Are Mobile In-Car Communication Systems Feasible? A Usability Study

Patrick Tchankue
NMMU – Summerstrand Campus (South)
University Way, Summerstrand, 6031
Port Elizabeth, South Africa
+27 41 504 2088
Patrick.TchankueSielinou@nmmu.ac.za

Janet Wesson
NMMU – Summerstrand Campus (South)
University Way, Summerstrand, 6031
Port Elizabeth, South Africa
+27 41 504 2323
Janet.Wesson@nmmu.ac.za

Dieter Vogts
NMMU – Summerstrand Campus (South)
University Way, Summerstrand, 6031
Port Elizabeth, South Africa
+27 41 504 2089
Dieter.Vogts@nmmu.ac.za

ABSTRACT
The issue of driver distraction remains critical despite efforts that aim to reduce its effects. In-Car Communication Systems (ICCS) were introduced to address visual and manual distraction occurring when using a mobile phone whilst driving. ICCS running on mobile phone have increased the number of people using ICCS as they can be installed at no cost and the quality of speech recognition on mobile devices is improving. Little research, however, has been conducted to investigate usability problems with mobile ICCS. This paper proposes a new model to address some of the issues found with current mobile ICCS, called the multimodal interface for mobile info-communication with context (MIMIC). This paper discusses the design and usability evaluation of a prototype mobile ICCS, designed using MIMIC. Several tasks were evaluated using different metrics including time on task, task completion, task success, number of errors, flexibility, user satisfaction and workload. Results obtained show that the users gave a good overall rating to the mobile ICCS, which indicates that users will easily accept such technology. Future work will include redesigning the speech user interface in order to address the usability issues found with the current prototype.

Categories and Subject Descriptors
H.5.2 [Information Systems]: Information interfaces and presentation – user interfaces, evaluation/methodology.

General Terms

Keywords
Speech user interfaces, mobile application, interaction design, driver distraction, spoken dialog systems, in-car communication systems.

1. INTRODUCTION
Driver distraction remains a serious issue as far as car safety is concerned. Driving while using a mobile phone can reduce the amount of brain activity associated with driving by 37% [12]. In the US, 18% of injury crashes in 2010 were reported as distraction-affected crashes [17]. Using a mobile phone while driving can delay a driver's reactions as much as having a blood alcohol concentration at the legal limit of 0.08% (USA) [24]. This result is valid regardless of the type of mobile phone used; hands-free or handheld. Making calls and sending text messages are the tasks that are most often initiated by drivers. It has been shown that sending a text message can cause more distraction than making a call. Text messaging creates a crash risk 23 times worse than driving while not distracted [19]. As an illustration, sending or receiving text messages takes the driver’s eyes from the road for an average 4.6 seconds; within this time the driver can travel the equivalent of an entire rugby field assuming that the speed was 90km/h.

Research on distracted driving in Johannesburg, South Africa by the AA found that 7.2% of drivers were holding and using their mobile phones while driving [21]. A large percentage of the people said that they knew that driving while distracted was dangerous, but did it anyway.

The introduction of ICCS is viewed as a means to reduce driver distraction caused by mobile phones. Several studies have shown that using a voice-activated interface can help in reducing driving errors [20; 23]. This technology, however, is not yet used by a large number of drivers. ICCS are not considered a standard option when buying a car, as there is an additional cost involved. In addition, young drivers who are more likely to use their phones while driving [14], cannot always afford this extra cost.

The smartphone market is growing exponentially and many mobile applications can be downloaded for free [6]. Drivers who use hand-held devices are four times more likely to be involved in crashes serious enough to injure themselves [11].

This paper will address the issue of lack of usability research on mobile ICCS. User-centred mobile ICCS can contribute in reducing driver distraction, because a poorly designed mobile ICCS can increase the driver’s cognitive workload increasing the vulnerability to distraction.

This paper is organised as follows: background information about ICCS and speech user interfaces are discussed in Section 2. Section 3 discusses the design of MIMIC as well as the implementation of the speech user interface. The usability evaluation of the speech user interface (SUI) is discussed in Section 4, while Section 5 presents the results of the user study and a discussion of these results is presented in Section 6. Section 7 concludes this paper.
2. BACKGROUND
This work is related to previous research on ICCS and speech user interfaces, which are reviewed in the sections below.

