ABSTRACT

Manuals were designed to provide support and information about the usage and maintenance of the vehicle. In many cases user’s manuals are the driver’s only guidance. However, lack of clarity and efficiency of manuals lead to user dissatisfaction. In vehicles this problem is even more crucial given that driving a motor vehicle is, for many people, the most complex and potentially dangerous task they will perform during their lifetime. In this paper we present a voice interfaced driver manual that can potentially fix the deficiencies of its alternatives. In addition we aim to provide a case for the integration of such technology in a vehicle to reduce driver distraction, increase driver satisfaction, and manual usability, while also benefiting Original Equipment Manufacturers (OEMs) in lowering costs and reducing the documentation process.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentations (e.g., HCI)]: User Interfaces, user-centered design, speech recognition, voice I/O
D.2.2 [Software Engineering]: Design Tools and Techniques – User Interfaces

General Terms
Documentation, Design, Human Factors.

Keywords

1. INTRODUCTION

Traditionally, vehicle technical documentation has assisted drivers when malfunctions occurred or when drivers needed support to perform tasks in which they were not familiar. Tasks such as starting the car engine, checking the oil-level, replacing a headlight bulb, or even setting up the clock are examples of topics in which a driver may look to a manual for instruction. However, in using paper manuals, drivers are limited. Depending on their ability to find the desired information and the clarity of the manual’s structure, searching a manual can be a frustrating and inefficient task for the driver.

On the other hand, advances in wireless technologies and Mobile Internet Devices (MIDs) allow for distributing real-time multimedia information to any portable device. This trend has also reached the automotive cockpits in the form of In-Vehicle Information Systems (IVISs). Digital media such as pictures, video and audio can now provide driver-relevant information, entertainment and productivity [1]. But multimedia information has also proven to be the source of major driver distraction since their interfaces are not optimized for driving performance [2][3][4], inputs and outputs differ greatly between models [5], and these complications affect drivers increasingly as they age [6].

Current interactions with a vehicle’s manual are unnatural, since many questions arise while the driver is operating the vehicle. It is obvious however, that performing two tasks that require visual
attention in a moving vehicle is not only distracting, but also dangerous [7][8]. The vehicle, therefore, needs to be parked while consulting the manual because conventional printed automotive documentation is focused more on the vehicle than on the driver. The printed design is in obvious conflict with the user’s primary goal, which is to drive, rather than to learn how to drive. This is a problem that was addressed in the late 1980s [9], but that has not been resolved.

A driver wants the ability to understand and solve problems when they occur and is willing to be educated in more efficient ways of driving and using vehicle functions. However, currently the driving process has to be interrupted to request information from the printed manual, which in some cases cannot be found at a hands reach, but in the trunk space or some other compartment of the vehicle. This discrepancy is believed to generate extreme dissatisfaction [10].

This paper presents a conversational voice interfaced user help that allows users to speak questions naturally. It is believed that the integration of such a system in a vehicle with the appropriate front-end interface will reduce user distraction while increasing user satisfaction. Additionally, it is hypothesized that such a system provides the opportunity to reduce the documentation process and lower the cost of user manual production centralizing the content in a cloud server.

2. BACKGROUND

Documentation of the deficiencies of traditional paper-based manuals can be dated back to the late 1980s [9][11]. Printed user manuals lack portability, often suffer inaccuracy or present outdated information. In the case of vehicle manuals, users find themselves consulting a manual that doesn’t reflect the vehicle. In addition, the growing complexity of vehicles makes it more difficult for OEMs to maintain manuals, keeping them accurate and up-to-date. Consequently, a number of studies emphasizing digitalization of the manual [11], improvements in the usability of manuals, and more user-centered documentation design, took place during the 1990s [10][12][13][39].

While these studies aimed for more efficient and clear manuals, the development of digital technologies also played a big role in the way manuals were constructed, consumed and understood. The steady increase of Internet usage and the influence of digital mass media produced a major shift into digital and online manuals. Most of the automakers now offer rich media contents through their official websites including user-oriented vehicle information, tutorials, and technology descriptors. This article will distinguish two kinds of Internet vehicle manuals, offline and online.

2.1 Offline Internet manuals

Offline manuals are digital copies of printed manuals, quick reference guides, or brochures that can be downloaded from an automaker’s web portal, usually in PDF file format. A survey of the websites of the 30 biggest automakers worldwide, including European, American and Asian companies was conducted, and it was found that the majority offer digital documents that can be downloaded either in the form of a schematic brochure or a complete copy of the printed vehicle manual. Although more than 50% of the websites provide this information publicly [14-20], some require registration and ownership of a vehicle through identification of a customer name and VIN number [21-24].

