ORIGINAL ARTICLE

Multi-Modal User Interfaces: Effects of Redundant and Complementary Approaches in High Stimuli Environments

Steven J. Kerr, Carmen Siau

Motorola Solutions, Innovation Design – Human Factors, Penang, Malaysia

ABSTRACT

Professionals such as public safety workers have to use communication devices as part of their job which can be high stress and involve demands on their cognitive resources that would be better applied to their task at hand. Multi Modal User Interfaces (MMUI), have been proposed as a way of supporting more flexible, efficient interfaces, appropriately conveying information to users whilst they are busily involved in their tasks. In this paper we describe a usability study dealing with two types of MMUI environments, redundant and complementary, under a controlled driving simulation environment and explore how these two approaches plus increasing modality impacts user interactions with their primary task of driving plus their secondary task of communicating.
Results show that increasing modalities potentially help users communicate more effectively in high stimuli environments though modality conflict has to be considered when the user cannot ignore it e.g. audio instructions on driving whilst talking to someone. Whilst complementary modalities show a slightly higher cognitive load rating, in general there appears little difference between redundant and complementary approaches.

Keywords: *multimodal interfaces, communication devices, cognitive load, usability*

INTRODUCTION

Professionals are utilizing ever more complex communication devices in high stress environments where High Velocity Human Factors (HVHF) is an important consideration in the design of said devices e.g. public safety workers. Communication devices need to be able to allow the user's cognitive resources to be focused on the job at hand and not overly on using the device. Communication devices will need to alert the user to incoming information and then convey that information in a way that does not overly distract the user. Multi-modal systems (MMS) and Multi-Modal User Interfaces (MMUI) are proposed as being an optimal method of achieving a more seamless interaction experience that can achieve the above requirements.

What are Multimodal systems?
MMS process two or more combined user input modes – such as speech, pen, touch, manual gestures, and body movements – in a coordinated manner with multimedia system output, e.g. sound, visual graphics, haptic vibrations etc. Multimodal user interface (MMUI) design is inspired by the goal of supporting more transparent, flexible, efficient and powerfully expressive means of HCI (Human-computer interaction). Some examples of MMUI include Match[1], an interactive city guide that uses speech, pen gesture and hand writing inputs and has synthetic speech output as well as dynamic graphics. Similarly Pocket Navigator[2] helps give directions, this time using haptics to relate to the direction the user should move.

Benefits of MMUI
Advantages of using MMUI are that is more intuitive, easier to learn and use and therefore require less training for various applications that can accommodate more adverse usage conditions. Due to different modalities fitting different tasks better, then it can be flexible and the system can also be robust, using mutual disambiguation (more than one recognition system) to determine user intent and not suffer inadvertent activation. MMUI systems can be faster, more efficient, support new functionality and therefore be more engaging to users[3,4]. Further advantages include redundancy, higher information bandwidth[17] and can aid users where cognitive resources need to be split across information sources[6].

These benefits are important to users of communication devices where audio/visual information is being given to a user in high stress situations e.g. public safety workers and in environments which might be noisy or visually distracting. Visual modes could include text, graphics, animations etc while auditory modes could include speech and non speech sound. Gesture interaction and haptic feedback on devices is becoming more common and this additional modality helps devices become more multi modal. The authors are interested in finding out more about how MMUI could help if applied to communication devices by users in scenarios described above, especially public safety users. Therefore we decided to do further research to understand different types of MMUI, the pros and cons around such systems to inform any potential usability study around communication systems.

BACKGROUND

Most multi modal systems currently incorporate audio, visual and haptic output, with gesture, speech and direct control as input. Kortum[7] looks at the pros and cons of these individual elements e.g. haptic alerts can be useful when a user cannot see or hear well, though may be limited in the amount of information they can convey[8], auditory interfaces can be very informative though are temporal[9]. But when you combine modalities you can limit the downsides and increase the benefits. MMUI, allow humans to communicate more naturally, permitting multiple styles of interaction[10].

