Proceedings of the 5th International Conference on Automotive User Interfaces and
Interactive Vehicular Applications (AutomotiveUI ,13), October 28–30, 2013, Eindhoven, The Netherlands.

Mostly Passive Information Delivery in a Car

Tomáš Macek, Tereza Kašparová, Jan Kleindienst, Ladislav Kunc, Martin Labský, Jan Vystrčil
IBM Research
V Parku 2294/4
Czech Republic
+420234201490
tomas_macek@cz.ibm.com

ABSTRACT

In this study we present and analyze a mostly passive infotainment approach to presenting information in a car. The passive style is similar to radio listening but content is generated on the fly and it is based on a mixture of personal information (calendar, emails) and public data (news, POI, jokes). The spoken part of the audio is machine synthesized. We explore two modes of operation. The first one is passive only. The second one is more interactive and speech commands are used to personalize the information mix and to request particular information items. Usability and distraction tests were conducted with both systems implemented using the Wizard of Oz technique. Both systems were assessed using multiple objective and subjective metrics and the results indicate that driver distraction was low for both systems. The users differed in the amount of interaction they preferred. Some users preferred more command-driven styles while others were happy with passive presentation. Most of the users were satisfied with the quality of synthesized speech and found it sufficient for the given purpose. In addition, feedback was collected from the subjects on what kind of information they liked listening to and how they would have preferred to ask for specific types of information.

Categories and Subject Descriptors

H5.2. Information interfaces and presentation: User Interfaces.

General Terms

Design, Economics, Experimentation, Human Factors,

Keywords

Interactive, Radio, Voice, Information delivery, User Interface

1. INTRODUCTION

Most personal assistants present in today's smart phones (e.g. Apple's SIRI [8], Google Now [9] or Dragon Mobile Assistant [14]) are driven by the activity of their user. The user asks for information (e.g. **H**uman: "What is the traffic ahead?") and the system responds (**C**omputer: "Traffic jam in five miles"). On the other hand, a radio-style presentation constitutes a completely opposite approach where the user passively listens to information delivered by the system. In the case of conventional radio stations, users manifest their preferences simply by tuning to a different station.

The amount of driver distraction is of primary concern when developing automotive UIs. While cell phone usage is generally perceived as dangerous and is regulated in most countries, listening to radio is considered safe. Some studies [6] report that listening to the radio even improves safety of long distance driving by keeping the driver alert. One of the reasons is that the information carried by radio is not critical for the driver and therefore does not require high level of concentration. The user can stop listening at any moment without even subconscious feeling that it might not be possible to continue the activity later.

In this study we explore the method of a mostly passive approach to presenting information in a car. The passive style is similar to listening to a radio. The presented audio stream is a mixture of music and spoken information. The spoken information is machine synthesized by a high quality TTS (Text To Speech) module.

The goal of the study is to present the collected evidence about how users perceive this type of presentation, what are their preferences and how they wish to interact with the system.

2. RELATED WORK

Implementations of customized audio stream generation can be found in many music delivery services including Spotify, Pandora, or iTunes. Music selection schemes range from manually created play lists to automatic systems. These services reflect user preferences and also involve a social aspect (e.g. recommendation or preferences of friends). Recently, Microsoft and Google [9] announced their improved services as well.

Though the above mentioned services often focus on music alone, other providers like large news agencies BBC, CNN and others provide their content in a form of podcasts. The textual information is translated to spoken form by professional speakers.

The systems providing multimedia content on demand are becoming available in today's cars; for instance Pandora can be played in vehicles including BMW, Buick, Ford, GMC, Honda, Hyundai and many others. Many in-car infotainment systems also adopted the use of voice commands and utilize information from the Internet [12][15][10][11].

Audio books are one particularly popular form of entertainment on the go. Machine translated books are available but rarely used. We can hypothesize that the users are more sensitive to quality of voice for this type of content.

AutomotiveUI'13, October 28-30 2013, Eindhoven, Netherlands.