2.2 Online Internet manuals

Online manuals typically present information enhanced with hypertexts, images, video, audio and flash-based interactive animations to instruct users on key technology advances, ergonomics, infotainment, security, services, navigation and integration of consumer electronics. While some automakers make public only a technical lexicon as a reference of the branded technology features [25-27], others include an array of video tutorials [14][28] and even interactive material that require active participation of the user [29][30], with the subsequent increase of interest and learning performance. Internet development and the adoption of Web 2.0 standards have led to a shift towards increased user satisfaction, trying to improve the drawbacks of paper-based manuals.

Although in-car communications are improving, due to advances in long distance wireless communication technologies, such as WiMAX [31][32], online rich-media content is currently not capable of meeting user expectations in a moving vehicle. At least until telecom companies provide the needed infrastructure. Furthermore, web rich-media manuals present enormous adaptability issues when it comes to user interfaces (screen size, fonts, input methods), compatibility (audio and video codec, updates, flash integration) and security concerns, as attacks to vehicle systems are demonstrated to be uncomplicated [33][34]. Automakers have, however, found a way to deliver rich-media user help content in the vehicle.

2.3 Multimedia user manual DVDs

Chrysler recently decided to integrate a user manual DVD in the 2010 Chrysler, Jeep and Dodge vehicles [35], substituting the printed manual. The manual can be consulted via the IVI system. Other auto brands have opted for a similar alternative as a way of delivering rich-media content to the consumer on a standard inexpensive platform. However, this solution still has limitations as this media is presented on a static platform that has no integration with the vehicle.

2.4 In-vehicle user manual applications

In 2008, BMW introduced a new concept, the Integrated Owner’s Manual, IBA for the German acronym [36]. The BMW approach consists of an in-vehicle application that delivers rich-media content through the navigation display and is browsed via the iDrive controller located next to the gearshift. Many services and features integrated in the vehicle are linked to the manual for reference making it accessible through a gesture on the iDrive controller. This is particularly useful with new built-in technologies like the BMW Night Vision Display or hybrid power distribution systems, which drivers may have never seen before, but also when warnings and check-control alarms flash on the dashboard. Quick access to the user manual provides video animation descriptions as well as image and text explanations on how to behave in these situations.

Although vehicle user help systems have gained improvements in usability and a broader media distribution due to digital media, many of the current commercial systems have not provided an effective and natural way of consulting the vehicle user manual while using the car.
Research on auditory display for driving has been mainly focused on collision warning signals [1] and only recent studies on in-vehicle information presentation have examined other actions such as using cell phones [2], navigation systems [37] or browsing infotainment content. The large effect that speech commands, and voice interfaces in general, have on the usability of the vehicle has been reported in a study [38] that points out how many real-life users might not consult the manual.

3. MOTIVATION

Our work with the voice interfaced user manual aims to provide a driver-centered interaction method for vehicle documentation by introducing an interactive driver assistant that can process natural speaking. This method could possibly decrease the cognitive load of generating questions by making the system smart enough to understand any natural variation of the question. It is believed that drivers will be more willing to consult the use of features of their vehicle when they are provided with a convenient, non-distractive interface that does not require extra cognitive effort. For the automotive industry it is crucial to overcome the “underuse” of their products that happens when there is a mismatch between the features provided by the car and the features utilized by the driver.

An increasing number of drivers with varying backgrounds and skills are exposed to extensive sets of in-vehicle technologies. Some of these technologies and systems are crucial for driver security, communication and driving efficiency. While a large number of features can be operated without the need of any actions by the driver, other systems need to be manually triggered, therefore should be understood before operating the vehicle.

Manuals are the primary and in most cases only source of instruction and documentation that is available to the driver in daily situations. Therefore it is important for vehicle manuals to be clear, efficient, and usable. Paper-manuals are not designed for use while driving, however situations that require the support of the user manual occur during this time. In addition, it is hypothesized that natural interactions will reduce driver distraction and will increase user satisfaction. In the following section, a voice user interface that allows for natural interaction with an in-vehicle user manual is presented.

4. VOICE INTERFACING THE VEHICLE MANUAL

The concept and design of a voice-activated vehicle manual is presented here. Instead of flipping through a paper manual to learn about the vehicle, a user can simply ask the system verbally a direct question. The voice interfaced vehicle manual takes user input in natural form, i.e. no need to convert queries to something more suitable, and returns to the user a set of results. The manual’s database architecture was developed using the Answer’s First methodology. The VoiceXML Server was used to house the application and the front-end Voice User Interface (VUI) was developed using VoiceXML.

4.1 Overview of Answers First and iTECH

Answer’s First (A1) is a conversational question answering methodology that removes the need for language preprocessing [40]. A1, unlike other methodologies, allows users to speak questions naturally without the need to translate or modify questions before query execution. A1 was first developed for use in iTECH, a conversational interactive agent developed to provide technical assistance. The term iTECH stands for Interactive Technology Assistant. The job of the assistant is to converse with the user of the system, in this case the system being the car. iTECH prompts the user to ask a question and returns an answer.

