Designing a Mobile, Context-Aware In-Car Communication System

Patrick Tchankue, Janet Wesson and Dieter Vogts
Department of Computing Sciences
Nelson Mandela Metropolitan University, P. O. Box 77000, Port Elizabeth 6031
Tel: +27 41 5042323, Fax: +27 41 5042831
Email: {Patrick.TchankueSielinou, Janet.Wesson, Dieter.Vogts}@nmmu.ac.za

Abstract- In-car communication systems (ICCS) are intended to minimise driver distraction when drivers engage in phone-related tasks whilst driving. Visual and manual distractions can be significantly reduced with the use of speech input and output. However, ICCS are still not widely adopted, partially, due to their high cost. Furthermore, existing ICCS do not use the vehicle's contextual information. This paper discusses the design of MIMIC (Multimodal Interface for Mobile Info-communication with Context), a collaborative mobile application with a speech-based user interface that uses different phone sensors and web services to determine the context of the current driving situation. Contextual and collaborative information is used to determine the distraction level of the driver. Based on the distraction level, different adaptation effects are applied in order to reduce the driver distraction. Example adaptation effects include delaying or cancelling incoming calls and text messages and sharing the driver status with the caller. The design of MIMIC is discussed, together with its architecture.

Index Terms — Context-aware Interfaces, In-Car Communication Systems, Neural Networks, Speech User Interfaces, Distraction level.

I. INTRODUCTION

Spoken dialogue systems are becoming popular especially in the automotive industry. Several studies have shown that hands-free and eyes-free modalities, such as speech, can reduce driver distraction when drivers use their mobile phones while driving [28].

Driver distraction remains a serious issue as far as car safety is concerned. Statistics from the Automotive Association (AA) in South Africa confirm this fact for developing countries. Research on distracted driving in South Africa by the AA found that 7.2% of drivers were holding and using their mobile phones while driving [21]. This study was conducted in Johannesburg during peak time. A large percentage of the people said they knew that distracted driving was dangerous, but did it anyway.

Driver distraction is more serious amongst certain age groups. According to the National Highway Traffic Safety Administration, younger drivers aged 16 to 24 are more likely to use a hand-held cell phone [19]. This implies that, research on reducing driver distraction should focus more on this category of drivers. Despite several solutions implemented, the problem remains crucial worldwide. The percentage of drivers who use a mobile phone (texting or manipulating it in some way) while driving increased to 0.9% in 2010. Effects of driver distraction vary depending on gender. The level of hand-held mobile phone usage is higher amongst female drivers than male drivers [31].

Sending text messages whilst driving is a serious cause of driver distraction. When sending a text message while driving, drivers experience manual, cognitive and visual distraction. A study revealed that over $^{1}/_{3}$ of drivers (37%) in the USA have sent or received text messages while driving, and 18% said they did it regularly [11].

ICCS are not considered a standard option when purchasing a vehicle, and there is an additional cost involved. Young drivers, who are more likely to use their phones while driving, cannot always afford this additional cost.

The fact that a driver may not be visually distracted does not guarantee that they will be free from other forms of distraction (cognitive, manual, auditory, etc.). The difficulty of changing driving situations and the environment itself require more attention from the driver.

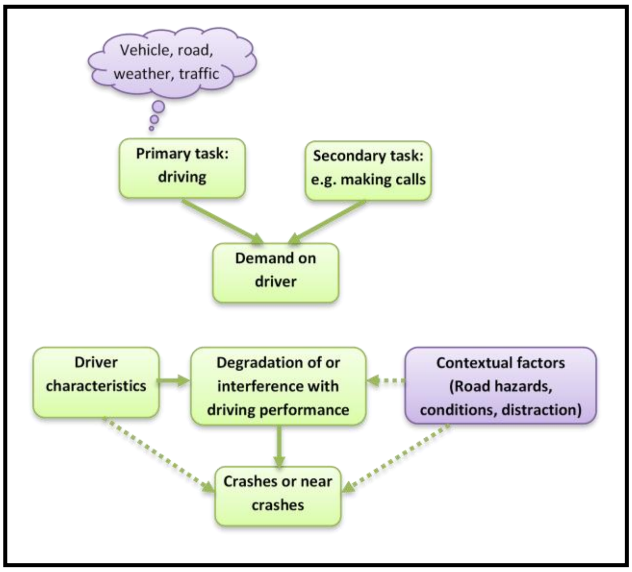

Figure 1: Sources of demand on driver and their safety relevance (Adapted from [3])