ACM 978-1-4503-2478-6/13/10 ...$15.00
http://dx.doi.org/10.1145/2516540.2516577

3. CONCEPT

This paper proposes the use of enhanced radio-style UIs for in-car infotainment systems. The presented approach mimics conventional radio broadcast but presents information from both public and private sources (e.g. customized music, news and weather combined with calendar and email summaries). Secondly, the user is allowed to verbally interact with the UI e.g. by requesting a topic change, by skipping backward or forward in the presentation, or by asking questions about the content presented. The voice interaction is, however, only a complement of the passive infotainment stream, not the major driving mechanism of the system.

The information is presented to the user using synthesized voice. The approach relies on the quality of speech synthesis and willingness of the user to accept it but permits high flexibility in selection of the information to be presented.

The presented information stream is sparse with the intervals between the spoken segments filled with music. The content is configurable and can be adapted both automatically based on observing habits of the user or via voice commands. The audio stream is accompanied by a very simple visual presentation which should make it easy for the user to recall the UI status at any time. The UI only presents static text or pictures in order to minimize visual distraction.

We summarize selected presentation patterns below:

Status summary. User is given a brief summary of what is new; e.g. C: "It is 5:30 PM, sunny 23°C, you have 5 new messages and one missed phone call. The next meeting in your calendar is: Update with boss in 1 hour". The way how and when the summary is presented is customizable (e.g. after starting the car in the morning or every time the user enters the car).

Context-aware information. The system provides information based on immediate context: e.g. when fuel level is running low and the car passes by a preferred gas station (e.g. C: "We have fuel for the next 20 miles and a Shell station is approaching").

Guide style information. Guide style is another example of context-aware information. When the system detects an unusual route, it informs the user about important sight-seeing points; e.g. C: "The castle on your left is named Karlstejn".

Multiple moderators. Two voices help the user distinguish between different types of presented information; e.g. news are read back by a different moderator than calendar and messages.

Presenting information at the right time. The announcement of certain types of events (emails, SMS, Facebook) is queued as an item to be presented by moderators instead of interrupting the radio stream. When road conditions are recognized as difficult, less distracting content is presented (music or even silence).

Moderators talking among themselves. E.g. C1: "Do you think that Thomas knows that we are running out of gas?" C2: "Oh yes, he must know. He'll stop in a minute; there is his favorite OMV in one mile".

3.1 Discussion

As stated before, the presentation style was tailored for the in-car environment, bearing in mind not only usability and efficiency aspects but also aspects of driver distraction. It can be disputed whether a graphical user interface (GUI) is needed for such system at all. GUI provides a parallel form of presenting information which is more persistent and well complements voice output. Users can double-check what has been said or retrieve the information with some delay. It improves the feeling that information can be retrieved later. On the other hand, it introduces visual distraction (see e.g. [7]).

The main benefit of content creation on the fly using speech synthesis is that the system can quickly adapt to user feedback. It can also reduce the redundancy of information based on knowing what has already been communicated to the particular user. Unlike standard radio broadcasting, which keeps repeating the top news, the proposed system presents most pieces of information just once.

The user can however always request the last piece of information to be repeated or presented in a more detailed form using voice commands like H: "Repeat it" or H: "What was that about weather?". This option should remove potential stress related to the concentration on a topic of interest which could vanish completely when not paying constant attention. When using the request-response systems, the user has to know what domains are covered by the system. On the other hand, the Radio continuously presents information from covered domains, acting to some extent as a teaching tool, teaching by way of examples showing what can be retrieved.

4. IMPLEMENTATION

We used the Wizard of Oz (WOZ) technique to evaluate the basic concepts described in the previous section. For the purpose of formal evaluation, we created a set of audio-video recordings covering several infotainment scenarios and a set of prefabricated responses to anticipated user requests. Confirmation prompts were created to provide quick feedback to the user prior to finding a suitable audio-video snippet which should be played in response. Music tracks were shortened to approximately 40 seconds to fit the testing time limits. The length of prompts varied according to the topic (from 2s to 15s). The video part of a recording was presented on a 7" monitor (MIMO720F) located to the right of the steering wheel. The video playback was controlled by the test supervisor from behind the scene. We used the WinAmp media player with a control panel located on a separate screen so that the user could not see the WOZ interventions.