Results from studies in iTECH suggests that A1 performed well given longer natural queries and returned the correct answer despite a large number of unique queries [40]. Unique queries are defined as questions that participants asked iTECH that were not in the original list of questions stored in the database. The study revealed the flexibility of the A1 methodology as among the 298 spoken queries submitted to iTECH, up to 115 were unique from each other and from the questions in the Questions Table [41]. Despite the large number of unique queries iTECH yielded a success rate of 89.60% [41]. This pointed out Answer’s First methodology as the optimal for our system, since one of the main goals of this conversational vehicle manual is to reduce the cognitive load of producing structured queries offering the user to verbally consult the manual in the most natural way of speaking.

In other language processing methodologies, certain terms, such as stop words (the, who, that, etc.) are often removed before query execution. In these methodologies a user’s original question is changed into a question that is suitable for execution or that the engine can understand. The question, “How do I open the car door” may be reduced into “open car door” before the query is executed. In A1 however, all terms are taken as relevant information that further define the user’s needs, often lost in the process. In addition, the A1 methodology suggests that the order of the words in a question does not influence the interpretation of the original query. Instead, queries or user questions are broken into bigrams, and these bigrams are used to reconstruct a user's utterance from a storage location.

4.2 Design: What drivers want to know

To begin the design process key features of the BMW 7 series were identified. These features were chosen from the perspective of the driver, based on what the research team considered to be relevant functions of the vehicle to be used while driving (see Table 1).

Features were mapped to a page or pages in the online 2010 BMW 7 Series owner’s manual. The pages of the vehicle manual were manually scanned, in teams of two to three researchers, and the page numbers on which a feature was discussed was recorded. If a feature was discussed on more than one page, each page number on which it appeared was recorded. If more than one feature appeared on a page, the page was recorded under each feature. The teams later came to a consensus, in a group meeting, on the page numbers recorded. This process created a many-to-many relationship between features and pages.

Through an informal brainstorming session, each group identified and recorded all the questions related to each feature. This group of questions included all questions that participants in the brainstorming session thought that a user might perform about a particular feature. The questions were later input into the Questions table.
Table 1. Selected BMW 7 Series features for driver usage

<table>
<thead>
<tr>
<th>Driving Programs</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC, Active Cruise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control with Stop &amp; Go</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSC, Dynamic Stability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDC, Hill Descent Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed Limit Setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sport Program</td>
<td>xDrive</td>
<td></td>
</tr>
<tr>
<td>Infotainment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-up Display</td>
<td>Music Search</td>
<td></td>
</tr>
<tr>
<td>Navigation</td>
<td>Radio Search</td>
<td></td>
</tr>
<tr>
<td>Voice Commands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive Light Control</td>
<td>Emergency Call</td>
<td></td>
</tr>
<tr>
<td>Emergency Request</td>
<td>Fuel Consumption -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Efficient Dynamics</td>
<td></td>
</tr>
<tr>
<td>Night Vision</td>
<td>Service Road</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assistance</td>
<td></td>
</tr>
<tr>
<td>Tire Pressure Monitor</td>
<td>Warning Message -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Check Control Display</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>Gentleman Function (Seat Position)</td>
<td></td>
</tr>
<tr>
<td>Air Conditioning</td>
<td>Wiper System</td>
<td></td>
</tr>
<tr>
<td>Heating</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3 Database Design using Answer’s First

After generating a list of questions and page numbers for each feature, the Answer’s First methodology was applied. Each question for each feature was decomposed into bigrams or “terms”. For example, the question “How do I open the car door?” was decomposed into 6 bigrams: “How do”, “do I”, “I open”, “open the”, “the car”, “car door”. These bigrams were stored in the Terms table. Questions in natural form were stored in the Questions table, and the page numbers on which the answer to a question could be found were stored in the Answer’s table. Figure 1 shows the relationships between the tables in the database.

When a query is executed against the database, it is first broken into bigrams. The bigrams are compared to word pairs in the Terms table to find those terms that match the bigrams represented in the user’s question. Each question in the Questions database is then ranked based on the concentration of matched bigrams in the Terms table. The question with the highest concentration of matched bigrams is ranked first; the second question is ranked second, and so on. The result is a list of questions ranked by frequency of matched bigrams.