Driving can be a complex task. Studies classify the driving task into three levels; namely strategic, tactical and operational [17]. Additional or secondary tasks may interfere with the primary task. Figure 1 depicts several demands that the driver experiences. Contextual factors (vehicle, weather and traffic) can seriously affect driver performance and lead to either a near-crash or a crash.

According to Wickens [30], tasks that share the same pool of resources interfere with each other. Attention theory also supports the fact that two concurrent tasks with different modalities will have less interference than two tasks with the same modality [10]. This means that combining manual-visual tasks and a speech task would result in less cognitive load than two concurrent visual tasks.

This paper is organised as follows: background information about ICCS, context-aware applications and adaptive interfaces are discussed in Section 2. Section 3 discusses the design of MIMIC. The Context-aware module and Inference engine are discussed in Section 4, while Section 5 details the implementation of MIMIC. Section 6 concludes this paper.

II. RELATED WORK

The design of a model for mobile, context-aware ICCS requires a better understanding of several concepts. These include ICCS, driver distraction, context-aware computing and adaptive interfaces.

A. In-Car Communication Systems (ICCS)

ICCS are systems that can help drivers to interact with a Bluetooth-enabled phone paired with the system. ICCS perform typical communication tasks such as calling or sending text messages to phone numbers or contacts. Shortcut commands such as calling the last incoming number or redialling the last outgoing number are often included. ICCS are a subset of a wide range of in-vehicle systems called in-vehicle infotainment systems (IVIS). These often include the following sub-systems [18]: navigation, car information, safety, entertainment and communication systems. Today, vehicles can also communicate with other vehicles, sharing information on highway alerts and emergency systems, to provide a safer journey. State-of-the-art navigation systems integrate real-time traffic data, customised points of interest and friends' locations, even enabling access to social networking in the car (e.g. Facebook and Twitter) [22].

Currently, almost all car manufacturers provide an IVIS including an ICCS on at least one model of car. However, this often comes with extra costs. For example SYNC [9] from Ford costs $395. This extra cost limits the access of safety-enhanced technology to young drivers who cannot afford it, but who are also more likely to use a mobile phone while driving [33].

Mobile ICCS are mobile applications that can be downloaded to mobile phones. These applications are often voice-activated. Examples include DriveSafe.ly, Siri [4], StartTalking [1], Speaktoit and Voice Talk (Vlingo). Most of these applications are not fully hands-free and 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*".

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 mobile ICCS reveals that most mobile ICCS are not aware of the driving environment.

B. Driver Distraction

Driver distraction can be defined as a phenomenon occurring when the driver's attention is, voluntarily or involuntarily, diverted away from the driving task by an event or an object to the extent that the driver is no longer able to perform the driving task adequately or safely [12, 33].

Driver distraction still has dramatic consequences worldwide. According to the AA, on average 40 people die and 25 are permanently disabled on South Africa roads daily [21].

Driver distraction and driver inattention are two separate phenomena, although the first generally leads to the second. In case of daydreaming for example, driver inattention is separated from driver distraction.

Multiple Resource Theory (MRT) [29] is a model used to predict situations of loss of attention. The MRT divides resources available into three orthogonal processing dimensions called pools. These are: perceptual modalities (visual, auditory), processing code (spatial, verbal); and processing stages (perception and central processing, response). The theory suggests that two concurrent tasks using different modalities will have less interference than two tasks using the same modality [29]. The use of a mobile phone while driving involves at least one conflicting modality which is visual; this is why it can cause a loss of attention leading to driver distraction.