The visual part of the UI had a simple design. Figure 1 shows examples of the application GUI. When reading the news, the screen showed a single static picture related to the news. Played songs were accompanied by a picture of the CD cover and the interpreter name and song name. The screen contained only static background pictures for all other situations.

Figure 1 The examples of the graphical user interface

The formal tests where conducted in Czech with native Czech subjects. We used a Czech female speaker Zuzana of the Nuance Vocalizer for Automotive SDK v5.3.3 to synthesize all prompts.

5. EXPERIMENT SETUP

The primary task of the experiment was driving a low fidelity car simulator of the Lane Change Test (LCT) made according to the ISO 26022 procedure [5]. Subjects drove on a three lane straight road and changed lanes based on the directions given by signs on both sides of the road. One track of the LCT driving simulator was 3 km long and the subjects were asked to maintain constant speed of 60km/h, resulting in a 3 minute drive. Since this time was too short to evaluate the effects of simulated radio broadcast, we modified the standard procedure and asked our subjects to continue driving for 3 subsequent tracks. Adjacent tracks were connected to one another through 180° curve segments which were excluded from evaluation. Hereby, the time for conducting one experiment increased to 9 minutes. We also shortened all songs played by the Radio to 40 seconds which allowed us to fit more interaction into a single experiment. To analyze driving performance, we adapted the ideal path as described by ISO 26022. The evaluated driving performance statistics included the mean deviation of lateral car position (MDev), standard deviation of lateral position (SDLP) and response time to lane change signs. They were computed based on simulator logs for all three tracks of an experiment using custom scripts.

For each experiment, we performed one of the two tasks described below. The first task represented passive usage only. The subjects listened and optionally watched the application but they did not interact with it. The second task represented interactive usage. The users were asked to control the application using voice commands of their own choice at any time, as we wanted to learn what kind of information or actions they would typically request, and how they would ask for them. The topics for both tasks' scenarios were chosen to comprehend the most expected topics e.g. news, sms/e-mail, calendar, traffic info, weather, POI(gas station) and jokes. Each topic was interspersed with music.

We ran the WOZ tests with 13 users, each conducting both tasks (within testing). The order of tasks was counterbalanced to compensate for a possible learning effect. Prior to conducting the evaluated tasks, each subject practiced driving the simulator without any secondary task. They conducted three single track rides. The recording of the first one was not used and its purpose was purely for training purposes. The driving on the second track was used to adapt the ideal track for the particular user. The third track data were used as a reference. In addition to evaluating their objective driving performance using the above mentioned statistics, the test subjects were also asked to fill out several questionnaires after each type of task (undistracted, passive, interactive). We used the standard SUS [3] augmented by several questions and SASSI [2] to collect usability feedback. DALI [1] was used to collect subjective feelings of distraction.

6. EVALUATION

The tests were conducted with 13 subjects (5 female, 8 male), age 25 to 55. All participants had high school education or higher, all were drivers, 77% of them were driving at least several times a week. We asked the test subjects questions about their habits of using electronic devices in the car: 85% of participants listened to radio at least every other drive; 23% of the participants tuned the radio rarely (less then once per drive), 46% tuned at least 2 times per drive, 31% tuned more often.

6.1 Usability observations

We used SUS and SASSI questionnaires to collect usability evidence about the method. The users where asked to answer the questions shortly after completing both tasks.

System Usability Score (SUS) is a simple, ten question test giving a global view of subjective assessment of usability. On the scale 0-100 the application was rated 80.9 (standard deviation was 10) which is considered acceptable (or between good and excellent) according to [4].