Cognitive Load

"Cognitive load is the amount of mental resources needed to perform a given task. All user interfaces make cognitive demands on users. Users must master special rules of system use, learn new concepts, and retain information in short-term memory"[11]. If a UI is too distracting then it has been shown that this can lean to users making more errors and this in turn can make the user more distracted[5]. This is clearly an undesirable state for public safety users. MMUI systems have been considered as a way of helping users when they are in demanding situations as they can help with demands on a user's cognitive load. Short-term or working memory consists of multiple independent processors associated with different modes. As task complexity increases, so does the rate at which users choose to employ multimodal rather than uni-modal commands. According to Oviatt[4], analysis of users' task-critical errors and response latencies across task difficulty levels increased systematically and significantly as tasks got more difficult. In an experiment with a crisis management domain involving tasks of four distinct difficulty levels, the ratio of users' multimodal interaction increased from 59.2 percent during low-difficulty tasks to 65.5 percent at moderate difficulty, 68.2 percent at high, and 75.0 percent at very high difficulty—an overall relative increase of 27 percent.

Mental Models

Multiple Resource Theory (MRT) is a theory of multiple task performance in high work-load environments e.g. a driver focusing on road, being on phone whilst also getting information from a sat-nav device that has practical implications in whether a human operator is able to perform in such environments[6].

Closely tied is the Theory of Working Memory which maintains that short term working memory consist of multiple independent processors e.g. one for visual and one for auditory information[5, 17]. Therefore an important element of MRT relates to whether there is conflict of resources in performing cognitively related tasks e.g. visually watching road whilst driving whilst also visually checking sat-nav for route information[6]. An important consideration around MRT therefore for public safety users are will an MMUI help users concentrate on their primary task when in high stress situation or will it just add to the clutter and make things worse?

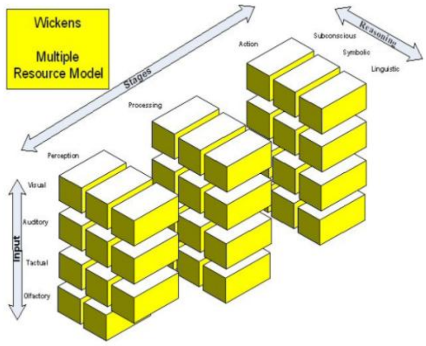

Fig. 1. Multiple Resource Model[6]

Mutual Disambiguation

Some input methods on their own such as handwriting and speech are error prone and there has been research on how to build techniques to reduce errors in these systems[13]. MMUI allows a system to have Mutual Disambiguation which means that recognizing one mode can help recognize another. The fact that recognition in one mode may have in enhancing the recognition in another mode is a positive attribute that can be illustrated in the following example (see figure 2):

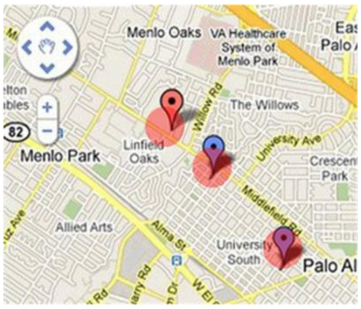

Fig. 2.

A user may say "Place three emergency hospitals here <point> here <point> and here <point>" provides a multimodal system with multiple cues. As the spoken word "three" can be matched to the number of objects to be placed so there is a level of redundancy that can allow a system to better understand a user's intention and check against those actions. A well-integrated multimodal system can yield significant levels of mutual disambiguation. Further examples include systems that integrate speech and lip movement recognition, or a system where a user both writes on a shared document whilst speaking the same terms. The fusion of speech and lip movement are tightly coupled modalities which helps make a system more robust, by enhancing recognition accuracy even when one signal is compromised. Semantically rich modalities such as speech plus pen input use late fusion techniques to employ multiple independent recognitions. Each recognizer can be trained independently over large amounts of uni-modal data, which is usually more readily available.[10]

Redundant, Complementary & Cross Modal Interactions
With MMUI, users naturally make use of the different characteristics of each modality to deliver information to an interface in a concise way. This tends to be either redundant or complementary multimodal input. Redundant inputs are when a user inputs the same data in two different modes e.g. User speaks the same word they are writing down. The degree of redundancy is affected by the level of cognitive load users' encounter while performing a task. Complementary multimodal inputs are when a user has to use two different complementary inputs to make their intentions known, e.g. when a user draws a circle on the map and speaks "add hospital".

Systems are obviously helped by redundant inputs as it gives the system 2 chances to recognize input and cross check. During interpersonal communication, people will often give spontaneous speech and manual gesturing involving complementary rather than duplicate information between modes, e.g. say "let's go over there" and point at place they wish to go, but they may also use redundant input if they think a person (like a system) will not understand them e.g. mouthing "check please" whilst making hand gesture for check when in a restaurant.[10]

From a system output point of view, if a user receives the 'same' information content across invoked sensory modalities, this is known as cross modal interaction[14]. Cross-modal interaction refers to situations where characteristics of one sensory modality may be bi-directionally transformed into the characteristics of another, e.g., audio ↔ tactile[15]. This gives the advantage of redundancy, such that if one sensory modality is knocked out (e.g. due to noisy environment) information is still received. This would have benefits over output systems that required user to pay attention to more than one modality to get all relevant information and thus reduce risk of information processing overload.

