VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

# A MULTIMODAL INTERACTION FOR MAP NAVIGATION AND EVALUATION STUDY OF ITS USABILITY

Nor Azman Ismail and Yee Yong Pang
Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
E-Mail: azman@utm.my

## ABSTRACT

To meet the challenges of ubiquitous computing and an increasingly map usage population; researchers have been trying to break away from the traditional modes of interaction. Over the past decade years, researches in this domain suggest that Multimodal User Interfaces (MUI) now provide maturity and affordable opportunities, which may be appropriate for society transformation on the interaction styles. We have developed a MUI prototype application, called MapNI, to help users carrying out everyday activities such as navigating a map. MapNI use user-defined hand gestures to perform a different range of tasks via a map navigation interface. This paper describes the MapNI development and reports its usability evaluation. We conclude that this inclusive technology offers some potential to improve the independence and quality of life of society, although there remain significant challenges to be overcome.

**Keywords:** multimodal interaction, map navigation, vision-based hand gesture, speech recognition.

## INTRODUCTION

Given the recent popularity, and tablet interfaces, and movies like "The Minority Report", the gestural interfaces offer an alternative solution for windows, icons, menus, and pointer (WIMP), in that they do not restraints by traditional interface paradigms (Hansen, Hourcade, Virbel, Patali, & Serra, 2009; Wilson, 2004). Gestural interfaces can be even more direct than their graphical user interface counterpart. With a touchscreen or tablet, the item that users touch is the object that they are interacting with. The advantages in these interfaces, for example, pinching to zoom the maps are just feeling so natural, the metaphor is very clear, the reaction is immediate, and the result is predictable (Grandhi, Joue, & Mittelberg, 2011). Therefore, in gestural interfaces, like in all others, the applications work best when the feedback is clear, and the interface provides hints through metaphors or visual signals that help users understand what they can do within the user interface (Neale & Carroll, 1997).

Extend from it, previous researchers have shown that gesture based and speech is the popular input modalities interactions in the map navigation (Jacobson & Sam, 2006) (Kamel Boulos *et al.* 2011) (McClymont, Shuralyov, & Stuerzlinger, 2011). This recent existing technology such as vision-based hand gesture recognition allows humans to interact with the computer naturally without wearing any mechanical devices on the body (Pang, Ismail, & Gilbert, 2010). On the other hand, the audio interaction provides a considerably natural and intuitive method in navigating the interface. For instance, speech dialogue is generally acknowledged good in handling, quick and routine interactions, like searching address or location. Communication using speech has been a rich channel just like communication among human to human (Krum, Omoteso, Ribarsky, Starner, & Hodges, 2002). The single modality interaction, however, is insufficient and inadequate to cover the entire usage context for certain systems in term of interaction. For example, a map navigation system is more complex for a user to give the coordinates by only speech (e.g., "I want to go one and a half centimeters to the right of the school"), but it is easier and more precise to use a deictic gesture to point out the coordinates.

The convention map navigation system interface often encounter the capability limitation of the device that lead to difficult for users to interact effectively, but also forced users to spend more effort to exchange information with the computer. Traditional devices such as a keyboard input though there are many possibilities, however, to remember each command will cause the cognitive burden and the device itself lack of flexibility. On the other hand, the mouse is very simple and convenient, but its own standard device with three buttons cause its functions very limited and can only perform very simple instructions. Moreover, wearable devices require prerequisites such as clean backgrounds, noise free environments, and costly high-end cameras are strictly demand which create an unnatural interaction experience with the users. And because the intuitive and flexible user interface usually requires multiple modalities, innovative input technology integration is becoming one of the solutions. Hence, the first challenge is to investigate the users' preferences modality of map navigation interface which can enhance the quality of user interaction by providing an intuitive and flexible way of multimodal interaction that does not support in existing map navigation interface.

Multimodal interaction that provides multiple modes for the user has a better performance in map navigation interface because it increases the navigation performance and its naturalness for mankind. Therefore, this study proposes multimodal map navigation interfaces; which considers the human-computer interaction perspectives that include the people, technology and context.

## RELATED STUDIES

Until recently, researchers started to study on the techniques that can combine with computer vision or sensory devices to reach a real wireless environment for gesture recognition system. For example, web camera, a type of computer vision devices, has become popular as a gesture input device in many domains with great potential to be applied in daily activities (Kamel Boulos *et al.* 2011). A recent prototype operator with motion control using natural interaction towards map simulator had been created by Kamel Boulos *et al*. (2011). They formulated their own specified gesture dialogues implemented into third party hardware and software. In this prototype, it allows user to apply single hand, both hands, and body gesture languages to represent different functions. For example, 'pan' is engaged with one hand that moves around the map in any direction and etc. However, the shortcoming of this prototype is that it does not have the speech interaction, even though speech is more natural in multimodal interaction.