Several types of driver distraction can occur. These are visual, manual, cognitive and auditory distractions. Visual distraction, also referred to as eyes-off-the-road, occurs when the driver looks away from the road. Manual distraction occurs when, for some reasons, the driver removes one or both hands from the steering wheel. Auditory distraction occurs when the driver is disturbed by internal or external noises. Finally cognitive distraction, also referred to as mind-off-the-road, occurs when the driver thinks about something that has nothing to do with driving. Although speech-enabled ICCS help in reducing most driver distractions, cognitive distraction is hard to overcome.

Driver distraction, as a multi-dimensional phenomenon, is difficult to reduce. Taking into account contextual factors that may distract the driver could be a solution to driver distraction.

C. Context-Aware Applications

According to Dey [7], context-aware applications are computer software using contextual information in order to adapt their behaviour. The adaptation may be in terms of user interfaces (information presentation), algorithms and others parameters that have an effect on the task being performed. Earlier definitions described context-aware software as applications that adapt according to the location of use, the collection of nearby people, hosts, and accessible devices, as well as to changes to such things over time [23].

1) Models to Represent Contextual Information

Context-aware applications can be very complex to design. Several approaches are used to model context-aware systems. Choosing the right approach will significantly reduce the complexity of the system. These approaches include the following: the *key-value model* which represents information as a key-value pair; the *Markup scheme model* which uses languages similar to XML (eXtended Markup Language); the *Graphical model* which uses Unified Modelling Language (UML) diagrams; the *Logic-based model* which expresses data as a set of facts that follows rules; and the *Ontology-based model* which defines concepts together with their relationships.

The structure of the context information influences the choice of algorithms used to determine the current context. Two categories of algorithms can be found in the literature: rule-based and machine learning algorithms.

2) Rule-based Algorithms

Rule-based algorithms are used when data are represented in a very structured way (Logic-based model). In such systems, the inferring process uses a set of *facts*, and knowledge captured as *rules* applied to these facts to draw conclusions, given a set of observations. The accuracy of an inference is based on the quality of both the fact and the underlying rules.

Rule-based systems are programmed declaratively. The programmer specifies a set of conditions and actions, leaving it to the system to work out how to accomplish them. The order in which the logic is specified is not important. Declarative programming provides a higher level of abstraction than procedural programming.

3) Machine Learning Algorithms

Several techniques can be used to discover a context given a set of contextual information. These include the following: neural networks, Bayesian networks and support vector machines. *Neural networks* are a set of supervised learning methods. Artificial neural networks connections represent axons and dendrites in biological neurons [2]. *Bayesian networks* are based on a probability model derived from Bayes' Theorem [5]. The use of Naive Bayes classifiers has previously been proposed for deriving context classifications from sensor data [20]. *Support Vector Machines* (SVM) are algorithms that were developed for pattern classification but have recently been adapted for other uses, such as finding regression and distribution estimation. SVM have been successfully used to detect driver distraction [15].

D. Adaptation Effects

The main purpose of context-aware systems is the adaptation that occurs based on the current context. In the literature on context-awareness, after having inferred the context, the next step is to apply the desired changes that match the current context. This step is applied by an *Actuator*. Context-aware applications consider several channels for adaptation, namely presentation, information and processing. The change of the presentation has to do with the User Interface (UI). Colours or size of fonts can be adjusted to suit the current user or environment. Secondly the content of the information to be conveyed may be adjusted. Detailed information can be removed when addressing novice users. Finally the algorithms used to process the data can be changed; the speed of the device can be calculated using geographical information in case the GPS does not provide it.

