Developing a User-Centered Mobile Service Interface Based on a Cognitive Model of Attention Allocation

Julia Niemann, Volker Presse, Jessika Reissland, Anja Naumann

▶ **To cite this version:**

HAL Id: hal-01055461

https://hal.inria.fr/hal-01055461

Submitted on 13 Aug 2014

Developing a user-centered mobile service interface based on a cognitive model of attention allocation

Julia Niemann[1], Volker Presse[2], Jessika Reissland[3], Anja Naumann[1],

[1] Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7,
10587 Berlin, Germany
Julia.Niemann@telekom.de, Anja.Naumann@telekom.de

[2] Berlin Institute of Technology, Chair for Technology and Innovation Management,
Sekr. H 71, Straße des 17. Juni 135, 10623 Berlin, Germany
Volker.Presse@tim.tu-berlin.de

[3] Department of Human-Machine Systems, TU Berlin, Franklinstr. 28/29, 10587 Berlin,
Germany
jre@mms.tu-berlin.de

Abstract. Mobile devices have conquered our daily life. They are carried and used at almost all times and in any situation sometimes even against legal restriction - e.g. in the car. Services running on these devices (e.g. email, text messages, etc.) include graphical (GUI) and voice interface (VUI) causing visual distraction for the driver although they could solely be operated by speech in- and output. As a result services should adapt their interfaces due to the specific usage scenario. The aim of this work is therefore to develop design recommendations based on a cognitive model for the voice user interface taking the particularities of the specific scenario (e.g. reduction of off-road eye gazes) into account. We assume that distraction is mainly due to an increased effort and a decreased expectancy of the VUI compared to the GUI. Design recommendations derived from these considerations will be described by means of a concrete example.

Keywords: Mobile services, in-car infotainment, car services, voice-user-interface

1 Introduction

Nowadays there is an enormous demand for all kinds of services that can be used on portable as well as mobile devices satisfying the users' needs for communication, entertainment, and mobility. These devices and their use are an integral part of today's life in industrialized societies. As a matter of fact it seems to be obvious that these services (e.g. email, music, news feeds, text messages, etc.) also find their way into the vehicle and are more and more used even while driving – careless of any legal restrictions. This leads to the point that the development and design of mobile

and portable services which could also be used in the car should take this specific usage scenario into account.

Within this scenario the driver has to deal with several tasks and is confronted with various interfaces. First of all the interface provided by the car in order to fulfill the primary task [1] of driving the car, secondly integrated devices such as radio, navigation (GPS) devices or high-end entertainment systems and finally portable and mobile devices such as mp3 player, portable GPS devices and mobile phones. All these systems demand attention from the driver in order to be used whereas the primary driving task has to be fulfilled by the driver at all times. This poses on the one hand a large challenge to designers and developers of in-car services, since distraction from the primary driving task has to be kept as low as possible. On the other hand, the specific safety-related driving context also implies that the usage of the infotainment system might generally be interrupted in more demanding or even critical driving situations.

2 Challenge

Usually, services are developed due to design guidelines and ease-of-use principles to improve the use of the service in general and these prototypes are then tested within a driving situation and adjusted afterwards.

Within this paper the authors are proposing an approach based on a cognitive model which takes human capabilities into account in an early phase of the design process. An interdisciplinary team work (consisting of Psychologists, Designers and Engineers) is required to enable user centered engineering in order to create services which minimize the distraction caused by the visual sources [2, 3] and avoid costly adaptation in the late phases of service development.

If attention is needed to be divided between two tasks, Wickens [4] states that those tasks will interfere less if they demand different cognitive resources. Since driving as the primary task is visually and manually demanding, it is most suitable to use speech as the main interaction modality for secondary tasks like interacting with in-car infotainment systems. In line with that, Vollrath & Totzke [5] showed that processing acoustic information while driving has less negative effects on the driving performance than processing of visual information. While attending the auditory channel, drivers can keep their eye gaze on the road. But do voice user interfaces (VUI) actually minimize glancing at in-car displays? Kun et al. [6] examined the influence on visual distraction of the presence of a graphical user interface (GUI) of speech-based portable navigation devices (PND). In case of an available display, more visual distractions compared to a speech based PND without a display were identified although the secondary navigation task was accomplishable with speech interaction only. This leads to the question why users still focus some attention on the GUI in spite of a VUI with which their task is solely achievable. Even more important is the subsequent question how to avoid these visual distractions.

One possibility to avoid eye glances towards the GUI is to cover the display while driving. But this is contradictory to the fact that drivers prefer to have a GUI to get visual information [6]. Hence, taking the GUI away is not a solution. Furthermore, it

is not clear what the reason for the glance to the GUI is. What kind of benefits does a graphical user interface provide compared to a voice interface? Based on a task analysis of the primary and secondary tasks while driving, mainly two situations seem to play an important role for requesting the visual feedback of the GUI: reorientation of the user to figure out at which process stage of the actual task he is and which input possibilities he has. The need for reorientation is a main characteristic of an in-car secondary task as the interaction might be interrupted by the primary task.