4.4 Voice User Interface Design and Implementation

VoiceXML was used to design the front-end voice user interface. VoiceXML is a version of XML designed for creating voice activated dialogs. A combination of VoiceXML and XML was used to write the application and speech grammar. Voxeo was used as the developer platform [41]. For the purposes of our prototype the following scenario was used to define a user's interaction with the manual system:

- User Calls the System
- System: Hello, How May I Help You?
- User: How do I search music?
- System: Okay. Is your answer on page 99?
- User: No
- System: Is your answer on page 104?
- User: No
- System: Is your answer on page 234?
- User: Yes
- System: Okay. Thank you for using the vehicle voice manual. Is there any other way I can assist you?

When the user calls the system, it prompts the users to ask a question (See Figure 2). The query is executed against the database and the top three likely results are returned. The question with the highest concentration, or frequency, of matched terms (bi-grams) is returned first. The question with the second highest concentration of matched terms is returned second, in case the first answer was not correct, and the question with the third highest concentration of matched terms is returned last. If none of the answers returned provide the user with a correct answer then the user is again prompted to rephrase the question.
To allow for natural interaction, it was important that the user had the perception that the manual could answer any question it was asked. This influenced the design of the speech grammar. The speech grammar determines what the user can and cannot say. The voice-user interface application will only recognize utterances defined in its grammar. This implies that a voice-user interface designer must anticipate a user’s utterances and include all possible utterances in the grammar.

Because of the large corpus of possible utterances, anticipating the user’s utterances would prove to be a tedious and inefficient task; therefore, instead of a static grammar the researchers created the grammar dynamically from bigrams stored in the Terms database. This design decision is based on a premise of the Answer’s First Methodology that the order of the words within a statement or question does not influence the interpretation [40]. This meaning that a user may ask the same question in several ways but always get the same result.

By generating questions dynamically using bigrams, a grammar that recognizes a large number of utterances with minimal effort, can be provided. The questions “How do I open the car door”, “Open the car door”, and “Open the car door how do I”, all can be mapped to the same set of bigrams (See Figure 3).

Using the terms table an algorithm was developed to create a grammar that includes all possible questions formed from all combinations of bigrams in the database. Because the grammar was dynamically generated using the Terms table a larger variety of anticipated questions was available. Therefore, the system is more likely to recognize a question that was not originally in the Question’s Table.

5. Conclusions and Future Work
In this paper we introduce a working prototype of a conversational voice interfaced vehicle manual using the Answer’s First methodology. The system allows users to interact naturally and returns the most relevant answer. Similarly to previous results, we anticipate that the use of the Answer's First methodology will yield high accuracy when naturally spoken longer queries are executed and the flexibility of the system will increase user satisfaction through the voice user interface in the vehicle.

Our future goal is to begin user studies to test the accuracy of the system. In addition development will be done to integrate the voice-interfaced manual into the dash of a car providing not only the page reference in the printed manual, but actual spoken answers. Therefore, improvements will also be made on the front-end. In addition, further research will be conducted towards a context-sensitive interactive manual, see Figure 4.
Two main research areas of interest will be followed, the personalization of the user’s interaction with the manual, and the context-awareness of the vehicle. Within the first field of interest the goal is to provide the system with the ability to learn from continuous interaction with the driver and from the driver’s usage of the vehicle. Data collection of driver’s performance would allow for education of the driver on efficiency, security or improved usage of the vehicle features. A complimentary system improvement should focus on the context-awareness of the automobile given by data collection of the vehicle’s surroundings and geographical location. Management of out-of-vehicle data would provide the driver with relevant information like traffic and weather conditions, points of interest, or regulations on unfamiliar environments. The ultimate scope of this interactive driver assistant is to make the user more aware of the vehicle and the vehicle more aware of the surroundings.

From an architectural standpoint, the voice user interface is a first step into a centralized digital vehicle manual production system as it presents the opportunity to have the vehicle manual integrated in the car while accessing a back-end server that enables flexibility in the deployment. The further development of this centralized system will benefit OEMs and will allow them to lower the documentation process and costs of user manual deployment from the cloud server to different media. Furthermore, this infrastructure would easily permit the personalization of vehicle manuals via VIN or customer login identification. Users would be provided with content that specifically fit the features installed in their vehicles. Any modification from the OEM, or the user side, would automatically take place in the central system and deployed over-the-air to the final customer who could consult in every moment the correct version of the manual matching his vehicle. In addition, this solution would inevitably support the Go Green efforts to protect and improve the sustainability of our environment lessening paper consumption.

6. ACKNOWLEDGMENTS
This material is based in part upon work supported by the National Science Foundation under Grant Number CNS-0955763. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES


[37] Kujala, T. 2009, “Efficiency of visual time-sharing behavior- the effects of menu structure on POI search tasks while driving”, In *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Essen, Germany, 63-70

[38] Chang, J., Lien, A., Lathrop, B., Hees, H. 2009, “Usability evaluation of a Volkswagen Group in-vehicle speech system”, In *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Essen, Germany, 137-142