MIMI [27] is an earlier ICCS that was designed by the authors to reduce driver distraction. MIMI adapts itself according to the distraction level, which depends solely on the vehicle dynamics (speed and the steering wheel angle). It has previously been discussed how contextual factors such as the weather and traffic may significantly degrade the driver attention [3]. Beside delaying or cancelling incoming calls and text messages, MIMIC can also share its status with a caller. This provides the caller with an explanation of why the call cannot be answered or the text message cannot be received.

III. Architecture of MIMIC

ICCS can be difficult to design as a large number of states are often involved. MIMIC consists of three main modules: input, dialogue and output. However the interaction mode (speech user interface) will also be discussed.

A. The Speech User Interface

The design of Speech User Interface (SUI) is subject to several challenges related to the application domain. When designing the SUI, the following list of requirements were taken into account [25]:

- *Speech recognition*: despite the progress made in speech technology, the recognition rate is still an issue. This issue is more critical on mobile devices because the processing is done remotely;
- *Dialogue*: the robustness and flexibility of the dialogue itself can play a critical role in the success of a SUI;
- *Environment*: a noisy environment can negatively affect the performance of the speech recognition;
- *Human cognition*: potential issues include low persistence, competition with verbal processing and limited working memory capacity;
- *User*: the user profile and demographics can affect the success of a SUI. Experts will not have the same results as novices; the same applies to native English speakers and people using English as a second or third language; and
- *Hardware*: the quality of the microphone used has an impact on the success of the SUI.

All these recommendations were taken into account when designing the SUI of MIMIC (Figure 2).

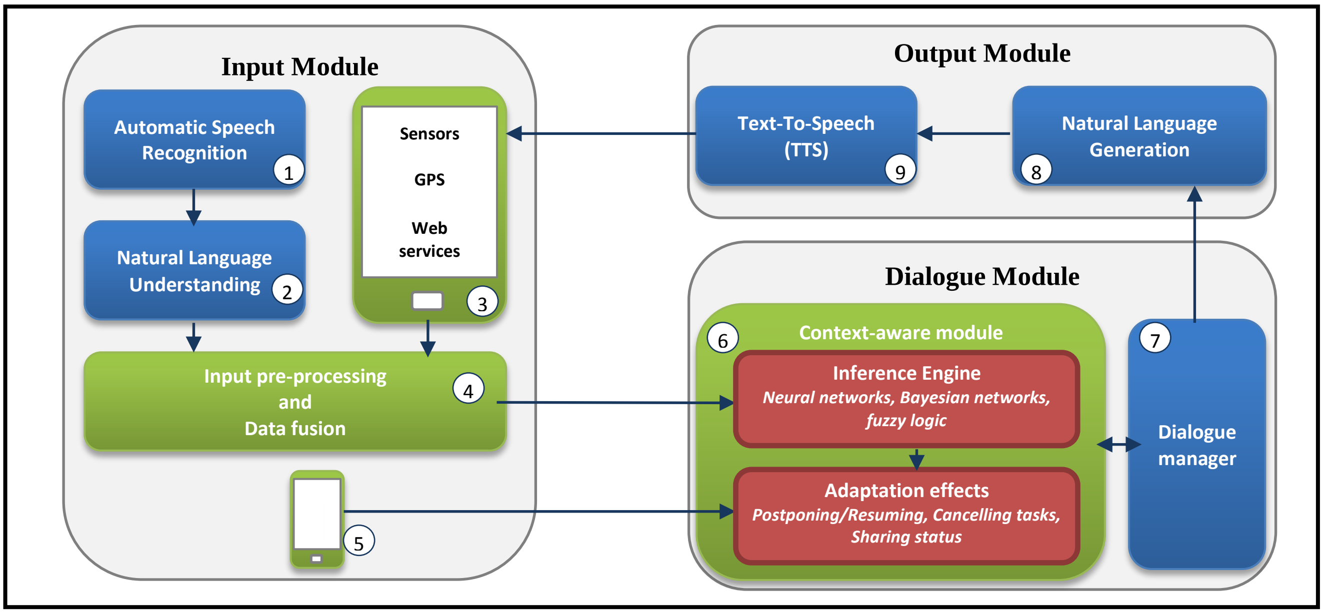

Figure 2: General Architecture of MIMIC

B. The Input Module

The Input Module is made up of several components. These include various sensors, the speech input and the peer's mobile phone.