2.1 In-Car Communication Systems (ICCS)
ICCS are systems that can support drivers in using their mobile phone whilst driving. The interaction with ICCS is often performed in a hands-free and eyes-free manner. ICCS perform classic communication tasks such as making calls or sending text messages to phone numbers or contacts. Shortcut commands, such as calling the last incoming number (call back) or redialling the last outgoing number (redial) are used to simplify and speed up the interaction. Today’s cars include a number of similar systems [16]: navigation, car information, safety and entertainment. Vehicles can also communicate with other vehicles, sharing information about highway alerts and emergency systems, to increase the level of safety. State-of-the-art navigation systems integrate real-time traffic data, customised points of interest and friends’ locations, enabling the access of social networking in the car [22].

Intelligent transportation systems are also used to prevent human errors that could lead to car accidents. These include systems such as traffic light detection, car-to-car communication to avoid traffic jams, intelligent speed adaptation and collision avoidance systems. Advanced driver assistance systems (ADAS) are increasingly included in new cars. These include automatic parking, traffic sign recognition, lane change assistance, driver drowsiness detection, adaptive cruise control and in-vehicle navigation system.

Today almost all car manufacturers provide an ICCS on at least one luxury model. However this often comes with extra costs. For example SYNC [8] from Ford costs up to $395. This extra cost limits the access of such technology, especially amongst young drivers who cannot afford it, despite being more likely to use a mobile phone while driving [26].

Several mobile ICCS have been launched over the past few years. Mobile ICCS are mobile applications that can be downloaded to mobile phones. These applications are often voice-activated. Examples of mobile ICCS include DriveSafe.ly, Siri [2], StartTalking [1], SpeakToIt, Voice Talk (Vlingo), S Voice (Samsung) and many more. Unfortunately most of these mobile applications are not fully hands-free; they require manual interaction for tasks such as push-to-talk. Some mobile ICCS, such as DriveSafe.ly, have to be activated before a journey so that they block incoming calls and text messages. Several smartphones use the same strategy when in “Driving Mode”, which has to be manually activated.

Although switching off calls and text messages is effective in minimising driver distraction, drivers are not likely to use such options unless they experience adverse driving conditions. A survey of the features of most mobile ICCS revealed that most mobile ICCS are not aware of the current driving environment.

2.2 Speech User Interfaces
Speech user interfaces (SUI), also called voice user interfaces, are what a person interacts with when communicating with a spoken language application [5]. SUIs have been used in a wide range of applications both in desktop and ubiquitous applications. The use of the speech channel is strongly encouraged in situations where the user’s hands are used for other purposes such as steering a car.

The design of SUIs is subject to several challenges related to the application domain. These challenges become more critical when designing for a domain such as the automotive domain. When designing a general SUI, the following list of requirements are often taken into account [25]: speech recognition accuracy, dialogue flow, reliability (ambient noise), human cognition (limited working memory), user (native speaker) and hardware (microphone). Some user studies have highlighted the importance of the user’s first language and background noise on the success of speech recognition on a mobile platform [13]. The word recognition rate was discovered to be inversely proportional with the level of noise. Native English speakers, however, managed to perform better than non-native English speakers under all noise conditions, especially with loud background noise.

Figure 1 depicts the typical components of a spoken language application. These include: an automatic speech recogniser (ASR) to perform speech to text conversion, a dialogue manager (DM) to control the interaction with the user, and a mechanism for conveying information to the user. More advanced systems incorporate modules for natural language understanding (NLU) and natural language generation (NLG) in addition to the previous components.

![Figure 1: Typical architecture for voice-activated application](image)

The successful design of a SUI depends on several aspects [5]. These include the design of prompts, the grammar used and the dialogue logic. Prompts have two main purposes. They cause the user to speak and they convey to the user what may be spoken. Synthesised speech is difficult to understand in noisy environments. That is why users will benefit from commands such as “REPEAT” that allow the user to have a sentence repeated by the system.

Several techniques are used to implement dialogue managers. These include finite-state based, frame-based and agent-based dialogue management [3]. With finite-state based dialogues, the dialogue is designed as a set of states that can change depending on the situation. This results in unnatural dialogues as the user has to provide information to the system only when the system needs it. The advantage of such techniques is the ease of implementation. Frame-based dialogues contain frames which, in turn, are made of slots. Several slots can be filled at once, leading to mixed-initiative dialogues where either the user or the system can initiate the conversation. Agent-based dialogues are more advanced; artificial intelligent agents are used to control the
dialogue’s flow. However the implementation of such techniques can be challenging and time-consuming.

3. DESIGN AND IMPLEMENTATION

The final goal of this research is to design an accessible and intelligent model for mobile ICCS. The mobile aspect of the model is the use of the mobile device itself whereas the intelligent aspect will be fulfilled by the Context-aware module. Figure 2 shows the proposed model which is comprised of several components interacting with each other. These include the Input module, the Dialogue module and the Output module.