There are researches that have been done to the extent of developing input modality among human and computer; however, the ideal of input modality should fulfill the requirements of an intuitive, natural and ergonomic interface design (Karpov, Carbini, Ronzhin, & Viallet, 2008). Some studies focused on the factors that govern the interface design rather than on the invention or improvement of the input modalities accuracy. Human interaction should always be in correspondence to human sensory system, which are sight, hearing, touch, smell and taste (Tzovaras, 2008). Thus, the multimodal interaction needs to be designed based on these senses. Body gesture and speech are the main communication methods, especially when human-human interaction is concerned and these have been the center of many multimodal interfaces related research (Pavlovic, Sharma, & Huang, 1997). Multimodal interaction provides multiple classes or modalities of interaction to a user. Multimodal interfaces can experience a decreased error rate with compared to single modal interfaces. Bolt's proposed an early example of multimodal interaction, "Put That There" which incorporated speech recognition and pointing hand gestures (Bolt, 1980).

A part that many users are not aware of is that the multimodal interaction has already been implemented for a long time ago before it has been proposed. For example, in kiosk machine (Johnston & Bangalore, 2004), game (Kamel Boulos *et al.* 2011), mobile phones (Oviatt, 2000) and so on. The advancement of hardware had also lightened the limitation of software. With high performance hardware, developing a complex application has become possible compared to a few decades ago. For instance, Microsoft Kinect is a cutting-edge motion sensing input device first introduced in the year 2010 to enhance the Microsoft Xbox. After that, many researchers have or tried to utilize Kinect in education (Villaroman, Rowe, & Swan, 2011), treatment, surgery, and so on to develop interactive prototypes in a shorter time. Kinect has jumped on the bandwagon because researchers do not need to develop similar devices from scratch anymore and it is affordable.

In general, the interaction between the user and an interactive map takes place in either unimodal or multimodal format which elaborates in Table-1. The unimodal format has only one modality, which can be further divided into three categories, namely, the physical sensor-based, visual-based and audio-based modality (Karray, Alemzadeh, Saleh, & Arab, 2008). The physical sensor-based modality is widely used in daily activities, such as the touchscreen on a smartphone (Kim, Kim, Bae, & Lee, 2006). The visual-based and audio-based modalities are less common (Fang, Chai, Xu, & Wang, 2009; Jacobson & Sam, 2006; Kamel Boulos *et al.* 2011).

**Table-1.** Related studies on map interactions.

| Interaction Style | Related Study | Pros/Cons |
|---|---|---|
| Hand gesture, body gesture | (Fang et al., 2009; Kamel Boulos et al., 2011; Pang & Ismail, 2012) | Passive methods; implicitly monitor users' movements to infer recognition (Oviatt, 2003). |
| Speech | (Jacobson & Sam, 2006) | Improves the accessibility and usability of an application, but the sensitivity to background noise may cause problems with the application. |
| Speech and pen-based (parallel) | (Cohen, McGee, & Clow, 2000; Doyle, Bertolotto, & Wilson, 2008; Jokinen, 2008; Oviatt, 1997) | Complement each other and increase the user performance of navigation, but pen-based input is neither intuitive nor natural to the user. |
| Speech and gesture (parallel) | (Krahnstoever, Kettebekov, Yeasin, & Sharma, 2002) | Close to the nature of human interactions, but the user navigation performance is restricted by the hardware limitations such as recognition accuracy. |

Next, from the studies of multimodal user interface support for map navigation interface, there are many studies on the multimodal interaction for map navigation in different input modalities such as haptic, hand gesture, body gesture, speech, speech and pen-based, gesture and speech. However, there is still more room in this research of multimodal interaction for map navigation because of the best knowledge in literature searching; there are no studies yet on the multimodal user-defined dialogue for map navigation interface. This field's multimodal interaction is still waiting for more explore.

## THE MAPNI APPLICATION: A PROTOTYPE

The architecture of the MapNI prototype is illustrated in Figure-1. Three module blocks is proposed, dealing with the input module sending raw data to interaction module which mapping the voice and hand gesture with the DM (dialogue manager) onto an interface command controller tailored to map navigation and, finally, display the feedback of such application in the application agents.