The speech input uses the mobile phone's microphone, the car's sound system or a hands-free Bluetooth microphone. The speech is collected and compressed by the device speech API before being sent to a remote speech server for processing. The speech server will return an n-best list which is ordered by descending confidence score. The list is sent to the natural language understanding sub-module in order to determine the result that best matches the context of the dialogue. A list of homophones and synonyms is used in order to avoid potential errors. Correct utterances are the ones that match the rules set by the designer.

Sensor input is also collected and pre-processed by the input module. Sensor information includes proximity, light, accelerometer, gyroscope and noise. The Global Positioning System (GPS) is also used to determine the location, the altitude, the speed and direction of the mobile phone; which implies the location, speed and direction of the car. Web services are also used to provide weather information (temperature, humidity and speed of wind.

Collaboration can provide an opportunity to reduce distraction [16]. In the case of MIMIC being used on both the caller and callee's mobile phones, information about the level of distraction can be shared between the two applications. This information can be used to decide whether the driver can safely handle an incoming event.

The Key-Value model is used to represent all information after pre-processing. This approach is easy to implement and can be used for rapid prototyping. However, context discovery can be difficult because there are few logical relationships between different contextual information.

C. The Dialogue Module

The Dialogue Module is responsible for interacting with the user. The dialogue manager (DM) is designed to decide on the next move of the system. The decision depends on the current context and the input, the task progress and the results from queries to a database.

The DM takes a semantic representation of the user's utterance, determines how the utterance fits in the overall context and creates a semantic representation of the system response.

MIMIC is a mixed-initiative dialogue system. This means that an action may be initiated by either the driver or by MIMIC. MIMIC may detect an incoming text message with the broadcaster receiver and prepare a frame that will be sent to the DM. On the user-initiation side, the driver may give a spoken command and the input module creates the appropriate frame, which is sent to the DM. From there, the frame is processed.

D. The Output Module

The Output Module consists of a light-weight natural language generation (NLG) module and a text-to-speech synthesis (TTS) engine. NLG is a process leading from a high-level communication goal to a sequence of communicative acts which achieve this communication goal. NLG prepares the text to be sent to the TTS engine so as to avoid any misunderstanding. For example, phone numbers need to be rewritten so as to prevent the TTS from interpreting these as a simple number.

IV. CONTEXT-AWARE MODULE

Context-aware systems follow a classic architecture. As depicted in Figure 3, sensor information is captured and combined. This is also called a Context Provider (CP) in other studies. The Inference Engine or Context Interpreter (CI) is responsible for processing the information received in order to extract the meaning (context). Lastly, the context is used by the domain application to adapt its behaviour. A similar approach was used to design a previous model for automotive systems [8].

A. Context Provider

MIMIC relies on sensor data as well as web services. Some context information can be duplicate or incomplete; hence a fusion mechanism is needed to aggregate data. For example, the orientation can be obtained from the compass or the GPS. In case the GPS does not provide accurate information, the compass will be used.

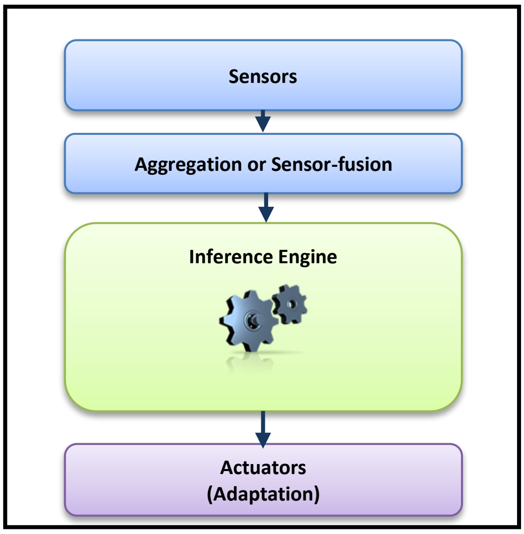

Figure 3: Structure of the Context-aware module

1) Sensors and Web Services

The proximity sensor is used to determine whether the device is in a container. This will help decide whether the light sensor can be used or not. Speed and direction are determined either by the GPS or by using the accelerometer and compass. Web services provide weather information such as temperature, speed of wind and current weather conditions).