Simultaneous or Sequential Users
Users when they cognitively process information in different modes may do so either simultaneously or sequentially. People may do either depending on the situation, but more often people have this inbuilt and will lean one way or the other[10]. The following table describes the differences between these two types of user:

Table 1- Differences between Simultaneous and Sequential system

Simultaneous	**Sequential**
Simultaneous integrators overlap multimodal elements at least partially	Sequential integrators do not overlap multimodal elements
In a speech-pen interface: Simultaneous integrators will begin speaking while still writing	In a speech-pen interface: Sequential integrators will finish writing before speaking
During demanding session (when dealing with errors or doing more difficult tasks), simultaneous integrators will more tightly overlap.	During demanding session (when dealing with errors or doing more difficult tasks), sequential integrators will further increase their intermodal lag.

The above shows that these different ways that a user may process information is important when designing systems for users in high stress or cognitively demanding situations.

Vehicle Interface Studies
Not surprisingly, in-vehicle systems have been studied to understand safety concerns of having attention divided between the systems and the road and MMUI have formed part of these studies. In particular numerous in-vehicle systems have been introduced that use auditory prompts and voice control features to avoid direct competition for driver's visual and manual resources. Numerous studies though have shown that even auditory based secondary tasks impact driver performance and so Yang et al[12] wanted to assess if auditory tasks induced similar levels of workload as visual ones. Participants were asked to do visual tasks in their mind whist following a car. Results showed that whilst there were similar levels of self reported workload, visual tasks had a greater impact of involuntary steering wheel movement. Recarte et al[16], showed that even without an interface, if the user has to perform a visual task in their head it can affect their attention to things like glancing at side mirrors which can have safety implications.

RESEARCH FOCUS

Research – Outstanding Questions
From the literature research, there were a number of findings regarding MMUIs, raising some important questions that need to be better understood when designing communication systems for users in high stress environments such as public safety workers:

1. How do MMUIs perform when in environments rich in stimuli (noise, visual distractions etc)? How to design the MMUI to accommodate this?

2. What will Cognitive Load be on different types of MMUI systems (redundant, complementary, mixed) or if you receive a feedback of the same modality as input in a I/O feedback loop e.g. you get a vibrate to acknowledge gesture or tone to acknowledge voice etc whilst also in different environments?

3. Even in normal real world scenarios does a MMUI help make users more efficient? (e.g. faster, less errors, obtain richer information to make better decisions).

4. Do Complementary or Redundant I/O systems fail or perform worse when user under pressure? Is there a trade-off? e.g. Complementary system might work best for normal usage, but then no good in pressure situation.

Pilot Study Outline
To help get some initial answers to the above questions, the authors looked to design a simulator that could expose a user to varying degrees of modal stimuli from the environment as well as include a communication system that had varying degrees of input and output modalities and run a pilot study to understand any usability or experimental design considerations for future experiments that would focus more closely on different use cases involving communication device users.

For this initial pilot study, a driving simulator was created that could incorporate increasing stimuli from the environment (engine noise, rain etc) and involve primary and secondary tasks that participants would be familiar with – driving a car whilst following navigation instructions whilst using a communication device to have a conversation with someone. The intent was to treat this as a usability study of a communication system to understand if design elements such as complementary or redundant modalities impacted a user's performance and how offering an increasingly multimodal system affected users as tasks became increasingly rich in stimuli to give an indication if MMUI could help with users in stressful situations.

PILOT STUDY

Use Case Development
Before developing the pilot test, use cases had to be identified that could potentially accommodate multiple modalities. A number of scenarios including a delivery service scenario and event managing scenario were brainstormed being mindful of potential modalities that could be used (both input and output), how often these would be experienced and that could be developed in a limited timeframe (due to staff resources). These use cases were then storyboarded, incorporating multi modal stimuli both from the environment and from the systems the participant would use.

From this exercise the delivery service scenario was chosen as this driving use case provided more possibilities to accommodate multiple modalities. Table 1 describes the input, output and feedback loop elements that could be incorporated in the simulation.