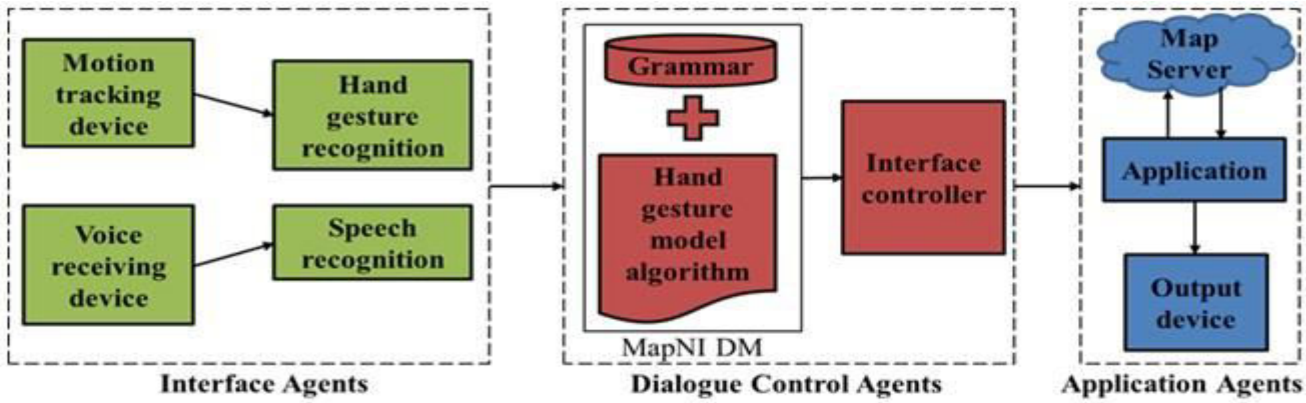


**Figure-1.** Architecture of MapNI.

The three blocks will be described further. The discussion will consider first the interface agents following by application agents and will close by aiming on the dialogue control agents' block, which plays the central role in the integration of the overall architecture.

### Interface Agents

In the depicted schema, the interface agents initial the motion tracking function and the sound receiving function. Both functions require an independent input device such as a leap motion controller to detect and track hand gesture movement, and microphone to receive the analogue sound in this study. Then, the parameter of hand gesture and sound signal were sent to their recognizer to further process. The hand gesture recognition processes the raw image frames that captured from the input device. The raw data will process in several stages, image pre-processing, tracking, and recognition (Yee Yong *et al.* 2010). In the other hand, the speech recognition mechanism first digitizing the analogue sound receiving from the input device to digital form signal. The digital signals will later analysis in language modeling and statistical analysis in which knowledge of grammar and the probability of certain words is used to increase recognition and improve accuracy.

### Application Agents

In application agents, the information that is sent from dialogue control agents in the application is responsible to provide the feedback to the user via the output device. The applications consist of web browser and the interaction interface. The interaction requests are sent through a web browser to the map server for receiving designation, location information, the latter, map server post back the requested data and render the content in the web browser. The updated map information will be displayed in an output device.

### Dialogue Control Agents

The responsibility for the integration of interface agents and application agents block is on the dialogue control agents, which is in charge of mapping words (speech) and movements (hand gesture) into meaningful interaction commands. The designed grammars and hand gesture repositories are stored in the dialogue manager (DM). This study proposes a novel dialogue manager called MapNI DM which specifically for map navigation interfaces. MapNI DM analysis and translate the meaningful input data to process in the interface controller. The interface controllers trigger certain desktop events to the application agents. Thus, the dialogue manager is one of the important components that link the input modalities to the output programs.

### Integration Strategy

In this section, the integration of the component lifecycle between the interface agents, dialogue control agents and application agents can be described according to their functions. There are four key components to success the interaction which including user, multimodality, interaction manager, and the view. Figure-2 illustrated the key component of interaction.


**Figure-2.** Component lifecycle of interaction.

The user is the stakeholder who interest to interact with the map navigation interface. The user requirement of the user is to have the knowledge of performing the multimodal interaction dialogue. During the map navigation interface interaction, it has only allowed one user to perform the multimodal interaction dialogue. The user either makes the meaningful hand gesture and speech dialogue or the meaningless dialogue that will be captured in the next process in this lifecycle.