2) Sensor-fusion

Several techniques are often used to perform sensor-fusion: these include the Dempster-Shaffer theory, Bayesian-networks and rule-based [32].

MIMIC's model uses a set of rules based on the accuracy of sensor data used to perform sensor-fusion. This technique is easy to implement, but may be subject to errors. Other techniques, such as Bayesian networks and dynamic Bayesian networks, can be trained to combine values obtained from several sensors into a single value. Similar approaches have been used in related research [24].

B. *The Inference Engine*

MIMIC's Inference Engine uses a neural network. Neural networks are suitable when dealing with unstructured and noisy data.

1) Architecture of the Neural Network:

Figure 4 depicts the architecture of the neural network used to infer the distraction level. This neural network contains three layers. Firstly, the input layer contains neurons that carry contextual variables such as the accelerometer (x, y and z), GPS (speed and bearing) and web services (current temperature, altitude etc.). All input data are pre-processed and normalised so as to minimise possibilities of errors. Secondly the hidden layer has a number of hidden neurons; future experimentation will determine the number of neurons which provide an optimum result. Finally, the output layer is comprised of a neuron which represents the distraction level.

2) Training Techniques:

The accuracy of a neural network depends on the training process. The training is an operation that consists of finding appropriate weights between neurons. Realistic training and test data containing some noise are used to train the neural network. Several techniques can be used to train a neural network. These include: back propagation, genetic algorithm and particle swarm optimisation (PSO) [2].

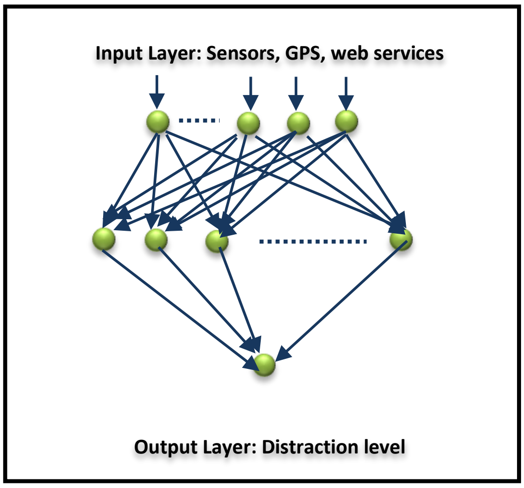

Figure 4: Neural network to infer the distraction level

V. IMPLEMENTATION

A simplified version of MIMIC (Figure 5) was implemented which did not include the Context-aware module. This was done in order to discover potential usability problems with the SUI.

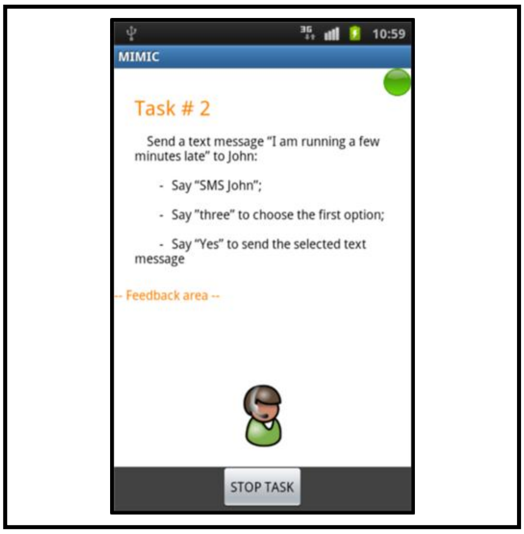

Figure 5: Graphical Simulation of MIMIC

The SUI was implemented on a Samsung Galaxy S2 running Android Gingerbread (2.3.3). The development was done using an Android Development Toolkit plugged into the IDE Eclipse. The speech is 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.