3.1.1 Input module

The input module contains several important sub-modules. These are the automatic speech recogniser (ASR), the natural language understanding unit (NLU) and the mobile phone itself. The caller’s mobile phone can provide valuable information to the dialogue module. The mobile phone of the caller (peer), if running MIMIC, can share contextual information obtained from sensors, GPS and web services.

The speech recognition for mobile devices is not performed on the actual phone because of the limited processing power of the phone. Instead, most mobile speech providers, such as AT&T, Nuance and Google, perform an online processing of spoken inputs [4; 9]. The ASR or speech-to-text module detects speech from the mobile phone’s microphone and returns the corresponding text. This text or list of possible results is passed to the NLU that will determine the semantics. The NLU uses a database of homonyms and synonyms to reduce the possibility of errors. The mobile phone is also used as a microphone.

The implementation of the ASR used by MIMIC was done using the Google Speech API. This API sends compressed recorded sounds in the cloud to Google speech servers. The speech servers process the sound and return the possible results ordered by confidence level.

3.1.2 Dialogue module

The dialogue module contains the Context-aware module and the Dialogue engine. The design of these modules is discussed below.

The Context-aware module is not yet implemented, but the Context-aware module will use sensor information to determine the current driving context. This module will comprise an inference engine and a module dedicated to determining the adaptation effects.

The Context-aware module will use a neural network to determine the distraction level. The distraction level is directly related to the current driving context, as a difficult driving context will imply a high distraction level.

The dialogue engine has been implemented. The dialogue engine is responsible for determining the next move of the dialogue and giving turns to the user. Several design schemes are often used in designing dialogue systems. These include finite state-based, frame-based and agent-based [3].

The state-based approach represents the dialogue as a set of states. The user has to provide data on each state in order to move to the next state. This approach is easy to implement but produces unnatural dialogues.

The frame-based approach represents the dialogue as a frame made of slots that need to be filled. An example is the frame “SMS” that has two slots: telephone number and the content of the message. Once the dialogue manager detects a frame, it has to fill it with any information coming from the user. This makes the dialogue less restrictive and more natural. There is no order in which data should be provided to the system.

Lastly the agent-based approach to design the DM module uses agents that are artificial intelligent components. This produces a highly flexible dialogue that can be easily adapted for any
application domain. The drawback is that it uses a considerable amount of processing power and it is complex to implement.

A frame-based approach was selected in implementing MIMIC. The following frames were implemented: CALL (number), CALL (contact), SMS (number, content), SMS (contact, content), REDIAL, REPEAT and CANCEL.

Prior experiments with the speech recogniser resulted in poor accuracy when dictating a full text message. Pre-defined text messages were chosen based on text messages that are likely to be sent while driving [8]. A few messages were chosen in order not to force the driver to remember a long list of options. The following four options were selected:

- I will call when I get there;
- I can’t talk right now, I am driving;
- I am running a few minutes late; and
- I am on my way;

![Figure 3: Graphical Simulation of MIMIC](image)

Figure 3: Graphical Simulation of MIMIC

The SUI was implemented on a Samsung Galaxy S2 running Android Gingerbread (2.3.3). Android was chosen because it is a growing mobile platform that is supported by a wide range of mobile phone brands. The speech was converted into text using the speech recognition API shipped with Android (android.speech). This API records and compresses a spoken input, then sends it to the Google speech servers which process it and return the results. In MIMIC, whenever a spoken input is recognised by the ASR, a “beep” sound is played as feedback.

The recogniser listener library (RecognitionListener) was preferred to the traditional recogniser intent (RecognizerIntent) library because the later needs to be manually activated. This would force the driver to use his/her hands to operate the system. The recogniser’s library is only available for devices running at least Android 2.2.

3.1.3 Output module

Outputs are processed by two components: the NLG and the TTS. A lightweight NLG module was designed and implemented in order to arrange data in a way that the user will have a better understanding. Examples of such output include telephone numbers and incoming text messages written using common abbreviations.

The text to speech engine syntheses the information that needs to be sent to the user. The default Android TTS was used to implement this component.

4. USER STUDY

Following the implementation of the mobile ICCS, a user study was conducted to investigate the feasibility of the system. The feasibility was evaluated using several aspects of the system. These include the usability, workload and user satisfaction. The impact of MIMIC on driver distraction was not investigated; this will be done in a subsequent study using a driving simulator. The aims of the user study and the research methodology are discussed in the following subsections.

4.1 Aim of the study

The goal of the user study was to investigate the feasibility of using a mobile ICCS. Participants used different strategies to make calls and send text messages.