3 Attention Allocation between GUI and VUI in dynamic Environments

According to Wickens [7] attention allocation in dynamic environments (like driving a car, or flying an airplane) is influenced by four aspects: salience, effort, expectancy, and value (so called SEEV model). These aspects affect the probability of drawing attention to specific regions (PA) in the following way:

$$P(A) = sS - efEF + (exEX + vV) \tag{1}$$

Salience represents the strength of the stimulus: the more salient a stimulus, the higher the probability of shifting attention to it. The effort of the user (e.g. long scanning paths) that needs to be invested to allocate attention to this stimulus inhibits P(A). Salience and effort are so called bottom-up processes, while expectancy and value are top down processes: if the expected bandwidth of a stimulus is high, the probability of attention shifting increases. Not only a high expectancy but also the importance of the stimulus (value) enhances P(A) in a top-down manor. Up to now, the SEEV model is evaluated for visual attention allocation [8] and can predict percent dwell times for different areas of interest, e.g. while car driving, with correlations from 0.65 to 0.98.

We assume that the effects of salience, effort, expectancy, and value on attention allocation do also play a role in the auditory domain. Therefore, theoretical considerations based on the SEEV model were taken into account to answer the question why the probability of attending the visual display (P(A)GUI) is obviously higher than the probability of shifting the attention to the VUI (P(A)VUI).

As "salience is a maximum for auditory events" [9], there is no benefit for GUIs over VUIs in this aspect.

Considering the importance of the inherent information of GUI, respectively VUI, they seem to have the same quality and therefore it is assumed that they have the same value in the first place. But detracting the visual attention from the road has high costs which in turn implicates a lower value for attending the GUI.

Hence, neither salience nor value give reasonable causes for an increased benefit of allocating attentional resources to the GUI despite safety-critical issues instead of attending the VUI.

Since visual information is constantly present, so that the driver can initiate an attention shift at any time the expectancy to get information that is needed via GUI is

very high. In contrast, a speech output is non-permanent and requires attention in the very moment it is presented, which reduces the expectancy to a value near zero. While in the special case of driving the interaction with the infotainment system is frequently interrupted by the primary driving task, the drivers expediently tend to use the expectancy high-valued GUI instead of the low-valued VUI for information acquisition.

Another difference between GUI and VUI is evident considering the aspect of effort, which is mainly expressed by time and cognitive effort required for attention allocation: a considerably high benefit for displays is the faster detection of information compared to voice outputs, because those require serial listening from the beginning to the end, while the relevant information cannot as easily be picked out as from a GUI. Furthermore, listening to a text generally is more time-consuming than reading a text [10]. Another aspect is that GUIs often provide more information compared to VUIs: different colours, fonts or grouping of information on the display transfers implicit information to the user (e.g. buttons in a specific colour or font mark input options). To present equivalent additional information via voice output, such kind of meta-information has to be provided verbally (e.g. "you have the following options"). This will increase the time needed for information acquisition and will also affect the cognitive workload for the user because he has to keep the verbal information in the phonological loop of his working memory [11]. This again might lead to annoyance and cognitive overload of the user. To avoid annoyance and increasing cognitive workload, it frequently results in design solutions of speech output prompts, which represent only parts of the whole information depicted on the GUI. In turn, this influences the value for the variable expectancy to receive information, because not all the information which is actually relevant for the user might be presented – the bandwidth of information is low. Usually these reduced ranges of prompts are limited to the actual interaction (like system information to give a brief orientation in the menu or a possible voice command that can be used to follow the main interaction path). Obviously there is a trade-off for designers of speech-based in-car infotainment systems: Increasing expectancy by presenting all the information of the screen via speech (increasing the bandwidth of information), which will lead to an increased cognitive and time effort, or decreasing the effort by presenting only a part of information which will decrease the bandwidth of information (decreased expectancy).

Summing up, voice outputs imply higher effort and lower expectancy, which might lead in a higher probability of drivers focusing their attention on the GUI of an infotainment system rather than on a VUI. Based on these considerations, implications for designing infotainment systems are deduced in the following part.

4 Design Recommendations

In a project of Deutsche Telekom Laboratories different mobile services were implemented on a G1 mobile phone. The application includes a graphical user interface as well as a voice user interface and can be solely operated by speech input and output. If the driver intends to give a speech input he has to press a button located

at the steering wheel to activate the automated speech recognition (push-to-activate button). In the development process the VUI was especially designed to gain a lower probability of attention shifts towards the display and as a result off the road. To achieve the objective it was intended to increase the expectancy and to decrease the effort. The "email" application was selected in order to derive design considerations and will be described in the following (Figure 1).