Multimodality component is the input modalities that contain functionality to capture audio and hand gesture movement. The input modality component captures the information from a user as directed by an interaction dialogue. In this research, the source that came from user is able to trace by the microphone and Leap Motion device. The microphone is the audio input device which received the speech of the user, capture the keyword from the user, filtering by using the grammar and finally convert speech to text. The leap motion controller is the gesture input device that responsible to track the hand gesture movement of the user.

The interaction manager is one of the most important components for managing a spontaneous human interaction with computer applications (Nakatsu, Nicholson, & Tosa, 1999). Based on the data received from the input modalities, the tracking situation and recognizing the human hand gesture and speech activity and action context, the interaction manager attempts to decide on if, when and how to present the response as requested by the applications. A protocol for interaction manager is developed following the principles and assumptions of user center interaction. The interaction manager plays the role in distinguishing predicted and spontaneous interactions and generating command reactions for map navigation interactions. The details of these functions are described below.

This study adopts the previous work on the defining dialogues for map navigation interfaces Pang and Ismail (2015). To explore the multimodal interaction in 2D map navigation interface, three types of gestures available with LeapMotion library are selected by using only one hand, right or left to simulate the find route, zoom, and pan functions. The tracking API embedded in LeapMotion library modifies the nature of Pointable, Hand, and SwipeGesture in order to perform the map navigation functionalities ("Leap Motion API Overview," 2015). These libraries' were selected because of the characteristic are similar to the defining dialogue. For instance, pointable can simulate the pointing and touch on the screen.

**Find Route - Pointable**

The Pointable class could be recognized as deictics gesture using index finger or one finger. To simulate the touch emulation, the Leap Motion software associates the touch zone based on a floating touch plane that follow to the user's finger movement ("Leap API: Pointable class," 2015). The Leap Motion predicts purposeful movements toward this plane as potential touch points. The Pointable class reports touch state with the property TouchZone and TouchDistance values. The TouchZone defines the values for reporting the state of a finger in relation to an adaptive touch plane. The TouchDistance is the value proportional to the distance between this finger and the adaptive touch plane. Figure-3 shows the touch zone and touch distance of using one finger. There are three states for the TouchZone, ZONE_NONE, ZONE_HOVERING, and ZONE_TOUCHING. If the finger is out of the hovering and touching zone, it assume as no finger detected by the device. If the finger moves close to the adaptive touch plane, it enters the "hovering" zone and the range of TouchDistance values are between 0 to 1 then return to the application. If the finger reaches or passes through the plane, it enters the "touching" zone with reports the range of values between 0 to -1 return to the application. The Pointable class is then implementing with the build in Touchless application to simulate as mouse to move the mouse cursor. The pseudo code to integration of speech with the touch emulation is shown in Table-2.

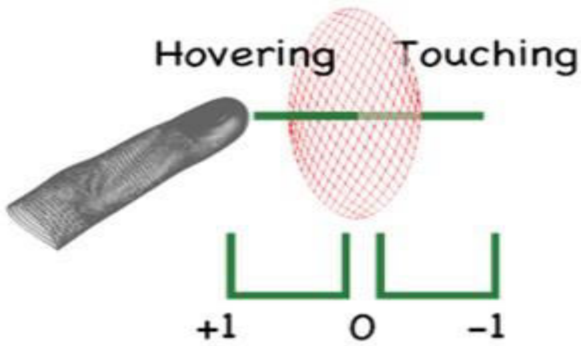


**Figure-3.** Touch zone and touch distance of finger ("Leap API: Pointable class," 2015).

**Table-2.** Touch Pseudo code.

| |
|---|
| If TouchZone equal to ZONE_HOVERING |
| If keyword "From Here" was heard |
| Set the starting location with simulate mouse left click |
| Else if keyword "To Here" was heard |
| Set the destination location with simulate mouse left click |
| Else if TouchZone equal to ZONE_TOUCHING |
| Set mouse left click simulate |

**Zoom - Hand**

The Hand class could be recognized as manipulate gesture includes palm and fingers. Figure 4 shows the sphere placing roughly at the palm when it was holding like a ball. Leap Motion software associates the sphere size based on the radius of a sphere fit to the curvature of the hand ("Leap API: Hand class," 2015). To simulate the zoom, the Leap motion device capture the hand frame by frame, the comparison of the sphere size of previous frame and the sphere size of current frame (there is a threshold value, e.g. previousSphereRadius – currentSphereRadius > threshold) will come out with two possible outputs, either zoom in or zoom out. The size of the sphere decreases when the fingers are curled into a fist, this will be considered as zoom out gesture and vice versa for zoom in. The pseudo code to integration of speech with the zoom is described in Table-3.