The recogniser listener library (*RecognitionListener*) was preferred to the traditional recogniser intent (RecognizerIntent) library because the latter needs to be manually activated. This would force the driver to use hands to operate the system.

Typical text messaging, call and pairing tasks were implemented. Users are able to send text messages by dictating a valid phone number or selecting a contact in the address book. The user then has to choose the predefined message that he or she wants to send. Text messages can also be read and abbreviations are converted into simple

English. Calls can be made using dictated phone numbers or contacts. It is also possible to redial the last outgoing number and call back the last incoming number.

VI. Conclusion & Future Work

Driver distraction remains a major concern; various electronic devices including mobile phones contribute to aggravating this issue. Voice-activated systems, such as ICCS, can help in reducing some aspects of driver distraction. Visual and manual distractions diminish significantly when using an ICCS. Despite the benefits of ICCS this technology is not yet largely available, partly because of additional costs involved. Contextual factors, such as weather and traffic conditions, play an important role in distracting the driver; however few ICCS use such variables in order to prevent the driver from being distracted.

This paper presented a mobile, context-aware model which aims to reduce driver distraction. MIMIC was designed to be a mobile application that uses sensor information and web services to determine the distraction level of the driver. Contextual information is used in order to derive the current context so that the ICCS can adapt its behaviour.

User studies will be conducted in order to investigate some aspects of MIMIC. These will include the usability of the SUI, as a poorly designed SUI can negatively affect the user experience. Secondly the impact of the model on driver distraction will be investigated. Experiments will be carried out to determine which techniques can best infer the distraction level and provide suitable adaptation effects.

VII. References

[1] AdelaVoice, "StartTalking." 2010. http://www.starttalking.com/. Accessed on the 2 March

[2] A. P. Angelbrecht, Computational Intelligence: An Introduction, 2 ed: John Wiley & Sons, 2007.

[3] L. S. Angell, J. Auflick, P. A. Austria, D. Kochhar, and L. Tijerina, "Driver workload metrics task 2 final report & appendices," Driver workload metrics task 2 final report & appendices 2006.

[4] Apple, "Siri Assistant." 2011. http://www.apple.com/iphone/features/siri.html. Accessed on the 2/2/2012

[5] T. Bayes, "An essay towards solving a problem in the doctrine of chances," Biometrika, vol. 45, pp. 293-315, 1958.

[6] T. Becker, N. Blaylock, C. Gerstenberger, and I. Kruijff-Korbayová, "The SAMMIE System: Multimodal In-Car Dialogue," presented at COLING/ACL 2006 Interactive Presentation Sessions, Sydney, Australia, 2006.

[7] A. Dey, "Understanding and Using Context," Personal and Ubiquitous Computing, vol. 5, pp. 4-7, 2001.

[8] M. Feld and C. Muller, "The Automotive Ontology: Managing Knowledge Inside the Vehicle and Sharing it Between Cars," presented at Automotive UI, Salzburg, Austria, 2011.

[9] Ford, "Ford Sync." 2008. http://www.fordvehicles.com/technology/sync/. Accessed on the 11/09/2009

[10] A. Gellatly, "The Use of Speech Recognition Technology in Automotive Applications," in Faculty of the Virginia Polytechnic Institute and State University, vol. PhD. Blacksburg, Virginia, 1997, pp. 203.

[11] Harris Interactive/HealthDay online poll, "Cell Phone & Texting Accident Statistics." 2011. http://consumer.healthday.com/Article.asp?AID=659288. Accessed on the 5/4/2012

[12] S. Hosking, K. Young, and M. Regan, "The effects of text messaging on young novice driver performance," presented at Distracted driving, Sidney, 2007.