Fig. 1: Screen shots of the email application

4.1 Increasing the Expectancy

As mentioned before, not all information that is represented on the screen is usually available by using speech input. This fact and also the single representation of voice output reduce the user's expectancy to get all information via the VUI. This in turn leads to a lower probability of focusing attention on the screen. To increase the level of expectancy, all information on the screen needs to be available for the driver so it can be requested anytime. To avoid a higher manual effort, an additional functionality was introduced to the push-to-activate button: while the speech recognition is activated with a short-term push of the button, a long-term push of this button will provide more elaborated verbal information. The aim is to represent all information which is present on the GUI – not only information that is related to the actual interaction. Thus, there are two different output levels:

- direct speech prompts for the main interaction path (system initiated speech output)
- long-term push prompts to get all information presented on the screen (driver initiated speech output)

This concept changes the way voice output is usually provided to the driver. Typically speech output is system-driven while in the system described the driver initiates the request of speech output. With the help of the long-term push on the push-to-activate button, the driver can get the desired information whenever he intends to do, which increases the expectancy of this information up to 100%. The user is now able to orientate him- or herself again after interruptions through the primary driving task. Furthermore, missing a system initiated direct speech prompt is no longer fatal, because relevant information can be repeated on the long-term push

prompt. This also reduces the cognitive demand of attending and memorizing the system output.

4.2 Decreasing the Effort

To present all the information via voice output increases the expectancy, but this in turn involves an increased time effort for speech output. In order to decrease that amount of time and to provide all the information presented on the screen for the driver without provoking an increased mental workload, there are three different kinds of design solutions that we developed for the VUI:

- non-speech sounds
- up-tempo speech
- acoustics instead of visualization (semantically enriched information representation)

Non-speech sounds. According to Brewster [12] "icons can present information in a small amount of space compared to text; non-speech sounds can present information in a small amount of time as compared to speech". There are three main types of nonverbal sounds: auditory icons, earcons and spearcons.

Auditory icons are metaphorical representations of a word or a concept [13]. Important for the use of auditory icons is a strong intuitive link with the word that should be presented through the auditory icon. The stronger the association with the word the less learning is required.

In the project presented, we tried to replace speech output with auditory icons whenever there was a strong and clear associated sound available. For example, instead of the speech output: "Your Mail has been sent" there is a wind sound played - comparable to an object flying through the air.

Earcons are short musical motives that represent menu items. They differ in the combination of rhythm, pitch, timbre, register and dynamics.

To achieve the same rapid information acquisition about the current status in the menu as in case of glancing at the screen earcons were used. For example: every menu item on the start screen (email, sms, telephone and news) has its own sound motive. The four sound motives exhibit a high differentiation. However, to represent the single steps of the main interaction path (e.g. writing an email) within the specific menu items, the particular sound was varied by changing the pitch. The pitch rises with every new step that brings the user to the end of the operation. Since absolute data for pitches is hard to perceive for humans, we present every increased pitch with the sounds of the preceded interaction steps. This in turn not only reduces the cognitive workload caused by detecting a difference between the sounds, it also enables the user to associate the sound with one of the interaction steps by counting the number of sounds. However, it is important to take into account that earcons "are abstract so their meaning must always be learned" [12]. Therefore the earcons that will be used for orientation in the menu in the email application initially need to be presented with a speech prompt.

Spearcons are speech audio cues that are sped up to the point speech is no longer

comprehensible. As the sound of the accelerated word is unique – like a fingerprint – it is very easy to produce sounds that are different enough from each other. Spearcons as well as earcons could also be used for orientation in the menu and also need to be learned before presenting them on their own.

One important aspect of non-speech sounds is the consideration of annoyance and user preference. Especially since in-car infotainment systems are comfort systems which add various requirements to the design in terms of acceptance and joy of use. To decide which kind of nonverbal sound will finally be applied to the email application as well as performance indicators, user acceptance has to be tested.

Uptempo Speech. Speeding up the paste of the voice output is another design solution for the VUI to decrease the time effort of speech output. This in fact should solely be used with content that the user has knowledge of. In the described project we sped up content that had already been heard via the direct system initiated speech output. For example: after opening a new email the user gets a direct feedback from the system. The voice output reads the email header via text -to-speech. If the user holds the push-to-activate button he gets all information on the screen and so the email header can be requested again, but this time it plays faster.