SASSI [2] is a test designed to assess speech applications. The results of the test are summarized in Figure 2 (larger means better for most of the factors except for 'Annoyance'). Some of the factors are influenced by the WOZ style of the experiment, particularly 'Habitability' was influenced by limited variability of the prerecorded prompts, 'Speed' and 'System response accuracy' was more assessment of the test conductor then the application concept. Ratio of 'Likeability' and 'Annoyance' indicates positive feedback from the users. The 'Cognitive demand' factor is discussed in detail in Section 6.2.

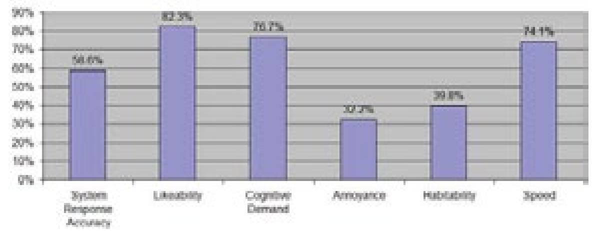

Figure 2 SASSI - Subjective Assessment of Speech System Interfaces

We monitored the frequencies of subjects requesting certain types of information or actions from the system while driving. We also asked questions about the requested functionality after the test. The questionnaire results are summarized in Figure 3. The test subjects rated on a 1-5 scale the functions they would like to have as part of the tested system. Not surprisingly, the test subjects valued travel-related information (traffic and POI) and communication and calendar features. Entertainment options apart of music and news where less valued.

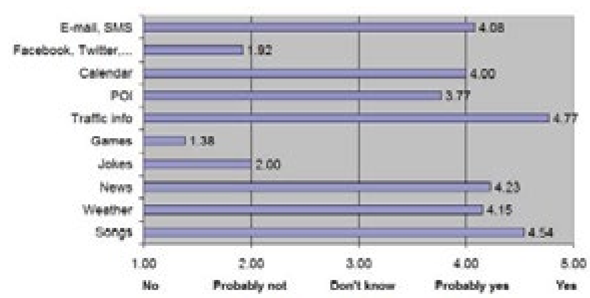

Figure 3 The topics which the testing subjects would like to have available as part of the system.

We also asked the participants whether they found it useful to have the audio broadcast accompanied by visual presentation on a screen. On a scale of 1-5, the need for screen was only rated 2.38, which suggests that an audio-only version of the UI may be a viable option (participants were however only exposed to a version with GUI). The frequency of utterances in active task depended on the participant. The natural form of utterances prevailed.

6.2 Distraction

Distraction is obviously a key aspect of in-car application assessment. We measured distraction both objectively by means of the LCT statistics and subjectively by asking the tested subjects questions about their feeling of distraction. Statistical significance is reported with α=0.05 in this paper.

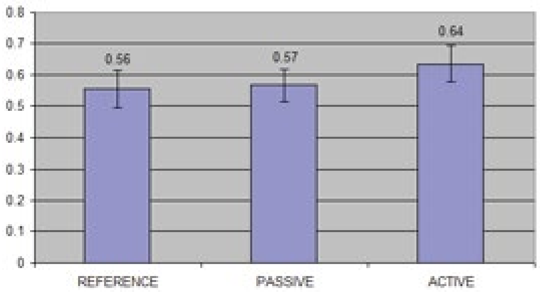

Figure 4 Distraction expressed by mean deviation during LCT

Figure 4 depicts the mean deviation (MDev) in meters measured separately for the 'Passive' and 'Active' modes of operation. As expected, the driver's distraction is slightly higher when performing a secondary task. The active task was more distracting then the passive one. However the impact of the secondary task was very low and we observed marginal statistical significance (p=0.048) when comparing the average MDev values between the 'Passive' and 'Active' tasks. We observed no statistically significant differences when comparing the values of SDLP and average reaction times.

We also asked the test subjects to fill in the DALI questionnaire after conducting each test. The DALI test provides subjective results but helps understand what the components of distraction are. The users rated overall user experience, not just the secondary task alone. The results are summarized in Figure 5. The differences between the passive and active modes are not statistically significant apart of the auditory demand.

