S., 2002. WIRE3: driving around the information super-highway. Personal and Ubiquitous Computing 6, 164–175.

Hart, S.G., Wickens, C., 1990. Workload assessment and prediction. In: Booher, H.R. (Ed.), MANPRINT. An Approach to Systems Integration. Van Nostrand Reinhold, New York, pp. 257–296.

Hawley, M.L., Litovsky, R.Y., Colburn, S., 1999. Speech intelligibility and localization in a multi-source environment. Journal of Acoustic Society of America 105 (6), 3436–3448.

ICBC, 2001. The impact of auditory tasks (as in hands-free cell phones use) on driving performance. ICBC Transportation Safety Research.

From: 〈http://www.icbc.com/library/research_papers/cell_phones/images/cellphones_impact2.pdf〉.

Jakka, J., 2005. Binaural to multichannel audio upmix. Master Thesis, Helsinki University of Technology, Finland.

James, W., 1890. The Principle of Psychology. Holt, New York, NY, 403pp.

Jin, C., Corderoy, A., Carlile, S., Schaik, A., 2004. Contrasting monoaural and interaural spectral cues for human sound localization. Journal of Acoustic Society of America 115 (6), 3124–3141.

Jones, S.D., Furner, S.M., 1989. The construction of audio icons and information cues for human computer dialogues. In: Proceedings of the Ergonomic Society's 1989 Annual Conference. Taylor & Francis, Reading, England, pp. 436–441.

Klante, P., 2003. Praxisbericht zur Gestaltung auditiver Benutzeroberflächen. Proceedings of the first annual GC-UPA Track, Stuttgart, Germany, pp. 57–62.

Kobayashi, M., Schmandt, C., 1997. Dynamic soundscape: mapping time to space for audio browsing. Human Factors in Computing Systems. In: Proceedings of the CHI 1997, pp. 194–201.

Lee, J.D., Caven, B., Haake, S., Brown, T.L., 2000. Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway. Cognitive Systems Laboratory, Department of Industrial Engineering, University of Iowa. From: 〈http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/27.PDF〉.

LePlâtre, G., Brewster, S., 1998. Designing non-speech sounds to support navigation in mobile phone menus. In: Proceedings of the International Conference on Auditory Display (ICAD2000), Atlanta, USA, pp. 190–199.

Llaneras, R.E., 2000. NHTSA driver distraction Internet forum: summary and proceedings. Driver Distraction Internet Forum. From: 〈http://www-nrd.nhtsa.dot.gov/pdf/nrd-13/FinalInternetForumReport.pdf〉.

Lucas, P., 1994. An evaluation of the communicative ability of auditory icons and earcons. In: Proceedings of the Second International Conference on Auditory Display, Santa Fe, USA, pp. 121–128.

McKnight, J., McKnight, A.S., 1991. The Effect of Cellular Phone Use Upon Driver Attention. National Public Services Research Institute. From: 〈http://www.aaafoundation.org/resources/index.cfm?button=cellphone〉.

NASA TLX for Windows. From: 〈http://www.nrl.navy.mil/aic/ide/NASATLX.php〉.

Navon, D., Gopher, D., 1979. On the economy of the human processing system. Psychological Review 86 (3), 214–255.

Openal, 2006. From: 〈http://www.openal.org/〉.

Pettitt, M.A., Burnett, G., Stevens, A., 2005. Defining driver distraction. In: Proceedings of World Congress on Intelligent Transport Systems, San Francisco, USA.

Pirhonen, A., Brewster, S.A., Holguin, C., 2002. Gestural and audio metaphors as a means of control for mobile devices. In: Proceedings of the CHI 2002, Minneapolis, Minnesota. ACM, USA, pp. 291–298.

Pulkki, V., 2001. Spatial sound generation and perception by amplitude panning techniques. Ph.D. Thesis, Helsinki University of Technology, Finland.

QUIS, 2006. About the QUIS, Version 7.0. From: 〈http://www.lap.umd.edu/quis/〉.

RACER, 2006. From: 〈http://www.racer.nl/〉.

Ranney, T.A., Mazzae, E., Garrot, R.,Goodman, M.J., 2000. NHTSA Driver Distraction Research: Past, Present, and Future. From: 〈http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/233.PDF〉.

Savadis, A., Stephanidis, C., Korte, A., Crispien, K., Fellbaum, K., 1996. A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction. In: Proceedings of Assets' 96. ACM, New York, USA, pp. 117–123.

Sawhney, N., Schmandt, C., 2000. Nomadic radio: speech & audio interaction for contextual messaging in nomadic environments. ACM Transactions on Computer–Human Interaction 7 (3), 353–383.

Schmandt, C., Mullins, A., 1995. AudioStreamer: exploiting simultaneity for listening. In: Proceedings of the CHI 1995. ACM, New York, pp. 218–219.

Sodhi, M., Cohen, J., Kirschenbaum, S., 2004. Mutli-Modal Vehicle Display Design and Analysis, University of Rhode Island Transportation Center, Kingston, RI, USA, A study conducted in cooperation with US DOT. 〈http://www.uritc.uri.edu/media/finalreportspdf/536103.pdf〉 (retrieved 11.11.2006).

Sodnik, J., Sušnik, R., Štular, M., Tomažič, S., 2005. Spatial sound resolution of an interpolated HRIR library. Applied Acoustics 66 (11), 1219–1234.

Stifelman, L.J., 1994. The cocktail party effect in auditory interfaces: a study of simultaneous presentation. MIT Media Laboratory Technical Report.

Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A., 2001. The role of driver distraction in traffic crashes. From: 〈http://www.aaafoundation.org/projects/index.cfm?button=distraction〉.

Tijerina, L., 2000. 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〉.

Vargas, M.L.M., Anderson, S., 2003. Combining speech and earcons to assist menu navigation. In: Proceedings of the International Conference on Auditory Display (ICAD2003), Boston, USA, pp. 38–41.

Vollrath, M., Trotzke, I., 2000. In-vehicle communication and driving: an attempt to overcome their interferences. Center for Traffic Sciences, IZVW, University of Wuerzburg. Germany. From: 〈http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/33.PDF〉.

Walker, B.N., Nance, A., Lindsay, J., 2006. Spearcons: speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the International Conference on Auditory Display (ICAD 2006), London, England, pp. 63–68.

Wenzel, E., Arruda, M., Kistler, D., Foster, S., 1993. Localization using nonindividualized head-related transfer functions. Journal of Acoustic Society of America 94 (1), 111–123.

Wickens, C.D., 1984. In: Parasuraman, R., Davies, R. (Eds.), Processing Resources in Attention. Varieties of Attention. Academy Press, New York, USA, pp. 63–102.

Wierwille, W., Tijerina, L., 1998. Vision in vehicles VI. In: Gale, A., Brown, I., Haslegrave, C., Taylor, S. (Eds.), Modelling the Relationship Between Driver In-Vehicle Visual Demands and Accident Occurrence. Elsevier, USA, pp. 233–244.

WordNet, From: 〈http://wordnet.princeton.edu/perl/webwn3.0?s=attention〉.

Young, K.L., Regan, M.A., Hammer, M., 2003. Driver Distraction: A Review of the Literature, Victoria. Monash University Accident Research Centre, Australia.