**Pan - SwipeGesture**

The SwipeGesture class represents a continuous swiping motion of a finger or hand ("Leap API: SwipeGesture class," 2015). To simulate the pan, there are several things need to consider such as the direction of swipe. The user might perform difference swiping action. The SwipeGesture provided the property of Direction which recorded the unit direction vector parallel to the swipe motion. It compares the components of the vector to classify the swipe as appropriate for the propose application. For example, to swipes horizontal, user compares the x value and when swipes vertical, user can compare the y value of Direction. Another issue need to consider is the speed of swiping. The intention of performing swipe gesture is important to this part. If the user performs the swipe gesture with no intention, normally the hand movement is slowed. Hence, the threshold for the speed of gesture is defined to observe the intention of user. Figure-5 shown the swipe gesture motion generated with a hand and Table-4 shown the pseudo code of Panning.

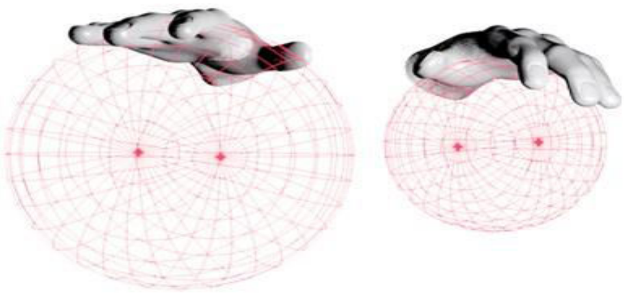

**Figure-4.** The sphere display at below palm ("Leap API: Hand class," 2015).

**Table-3.** The zooming pseudo code.

| |
|---|
| **If** previous SphereRadius - current SphereRadius greater than threshold value |
| **If** keyword "Zoom More" was heard |
| **Set** for the zoom out action twice. |
| **Else** |
| **Set** for the zoom out action |
| **Else if** previous SphereRadius - current SphereRadius smaller than threshold value |
| **If** keyword "Zoom More" was heard |
| **Set** for the zoom in action twice. |
| **Else** |
| **Set** for the zoom in action |

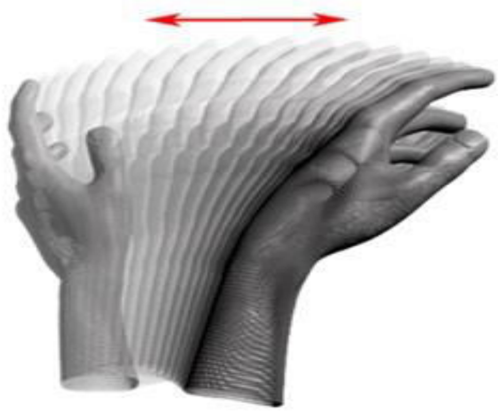

**Figure-5.** The swipe gesture motion of hand ("Leap API: SwipeGesture class," 2015).

**Table-4.** Panning pseudo code.

| |
|---|
| **Switch** *GetureType, SwipeGesture* was detected |
| **If** the swipe speed greater than the threshold value |
| **If** swipe direction x greater than swipe direction y |
| **If** the x value smaller than zero |
| **Set** keyboard arrow left to pan left |
| **Else if** the x value greater than zero |
| **Set** keyboard arrow right to pan right |
| **Else if** swipe direction x less than swipe direction y |
| **If** swipe direction y greater than zero |
| **Set** keyboard arrow up to pan up |
| **Else if** the swipe direction y less than zero |
| **Set** keyboard arrow down to pan down |
| **End Switch** |

**View**

View is an important and powerful component that virtually from the computer graphics is easily understood by the user. The broad range of graphics hardware and software systems is now available for applications in all fields. The main output device in a graphics system is a video monitor. Users are only able to take response when they see the presented information that contains GUI. The response contents of captured hand gesture and speech that processes in interaction manager will finally viewable in a display device through a screen. The interfaces of the view details will be explained in next section.

**MapNI Interface**

Figure-6 shows the interface of multimodal MapNI prototype of hand gesture and speech interaction. The interface and the functions are almost similar to that

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences



unimodal MapNI prototype except it is added with speech textbox. The function of speech textbox is to display the text that the user speech out within the available grammars.