After choosing this suitable use case, a few research questions were outlined to show a clearer direction of this study. (1) How do MMUIs perform in environments rich in stimuli (noise, visual distractions, etc)? (2) Which design elements of a MMUI potentially help? (3) Is cognitive load on a user affected by different multi-modal systems (redundant, complementary, and mixed) (4) Can MMUI help make users more efficient in terms of less errors and obtain richer information to make better decisions?

Table 2 – Descriptions and example of use cases components

	Description
Input	Actions triggered by users to carry out certain tasks with the system. From the delivery service use case, in order for user to get fastest route to destination, user has to pin point a location and then user get to input to choose the most ideal route. Examples of input in this use case are, touch input (while user is interacting with car dashboard), audio input (user makes voice command to the system), gesture input (mainly on car dashboard) or combination of above input modalities.
Output	After user pin-point a location, system will then show the traffic status and alternative routes on car dashboard as an output. Outputs are mainly from the systems and triggered by facilitator.
Feedback Loop	With any triggering action by users, a loop feedback is included to keep user alert while they are interacting with the system. Examples of loop feedback are, haptic (from car dashboard and steering wheel), audio feedback from system.

As discussed earlier, two multimodal systems of interest to investigate are redundant and complementary. The simulation therefore had to incorporate these elements within the user tasks. Storyboarding of the delivery service use case was further developed to include these systems in varying degrees of modal stimuli. In the use case, there are three main parts which each consists of three ascending level of system feedback (Table 2). According to the storyboard, Task 1 has the lowest level of stimuli, only visual feedback; on the second section, level of stimuli increased with audio feedback; Third section of the storyboard has the highest level of stimuli.

Simulator Development

The simulation had to be developed quickly and with limited resources. Therefore the simulator had to be controllable by the users with elementary input devices. Before building the simulator, a few explorations on development software was carried out to understand each program's capabilities. During the exploration, we came across three promising programs which were 3DRad (3D Rapid Application Development), Flash Develop and Scratch. Both 3DRad and Flash Develop have the capability of creating 3D environments. In 3DRad, it has a library of "intelligent" components (objects) which allows the developer to pick and choose while setting up the virtual environment without any coding. Basically, each independent object's default state is visually set by navigating the virtual space. Then the engine is started and objects' initial condition is evolved to achieve specific behaviours, like, for example, acceleration, attraction, friction, etc. However, 3DRad has limited objects in making a driving environment and appeared to need a lot of objects to be created (and therefore take a lot of time to create). Flash Develop is a free and open source code editor or an integrated development environment (IDE) which is mainly used to develop video games, web and desktop applications. This program allows open source coding and has a lot of flexibility in terms of the resulting applications running on various platforms like, for example on Microsoft Windows, Mac OS, Android or iOS. Although this program is versatile in creating a web simulator, insufficient time and resources might be difficult for a developer in handling the project in a short period of time.

Scratch is a different platform compare to 3DRad and Flash Develop. It is a free educational programming language in allowing user to create a 2D virtual environment by using block-based programming. In Scratch, it helps the developer gain an understanding of the fundamentals of programming before moving on to other programming languages. Without writing a code line, Scratch enables developers to program by dragging blocks from the block palette and attaching them to other blocks like a jigsaw puzzle (also known as drag-and-drop programming). These structures of multiple blocks are called scripts. Although Scratch can only be developed online, it allows developers to create, remix and collaborate with existing Scratch projects. Scratch's user interface is fairly straightforward: From left to right, in the upper left area of the screen is a stage area, featuring the live simulation and all sprites (layers) thumbnails listed in the bottom area. There are many ways to create sprites and backgrounds. In this project, sprites were drawn manually with "Paint Editor". Each sprite was drawn layer by layer (i.e. road, car framework, the dashboard, steering wheel, etc.). With a sprite selected in the bottom-left area of the screen, blocks of commands can be applied to it by dragging them from the Blocks Palette onto the right area of the screen (Script Area), containing all the scripts associated with the selected sprite.

Fig. 3. Interface of driving simulation

Throughout the process of system design, a few rounds of testing, reviewing and gathering requirements were carried out in improving the driving simulator. Lastly, the final system design is as shown at Figure 3. Figure 3 shows a simple interface of a first-person view driving, with a basic interior of a steering wheel, speedometer and a decorative radio. Also, information is displayed on the car windshield; this is to mimic a car head-up display (HUD). A user can interact with the simulator in a multimodal approach based on the instruction shown on the HUD with a combination of keyboard, speech recognition and text to speech. For instance, by pressing keys on a wireless keyboard, a user can control the simulated vehicle to accelerate, making turns and slow down. On the other hand, the driving environment and information on the HUD are controlled by the facilitator via a personal laptop and radio.