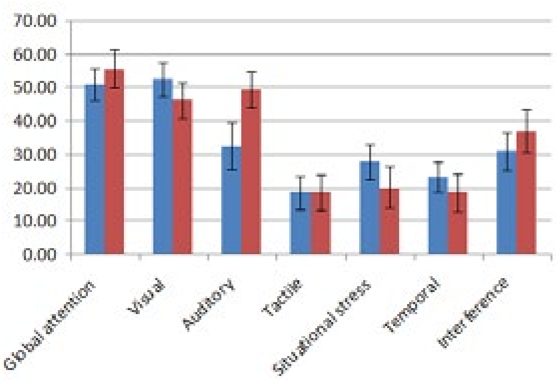

Figure 5 DALI results – the components of the distraction; passive and active modes shown in blue and red, respectively.

7. CONCLUSION AND FUTURE WORK

We introduced a UI concept of the "Mostly Passive Information Delivery" for in-car infotainment systems that enhances the common radio-style presentation with customizability and the possibility to interact with the infotainment stream in a limited manner. The WOZ study conducted on 13 participants provides encouraging results reporting both good acceptance and low distraction levels measured both objectively and subjectively. We also presented a summary of functions rated as useful by our test subjects. We believe that the presented approach can be viewed as a lower-distraction alternative to personal digital assistant systems, which makes it more suitable for use while driving.

8. REFERENCES

[1] Pauzié, A. 2008. A method to assess the driver mental workload: The driving activity load index (DALI) in *IET Intelligent Transport Systems* 2, 4. (Dec 2008), 315 – 322.

[2] Kate S. Hone and Robert Graham. 2000. Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). In *Natural Language Engineering*. 6, 3-4 (September 2000), 287-303.

[3] Brooke, J. 1996. SUS-A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. *Usability Evaluation in Industry, 189*, 194. London: Taylor and Francis

[4] Brooke, J.: SUS: A Retrospective, In *Journal of Usability Studies*, Volume 8, Issue 2, February 2013, pp. 29 - 40

[5] Mattes, S. 2003. The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In *Proc. of the Annual Spring Conference of the GFA/ISOES*.

[6] Dibben, N. and Williamson, V. J. 2007. An exploratory survey of in-vehicle music listening. In *Psychology of Music*, 35(4), 571-589.

[7] Cuřín, J., Labský, M., Macek, T., Kleindienst, J., Young, H., Thyme-Gobbel, A., Quast H. and König, L. 2011. Dictating and editing short texts while driving: Distraction and task completion. In *Proc. 3rd Intl. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 13-20). ACM.

[8] Apple Inc. 2013. *Use your voice to do even more with Siri*. http://www.apple.com/ios/siri/

[9] Google Inc. 2013. Google Now. The right information at just the right time. http://www.google.com/landing/now/

[10] Ford Motor Company. 2013. SYNC AppLink http://ford.com/technology/sync/features/sync-app-link/

[11] Mercedes-Benz UK Ltd. 2013. *COMAND online* http://www2.mercedes-benz.co.uk/content/unitedkingdom/mpc/mpc_unitedkingdom_website/en/home_mpc/passengercars/home/new_cars/models/c-class/_w204/facts/overview/communications.0003.html

[12] Daimler AG, 2013. *Mercedes-Benz presents new COMAND Online features at CeBIT show* http://www.daimler.com/dccom/0-5-7153-1-1580000-1-0-0-0-0-1-8-7145-0-0-0-0-0-0-0.html

[13] Road vehicles-Ergonomic aspects of transport information and control systems-Simulated lane change test to assess in-vehicle secondary task demand, International Standard ISO/DIS 26022:2010

[14] Nuance Reveals Dragon Mobile Assistant, 2012, http://www.dragonmobileapps.com/beta-launch/

[15] BMW, Message Dictation,http://www.bmw.com/com/en/insights/technology/connecteddrive/2010/infotainment/mobile_office_telephone/message_dictation.htm