[13] J. W. Jenness, R. J. Lattanzio, M. O'Toole, and N. Taylor, "Voice-Activated Dialling or Eating a Cheeseburger: Which is more Distracting during Simulated Driving?" presented at Human Factors and Ergonomics Society Annual Meeting, 2002.

[14] P. Langley, "User Modeling in Adaptive Interfaces," 1999.

[15] Y. Liang, "Detecting Driver Distraction," in Graduate College, vol. Philosophy Degree. Iowa: University of Iowa, 2009, pp. 151.

[16] J. Lindqvist and J. Hong, "Undistracted Driving: A Mobile Phone that Doesn't Distract," presented at HotMobile'11, Phoenix, Arizona USA, 2011.

[17] J. A. Michon, "A Critical View on Driver Behavior Models: What Do We Know, What Should We Do?" Human Behavior and Traffic Safety, pp. 485-520, 1985.

[18] Microsoft, "Microsoft Auto 3.1 Platform Overview." 2008. Accessed on the

[19] National Highway Traffic Safety Administration, "Distracted Driving 2009," Department of Transportation 2010.

[20] A. Ranganathan and R. H. Campbell, "A Middleware for Context-Aware Agents in Ubiquitous Computing Environment," presented at ACM/I-FIP/USENIX International Middleware Conference, Rio de Janeiro, Brazil, 2003.

[21] G. Ronald, "Demonstration: Distracted Driving." Johannesburg, South Africa, 2011.

[22] SAAB, "SAAB-IQon: Open Innovation in Car Infotainment." 2011. Accessed on the 5-4-2011

[23] B. N. Shilit, N. Adams, and R. Want, "Context-Aware Computing Applications," 1994.

[24] A. Singhalt, J. Luo, and C. Brown, "A Multilevel Bayesian Network Approach to Image Sensor Fusion," presented at Information Fusion, 2000. FUSION 2000, 2000.

[25] B. Suhm, "Towards Best Practices for Speech User Interface Design," presented at Conference on Spech Communication and Technology (Eurospeech), Geneva, Switzerland, 2003.

[26] I. Tashev, M. Seltzer, Y.-C. Ju, Y.-Y. Wang, and A. Acero, "Commute UX: Voice Enabled In-car Infotainment System," presented at Mobile HCI '09, Workshop on Speech in Mobile and Pervasive Environments SiMPE, Bonn, Germany, 2009.

[27] P. Tchankue, J. Wesson, and D. Vogts, "MIMI: A Multimodal and Adaptive Interface for an In-Car Communication System," presented at SATNAC 2010, Stellenbosch, South Africa, 2010.

[28] M. Vollrath, J. Maciej, and U. Niederee, "In-Car Distraction Study," Technical University of Braunschweig, Braunschweig 2008.

[29] C. D. Wickens, "Engineering Psychology and Human Performance," pp. 513, 1984.

[30] C. D. Wickens, Engineering Psychology and Human Performance, vol. 2. New York, NY: Harper Collins, 1992.

[31] M. S. Wogalter and C. B. Mayhorn, "Perceptions of driver distraction by cellular phone users and nonusers," Human Factors, vol. 47, pp. 455-467, 2005.

[32] H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang, "Sensor Fusion Using Dempster-Shaffer Theory," presented at IEEE Instrumentation and Measurement Technology, Anchorage, AK, USA, 2002.

[33] K. Young, M. Regan, and M. Hammer, "Driver Distraction: A Review of the Literature," Monash University, Accident Research Centre 206, 2003.

Patrick Tchankue received his MSc degree in Computing Sciences at the Nelson Mandela Metropolitan University and is presently studying towards his PhD in the same department. His research interests include context-aware interfaces, in-car communication systems, interaction design, neural networks, particle swarm optimisation and fuzzy logic.