Table 3 – Description of delivery service driver use case

Task	Description	Users Control		Feedbacks			Environ-ment
		Redundant	Complementary	Visual	Audio	Haptics	
1	David needs to get the direction to customer A house David wish to call customer A from Location A While David is on the phone, David needs to follow direction on HUD End phone call	Pin point the location AND THEN [Select traffic status OR Speech "***Tina, Show Traffic Status***"] AND THEN Zoom into the traffic	Press the button to initiate speech mode AND THEN "***Tina, bring me here***" AND Point on a specific point on dashboard	✓			Normal Driving
2	Suddenly a phone call interrupt David while he is concentrating on GPS navigation Pick up phone call End phone call Random notification pop out (Alert Notification)	[Press on a button to initiate Gesture OR Speech mode] AND THEN Gesture over dashboard to pickup call OR speech: "***Tina, pick up the call***"	Gesture over dashboard OR Press on a button to initiate speech mode AND "***Tina, pick up the call***"	✓	✓		Normal driving + Raining
3	Message pop in David view and listen to message Another message pop in David hold message More random notification pop out (Alert notification)	Gesture over dashboard to view message AND THEN Speech: "***Tina, view the message***"	[Press on a button to initiate Gesture or Speech mode AND Gesture over dashboard to view message	✓	✓	✓	Normal Driving + Music

***Note: (1) *Tina* is an intelligent system used in infotainment and represents a driver assistance system in this use case. (2) Visual feedbacks: Road information and notification on HUD (3) Audio feedbacks: feedbacks that usually in synchronous with visual feedbacks when certain action is accomplished. (4) Haptic feedback: A radio with vibrator is attached to user which controlled by facilitator.**

In order to enhance the simulation experience, an application was designed to imitate part of a car dashboard and this application was displayed on a 7-inch ASUS nexus tablet that could be placed in an appropriate position next to the driving controls. This application was created through Appnotch, which is a cloud-based mobile application builder with drag & drop controls. The idea of having a "car dashboard" is to replicate an intelligent partner system (we named this system *Tina*) which worked in association with the HUD to allow users to interact with the system by voice or touch (dashboard within reach) and receive information both at the dash and screen level visually. With the access to Appnotch library of templates, themes, styles and asset, this prototype is made up of three main screens. Generally, the "car dashboard" or *Tina* will assist user in suggesting the shortest route. Based on figure 4, from left to right, it is the main interface plus a map showing user's current location. On the same interface, *Tina* will suggest a few locations based on the delivery orders (name in Location A, B and C). Once the user chooses a location, *Tina* will then suggest a few routes with live traffic updates with estimated time of arrival. As a result, the facilitator will be responsible in making sure these two programs are running under control during the pilot testing.

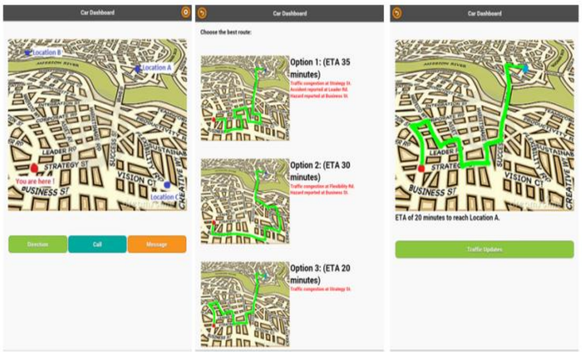

Fig. 4. "Car Dashboard" application user interface

Experiment Procedure
-Subjects

Participants were recruited internally from Motorola Solutions in Penang Malaysia. Ten individuals participated in the study, five males and five females. There is no prerequisite needed for this pilot study.