Acoustics instead of visualization. As mentioned before, by using different colours, fonts or grouping items together, a GUI can represent additional information to the user. In the described project, different voices and pauses were used to represent such indirect information. Buttons or interaction elements on the GUI were magenta coloured. By drawing visual attention to the screen, the user knows which next interaction options are available. Additionally, he or she knows which speech commands he can use devoid of being told "you have the following options…" (meta-information). As mentioned before, we do not only want to let the user know via direct speech prompts which are the next possible interaction steps to follow the main path of interaction (as it is frequently used in speech based infotainment systems). We also want to enable the user to hear all the speech commands he or she can use. By using a different voice for interaction elements than for other information, the information is implicitly clustered and does not require further explicitly verbal announcing. This has a positive effect on the time effort, because prompts generally shorten while the content of information stays the same.

Variations in the duration of pauses were used to represent the graphical grouping of elements in different locations on the screen. To give an example: in the email inbox the system pauses are shorter between reading out the possible interaction elements "replay" and "forward" compared to the length of the pause between these two options and the navigation interaction elements "next" and "previous".

Another way of presenting the implicit information given on the screen can be found in spatial representations of nonverbal sounds or speech. For example, instead of using different voices for interactional elements as well as other information, this content could be represented via different spatial alignments of these speech prompts. A critical aspect by using spatial representation of sounds for in-car infotainment systems could be seen in the stronger orientating responses towards the source location of acoustic cues rather than visual cues [14]. Since it is not intended to affect

these orientation responses while driving, it was initially renounced to semantically enrich sounds and speech prompts by spatial representation for the email application.

5 Conclusions and Implications

The present project focuses on the question of how to prevent drivers from allocating their attention to displays of in-car infotainment systems if a VUI is available. The design recommendations that were made by taking the SEEV-Model [7] into account allow the integration of human factors in an early phase of the engineering process. Furthermore in a next step the design recommendations are going to be tested in a driving simulator study to iterative verify if the design goals (to develop an intuitive and safe interface for in-car use) are achieved. Aim of further studies will be to answer the question, which of the above mentioned methods for decreasing the time effort can substitute graphical information best.

6 References

1. Bubb, H.: Fahrerassistenz - primär ein Beitrag zum Komfort oder für die Sicherheit? : Der Fahrer im 21. Jahrhundert, Vol. 1768. VDI Verlag, Düsseldorf (2003) 25-44
2. Wierwille, W., Tijerina, L.: Eine Analyse von Unfallberichten als ein Mittel zur Bestimmung von Problemen, die durch die Verteilung der visuellen Aufmerksamkeit und der visuellen Belastung innerhalb des Fahrzeugs verursacht wird. Zeitschrift für Verkehrssicherheit (1995) 164-168
3. Wierwille, W., Tijerina, L.: Darstellung des Zusammenhangs zwischen der visuellen Beanspruchung des Fahrers im Fahrzeug und dem Eintreten eines Unfalls. Zeitschrift für Verkehrssicherheit (1997) 67-74
4. Wickens, C.: Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science **3** (2002) 159-177
5. Vollrath, M., Totzke, I.: In-vehicle communication and driving: An attempt to overcome their interference. Driver Distraction Internet Forum sponsored by the United States Department of Transportation (2000)
6. Kun, A.L., Paek, T., Medenica, Z., Memarovi, N., Palinko, O.: Glancing at personal navigation devices can affect driving: experimental results and design implications. Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, Essen, Germany (2009)
7. Wickens, C.D., Goh, J., Helleberg, J., Horrey, W.J., Talleur, D.A.: Attentional Models of Multitask Pilot Performance Using Advanced Display Technology.

Human Factors: The Journal of the Human Factors and Ergonomics Society **45** (2003) 360-380

8. Wickens, C., McCarley, J.: Applied attention theory. CRC Press, Boca Raton, FL (2007)

9. Wickens, C., McCarley, J., Alexander, A., Thomas, L., Ambinder, M., Zheng, S.: Attention-Situation Awareness (A-SA) Model of Pilot Error. In: Foyle, D., Hooey, B. (eds.): Human performance modeling in aviation. CRC Press, Boca Raton, FL (2007) 213-239

10. Kozma, R.B.: Learning with Media. Review of Educational Research **61** (1991) 179-211

11. Baddeley, A., Hitch, G.: Working memory. In: Baddeley, A., Hitch, G., Bower, G. (eds.): Recent advances in learning and motivation, Vol. 8. Academic Presse, New York (1974) 47-90

12. Brewster, S.: Nonspeech auditory output. In: Jacko, J., Sears, A. (eds.): The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Lawrence Erlbaum Associates, Mahwah, NJ (2002) 220-239

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

14. Proctor, R., Tan, H., Vu, K., Gray, R., Spence, C.: Implications of compatibility and cuing effects for multimodal interfaces. Lawrence Erlbaum Associates, Mahwah, NJ (2005)