4.2 Research methodology

The methodology used in the user study will be discussed in the following sections. These include the selection of the participants, the apparatus, the procedure, the tasks performed and the metrics used.

<table>
<thead>
<tr>
<th>ID</th>
<th>Task Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T01</td>
<td>Please call the contact Maria</td>
</tr>
<tr>
<td></td>
<td>- Say “Contact Maria”;</td>
</tr>
<tr>
<td></td>
<td>- Answer “Yes”</td>
</tr>
<tr>
<td>T02</td>
<td>Send the text message “I am running a few minutes late” to John:</td>
</tr>
<tr>
<td></td>
<td>- Say “SMS John”;</td>
</tr>
<tr>
<td></td>
<td>- Say “three” to choose the third option;</td>
</tr>
<tr>
<td></td>
<td>- Say “Yes” to send the selected text message</td>
</tr>
<tr>
<td>T03</td>
<td>Redial the previous outgoing call:</td>
</tr>
<tr>
<td></td>
<td>- Say “Redial”;</td>
</tr>
<tr>
<td></td>
<td>- Answer “Yes”</td>
</tr>
<tr>
<td>T04</td>
<td>Please call Diana</td>
</tr>
<tr>
<td></td>
<td>- Say “Phone Diana”;</td>
</tr>
<tr>
<td></td>
<td>- Answer “Yes”</td>
</tr>
<tr>
<td>T05</td>
<td>Send the text message “I will call you when I get there” to Peter:</td>
</tr>
<tr>
<td></td>
<td>- Say “Text Peter”;</td>
</tr>
<tr>
<td></td>
<td>- Say “one” to choose the first option;</td>
</tr>
<tr>
<td></td>
<td>- Say “Yes” to send the selected text message</td>
</tr>
<tr>
<td>T06</td>
<td>Send the text message “I can’t talk right now, I am driving” to Janet:</td>
</tr>
<tr>
<td></td>
<td>- Say “SMS Janet”;</td>
</tr>
<tr>
<td></td>
<td>- Say “two” to choose the second option;</td>
</tr>
<tr>
<td></td>
<td>- Say “Yes” to confirm the message;</td>
</tr>
<tr>
<td>T07</td>
<td>Call a number :</td>
</tr>
<tr>
<td></td>
<td>- Say “Call 074 456 1245”;</td>
</tr>
<tr>
<td></td>
<td>- Say “Yes” to confirm the number;</td>
</tr>
</tbody>
</table>
4.2.1 Selection of participants
Research has shown that ten users are sufficient to discover approximately 95% of usability issues [18]; therefore ten (10) volunteers were recruited and completed the user study. The main selection criteria were the language spoken and the age of the participants. All the participants have a valid South African driving license with a strong computing background. Most participants had never used a mobile ICCS prior to the study. This was done to investigate the ease of learning of the system. The age of participants ranged from 18 to 29; this was chosen because the majority of drivers likely to use a mobile phone while driving are young.

4.2.2 Apparatus and procedure
The user study was conducted in a lab environment. A low-fidelity driving simulator was designed using the Lane Change Task (LCT) test software [15]. A video projector and a board were used to simulate the windshield of the car. Each participant was required to sign an informed consent form prior to the start of the evaluation session. A brief demonstration on how to use the system was given by the test moderator. During the demonstration, participants performed one of each type of task; that is, calling using a contact name, calling using a telephone number and sending a text message.

4.2.3 Tasks performed
Each participant was asked to perform several tasks. These included making a call and sending a text message (Table 1). When making a call, participants had three options: using a contact’s name, dictating the phone number or redialling the last outgoing number.

Telephone numbers used followed the format valid in South Africa. South African telephone numbers consists of 10 digits starting with a “0”.

4.2.4 Metrics
Performance and self-reported metrics were captured. These included the following:

Performance metrics:
- **Time-on-task** (seconds): the time spent in performing a task;
- **Error on task**: the number of errors made while performing a specific task;
- **Task completion**: indicates whether the task was completed or not; and
- **Success rate**: indicates whether the task was successful or not.

Self-reported metrics:
- **Overall satisfaction**: a SUS questionnaire [7] was used as an instrument to collect self-reported usability data;
- **Speech recognition and dialogue performance**: participants were asked to rate the accuracy of the speech recogniser as well as the flexibility of the dialogue; and
- **Workload**: mental, physical, temporal demand was collected in addition to performance, effort and frustration experienced by participants. The NASA TLX [10] was used.

5. RESULTS
The results are presented in three sections, namely performance, workload and satisfaction. Graphs and tables showing the results are discussed.