Equipment
The following materials were prepared for every test:
•A wall-mounted LCD television (simulated HUD) (Figure 5a)
•A wireless keyboards (user's controller) (Figure 5a)
•7-inch ASUS Nexus tablet (Car dashboard) (Figure 5a)
•Personal laptop (control station by facilitator) (Figure 5b)
•A pair of radios (one programmed with vibrator to give haptic feedback and another one for facilitator to trigger the haptics) (Figure 5c)
•Two sets of facilitator use case scripts (Redundant and complementary groups)
•Ten sets of feedback forms (Each five sets for redundant and complementary groups)

Fig. 5(a). Participant's simulator setup

Fig. 5(b). Facilitator's control station setup

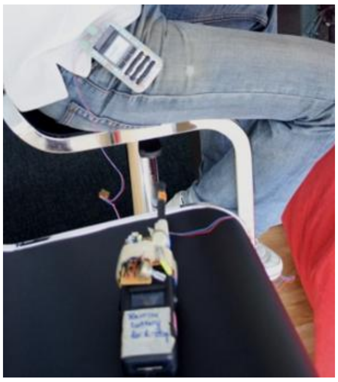

Fig. 5(c). Haptic devices

Procedure

There were 10 participants in this study. The participants did not need any special skill, prepare in advance or bring anything to participate. To reduce the causal effects of other factors, the following controls were applied:
i. Each participant was randomly assigned to one of the two groups (Redundant or Complementary).
ii. The participants were trained based on the multimodalities group they were in. (Redundant modality controls and feedback are different from complementary modalities)
iii. The order of each task was randomized for every participant.
iv. The participants worked in the same simulated environments.
v. The pre-experiment survey was started upon before the experiment. A NASA-TLX questionnaire was administered after each task. The post-experiment survey was started immediately after the participants had finished their final task.

The following procedure was followed for each of the experiments:

i. The details of the study were precisely explained to each participant.

Table 4 – Cognitive load questionnaire

Subscales	Questions	Rating
Mental Load	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?	Low, Medium, High
Temporal Demand	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?	Low, Medium, High
Performance	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?	Good, Average, Poor
Frustration Level	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	Low, Medium, High

ii. Participants completed a pre-questionnaire for demographic information.
iii. Each participant was then randomly assigned to either the redundant or complementary group.
iv. The participant completed a training simulation beforehand. In the training simulation the facilitator would explain the interface components and the multimodal interactions. The facilitator would then guide the participants through the experiments by controlling the simulator.
v. Each participant completed three tasks. The order of the tasks was randomly assigned.
vi. After each task, the participant completed the NASA-TLX questionnaire.
vii. Upon finishing the last task the participant completed a post questionnaire session.

Questionnaire
At the conclusion of each task, participants were asked to answer a set of questionnaire, which is derived from NASA-TLX Rating. NASA-TLX Rating is a subjective workload assessment tool. It allows users to perform subjective workload assessments on operator working with various human-machine systems. This rating procedure obtains an overall workload score based on a weighted average of ratings on six subscales. These subscales include mental demands, physical demands, temporal demands, own performance, effort and frustration. However, in this pilot study, only four subscales were used in the cognitive load questionnaire, which are, mental load, temporal demand, performance and frustration level as these related to the task in hand (physical and effort subscales were not relevant as no real physical activity taking place) (Table 4). To record each participants cognitive load on every task, this questionnaire was then provided after every task was completed.

RESULTS
This section summarizes and discusses the results obtained from the driving simulator. A main objective of this pilot study is to examine how redundant and complementary modalities impact user's interaction and mental load with their primary tasks. To validate this pilot study, results from the NASA-TLX questionnaire and post-test results were explored.
There were 10 participants for this study who performed a total of 30 tasks. Of the 10 participants, 5 participants were male and another 5 were female.

Table 5 shows the overall result of both groups, redundant and complementary, on three different tasks. Task 1 from both redundant and complementary modalities group was generally more complicated for participants. This is to be expected as they only received visual feedback on their tasks whilst performing the primary visual task of driving. Data from both redundant and complementary groups showed there were six out of ten participants who tended to miss the direction notification while their focus was occupied by other information on the HUD. The most likely situations where users missed driving directions happened where a singular feedback modality (visual in this instance) was too overwhelming for users. For instance, a driver can be reading a text message on the HUD while he/she needs to also concentrate on visual-based directions on the HUD; if the driver's attention was occupied by visual text messages, then the higher the chance their concentration drew away from the primary task. Overall scoring for Task 1 from cognitive load scores were that

Table 5 – Redundant and complementary group result

Task	Redundant			Complementary		
	Task Result	Cognitive Load	Score (n/3)	Task Result	Cognitive Load	Score (n/3)
1	3 out of 5 participants miss the direction notification while they are concentrating on the HUD information	Mental Demand	2.2	3 out of 5 participants miss the direction notification while they are concentrating on the HUD information	Mental Demand	2.6
		Temporal Demand	2.4		Temporal Demand	2.6
		Performance	2.6		Performance	2.8
		Frustration Level	2.4		Frustration Level	2.6
		Total (N/12)	**9.6**		**Total (N/12)**	**10.6**
2	Participants tend to forget to end the call after the conversation	Mental Demand	2.6	Participants tend to forget to end the call after the conversation	Mental Demand	2.0
		Temporal Demand	2.2		Temporal Demand	2.0
		Performance	2.2		Performance	2.2
		Frustration Level	2.6		Frustration Level	2.0
		Total (N/12)	**9.6**		**Total (N/12)**	**8.2**
3	• Although most of the participants find direction with audio more relatable, they find it interfering while there is an incoming message • Most of the Participants find haptics feedback is useful. However, there's one participant find haptic feedback is distracting.	Mental Demand	1.2	Although most of the participants find direction with audio more relatable, they find it interfering while there is an incoming message	Mental Demand	1.4
		Temporal Demand	1.4		Temporal Demand	1.4
		Performance	1.4		Performance	1.0
		Frustration Level	1.0		Frustration Level	1.4
		Total (N/12)	**5.0**		**Total (N/12)**	**5.2**

complementary modalities obtain higher cognitive load (score 10.6 over 12) than redundant modalities (score 9.6 over 12). In general, in the complementary group, all four subscales show a higher score than the redundant group by the range of 0.2 to 0.4.

In Task 2 (raining scenario), the level of stimuli was increased both in the environment (audible rain) and with the addition of audio system feedback. Users were also asked to perform an extra combination of modalities while carrying out the task (having to perform a gesture to end call). Unintentionally, users have the tendency to neglect this call end gesture action because it is often perceived as needless, though it could be argued if user had intent to make a call (rather than make a call because they were instructed to do so) with perhaps some kind of audio indication e.g. 'white noise' on dead line then the user may be prompted to remember to carry out the end call gesture. Based on the result on Table 4, the redundant group scored slightly higher for cognitive load than the complementary group. The addition of audio feedback though did appear to reduce the mental load needed when using a complementary interface.

As the task environment builds up, Task 3 has the highest level of stimuli. All participants find the audio feedback (short audio feedback follow by *Tina* voice over the direction, i.e. *short audio feedback* THEN "In 200 M please turn left" *Tina*) on navigating is very helpful. However, some participants commented that the number of audio feedbacks were too many and could interfere with important information. In Task 3, one participant found the haptic feedback too repetitive. The cognitive load scores again are almost identical showing little difference between redundant and complementary systems, but the scores for Task 3 are markedly lower than the other tasks, indicating that the combination of audio, visual and haptic feedback could be beneficial. It should be noted though that Task 3 did not involve a phone call but viewing and listening to messages. This also involved some decision making by the user and so they had more cognitively speaking to do in Task 3, though perhaps with less audio to listen to which occurred in the phone call in Task 2. Therefore a drop in cognitive load score might not just be due to the increased modalities, but having less audio to listen to, especially overlapping audio of call and navigation instructions which can be difficult for user to process.

The following sections are anecdotal feedback from both redundant and complementary groups:

i. Redundant Modality (Anecdotal feedback)

"I suggest that an intelligent system should be design in a way that it will prioritize tasks based on the situation" (Participant 2, redundant group)

"I feel like Task 2 (raining) needs improvement in terms of creating an immersive environment because I think at such situation, HUD information might be distracted by the situation." (Participant 3, redundant group)

*ii.*Complementary Modality (Anecdotal feedback)

"I think the End Phone Call function is redundant." (Participant 1, complementary group)

"I personally feel that haptics, visual and audio feedbacks are helpful in terms of alerting users. However, overload information or feedback will cause confusion. And, I think system designer should come out the most adequate number of feedbacks." (Participant 3, complementary group)

CONCLUSION

In general, cognitive load results would indicate there is no real difference between redundant and complementary modalities, but that with the addition of haptics to an audio/visual system then cognitive load could be reduced when the user has to deal with multiple stimuli from their task and environment and therefore shows the potential benefits of MMUIs.

It would appear that if trying to further study the usability issues of MMUI systems in real world tasks and applications then designers of the systems should be free to decide whether complementary or redundant modes are most appropriate at any one stage of interaction for a user. Results would also indicate that if too much conflicting stimuli is received by the user e.g. when listening to both a phone call and trying to listen to sat-nav instructions and whilst watching the road whilst also looking for visual directions then the user will find the task quite difficult. If a system could be more contextually aware to a user's circumstances then an adaptive interface could pick and choose which modality to give information to a user to help avoid cognitive resource conflict. Taking the findings of this study, next steps include the investigation of such adaptive MMUI interfaces. These interfaces would be adaptive not just in terms of which modality to use, but how to use that modality e.g. where should visual information be placed, where should audio streams come from and to use more sophisticated haptics which in themselves could give actual directions to user rather than just be an alert.