5.1 Performance

![Figure 4: Means of errors for each task (n=10).](image1)

Figure 4 shows the number of errors logged by the system. The highest number of errors was recorded for Task 3 (redial the last outgoing number). A further analysis of the log showed that the command “REDIAL” was often not recognised. The use of a constrained grammar could prevent the occurrence of such an issue. Task 7 consisted of making a call using the telephone number. Most participants had to try at least twice in order to successfully perform this task.

![Figure 5: Time (seconds) spent on each task (n=10).](image2)

Figure 5 depicts a box and whisker graph of the time-on-task. The minimum, the first quartile, the median, the third quartile and maximum are represented for each task. For each box, the first half represents values greater than the first quartile, while the second half represents values greater than the median. The lower whisker represents values ranging from the minimum to the first quartile, while the upper whisker represents values ranging from the third quartile to the maximum. Calling a contact (T01 and
T03) took less time than the other tasks; selecting and sending a text message to a contact (T02, T05 and T06) were performed, on average, within 45 seconds. This is reasonable because after recognising the command, the system has to inform the user about the available options so that the user chooses the preferred option. However on these tasks, there were a large number of people performing in the third quartile. Calling a contact using the telephone number (T07) was the most time-consuming task; most of the time was spent in getting the telephone number recognised correctly.

Table 2. Task completion and success (n=10)

<table>
<thead>
<tr>
<th>ID</th>
<th>Completion</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>T01</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>T02</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>T03</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>T04</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>T05</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>T06</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>T07</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

The completion and success rates of the tasks (Table 2) were encouraging for making a call using a contact name (T01 and T04). However the use of the shortcut command “REDIAL” resulted in only 60% completion rate (T03). This was due to the poor recognition of the command. Sending text messages using a contact name (T02, T05 and T06) were all completed with a success rate of at least 90%. Thirty percent of participants failed to complete T07, which consisted of making a call by dictating a telephone number.

5.2 Workload

A post-test questionnaire was completed after the end of tasks. The NASA TLX gave the following results (Figure 6).

The ease of use, integration and ease of learning were highly rated by the participants (median=4.00). This was followed by the willingness of the participants to use the system frequently (median=3.50) and the confidence that participants had towards the system (median=3.00). In addition, few participants found the system complex, cumbersome or thought they would need assistance to use it. Very few participants (median = 1.00) felt that they would need a lot of learning (training) to be able to use the system efficiently.

5.3 Satisfaction

The results of the user satisfaction (Figure 7) were generally high. The ease of use, integration and ease of learning were highly rated by the participants (median=4.00). This was followed by the willingness of the participants to use the system frequently (median=3.50) and the confidence that participants had towards the system (median=3.00). In addition, few participants found the system complex, cumbersome or thought they would need assistance to use it. Very few participants (median = 1.00) felt that they would need a lot of learning (training) to be able to use the system efficiently.

6. DISCUSSION

Overall the participants gave positive feedback after using the SUI of MIMIC. However some aspects of the prototype need to be improved. The high number of errors in performing Task 3 and
An analysis of the general comments given by participants highlighted the need for a barge-in strategy. A barge-in strategy enables the user to take control by interrupting the system. This will speed up the process of selecting a text message. When implementing the proposed model some restrictions of the ASR did not allow the integration of a barge-in facility. Other ASR engines will be investigated in order to overcome this issue.

7. CONCLUSIONS AND FUTURE WORK

Driver distraction is still a matter of concern throughout the world. ICCS are supposed to attenuate the effects of driver distraction but these systems are still not widely accessible. Some mobile ICCS are available on mobile phone application markets, but the usability of these applications has not been extensively investigated.

The use of ICCS is limited due to the high cost of this technology. Mobile ICCS present a solution to this issue as the number of mobile phones capable of hosting such applications has increased. In addition, few mobile ICCS are available for sale. The usability of mobile ICCS is a key factor to achieve the reduction of driver distraction, but little research has been conducted in this area.

This paper discussed the design of a model for mobile ICCS called MIMIC and the implementation of a prototype system. The system was found to be highly effective in sending pre-recorded text messages and making calls. The majority of the participants were willing to use such an ICCS in future. However the dictation of telephone numbers was a source of several errors that frustrated the participants. In addition, the system frequently failed to recognise the command “REIAL”.

Future work will include the implementation of the Context-aware module and experiments to select the most effective techniques to acquire, infer and use contextual information. A user study will be conducted in order to measure the effectiveness of the proposed model. This study will be designed using a driving simulator capable of providing contextual data such as weather, GPS, light and ambient noise.

8. REFERENCES