The authors would also like to expand the tasks to get more realistically stressful, either by incorporating the system into a real world driving task, but also by applying time pressure constraints. In this study, due to time limitations, users experienced variously demanding tasks as the stimuli and system modalities increased, but it would be beneficial if participants experienced all the tasks and environments with one set of modalities. This will likely need a participant to only experience one set of modality and then compare against other users with differing sets of modalities.

Research literature also indicated there are different types of users, sequential and simultaneous. This study was not able to differentiate between these users, so an appropriate assessment method will be looked for to screen participants beforehand to see if results differ between these groups and whether it is realistic for one system to be able to cater for both types.

ACKNOWLEDGEMENTS

Thanks to the participants from Motorola Solutions Penang for taking part in this study and thanks to UniMas, Kuching, whose internship support facilitated study.

COMPETING INTERESTS
There is no conflict of interest.

REFERENCES

1. Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker,

S., & Maloor, P. MATCH: An Architecture for Multimodal Dialogue Systems. *Association for Computational Linguistics* (ACL), 2002, 376-383.

2. Pielot, M., Heuten, W., & Susanne, B. PocketNavigator: Studying Tactile Navigation System In-Situ. *Association for Computing Machinery*. 2012. ACM 978-1-4503-1015-4/12/05.

3. Sebe, N. Multimodal interfaces: Challenges and Perspectives. *Journal of Ambient Intelligence and Smart Environments*, 2009 1, 19-26.

4. Oviatt, S. Multimodal Interfaces. Handbook of Human-Computer Interaction. 2002. Lawrence Erlbaum: New Jersey.

5. Oviatt, S. Human-centered design meets cognitive load theory: designing interfaces that help people think. *In Proceedings of the 14th ACM international conference on Multimedia* (2006, October). 871-880.

6. Wickens, C.D. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 2002 159-177.

7. Kortum, P. HCI Beyond the GUI, Morgan Kauffman Publishing, San Francisco, 2008.

8. O'Malley, M.K., Gupta, A. "Haptic Interfaces" HCI Beyond the GUI, Morgan Kauffman, 2008.

9. Peres, S.C., Best, V., Brock, D., Shinn-Cunningham, B., Frauenberger, C., Hermann, T., Neuhoff, J.G, Valgerour, N., Stockman, T. "Auditory Interfaces" HCI Beyond the GUI, Morgan Kauffman, 2008.

10. Barthelmess, Paulo, and S. L. Oviatt. "Multimodal interfaces: combining interfaces to accomplish a single task." HCI Beyond the GUI, Morgan Kauffman, 2008.

11. Cohen, M. H., Minimizing Cognitive Load, in Choenm M.H., Giangola, J. P., & Balogh, J. Ch9, Voice user interface design, 2004, Addison-Wesley Professional.

12. Yang, Yan, et al. "Exploring differences in the impact of auditory and visual demands on driver behavior." *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 2012. 173-177.

13. Mankoff, Jennifer C., and Gregory D. Abowd. Error correction techniques for handwriting, speech, and other ambiguous or error prone systems. GVU Technical Report, GIT-GVU-99-18, 1999.

14. Lewkowicz, D.J. Development of intersensory perception in human infants. Development of Intersensory Perception: Comparative Perspectives, 1994, Norwood, N.J.: Lawrence Erlbaum Associates.

15. Hoggan, E., and Brewster, S.A. Designing Audio and Tactile Crossmodal Icons for Mobile Devices. *In ACM International Conference on Multimodal Interfaces*. ACM Press, 2007, 162-169.

16. Recarte, Miguel A., and Luis M. Nunes. "Effects of verbal and spatial-imagery tasks on eye fixations while driving." *Journal of experimental psychology*: 2000, Applied 6.1: 31.

17. Sarter, N.B. Multimodal information presentation: design guidance and research challenges. *International Journal of Industrial Ergonomics*, 2006, 36, 439-445.

18. Dumas, B., Lalanne, D., and Oviatt, S. Multimodal Interfaces: A survey of principles, models and frameworks. *Human Machine Interaction*, 2009, 5440, 3-26.