**Title:** The title bar shows the name of the title for this prototype.

LeapMotion Status: the leap motion controller status display the current status of leap motion device, it is either connected or disconnect. When the leap motion device is not plug in or is not detected by the computer, the LM Status will display “Disconnected” and vice versa. The tracking status will indicate “On” when the hand is within the tracking angle of leap motion device and “Off” when is hand is off the tracking area.

**Indicator:** The indicator button is the hint to turn on/off the users’ indication during performing their hand gesture action. Table-5 shows the overall indicator available during the performance of hand gesture dialogue. The indicator divided into three categories; find route, zooming and panning.

**Action Textbox:** The action textbox present the action that the user performing. It is unlike the indicator, the action textbox is an active textbox that always shows the action that the user performance, it cannot be turning off by the user.

**Interaction Button:** The interaction button is the switch to turn on/off the map navigation action. The navigation action is enabling only when the particular button is turning on. This three buttons can turn on simultaneously.

**Speech Textbox:** The function of speech textbox is to display the text that the user speech out within the available grammars.

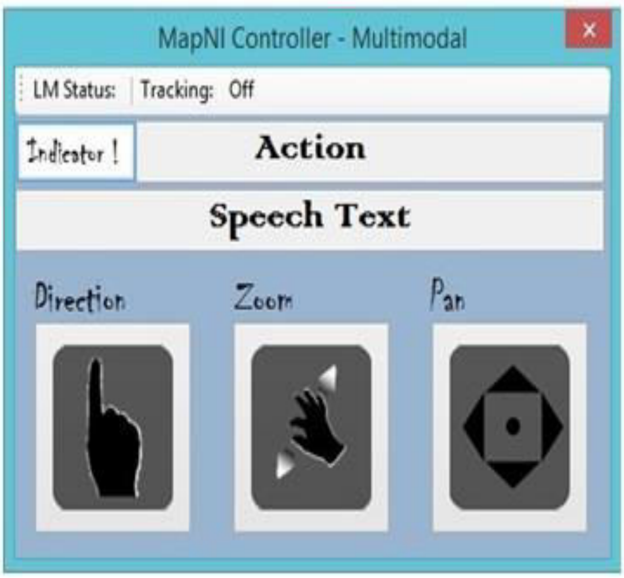


**Figure-6.** User interface of MapNI.

**USABILITY EVALUATION**

An evaluation study was carried out to test the user performance and usability of the purposed prototype. The time taken for participants to complete each task was recorded and the subjective rating on the usability was using the System Usability Scale questionnaire.

Twenty paid participants took part in the evaluation. Each of the participants were given some kind of token of appreciation for taking part in the evaluation. There are three tasks needed to complete as elaborate in Table-6. The time taken to complete each task will be recorded. After complete all tasks, they are asked to fill a System Usability Scale form to rate based on their favor. The modified SUS (Table-7) instrument has 10 statements that are ranging on a 5-point scale of strength of agreement (strongly agree to strongly disagree). Final scores for the SUS can range from 0 to 100, where higher scores indicate better satisfaction. Since the statements alternate between the positive and negative, participants spent some time to think to answer the survey carefully.

**Table-5.** Indicator of hand gesture available.

| Category | Available hand gesture dialogues indicator | |
|---|---|---|
| Find route |  Direction From |  Direction To |
| | Direction from | Direction to |
| Zoom |  Zoom In |  Zoom Out |
| | Zoom in | Zoom out |
| Pan |  Pan South |  Pan North |
| | Pan down | Pan up |
| |  Pan West |  Pan East |
| | Pan left | Pan right |

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences



**Table-6.** Task description.

| Task | Description |
|---|---|
| Find Route | To find the direction from Point A to Point B. |
| Zoom | Viewing the map more closely or more distant. |
| Pan | Move the map to Point C. |

**Table-7.** Modified SUS questions.

| The Modified SUS Statements |
|---|
| I think that I would like to use this application frequently |
| I found the application unnecessarily complex |
| I thought the application was easy to use |
| I think that I would need the support of a technical person to be able to use this application |
| I found the various functions in this application were well integrated |
| I thought there was too much inconsistency in this application |
| I would imagine that most people would learn to use this application very quickly |
| I found the application very awkward to use |
| I felt very confident using the application |
| I needed to learn a lot of things before I could get going with this application |

## RESULT AND DISCUSSION

The means and standard deviations of time completion task are summarized in Table-8. The panning task required more time (40.6 seconds) to complete as compared to other because panning consists of four types of movement compared to others. During the observation, users required more mental memory to switch between these four types of dialogue before performing the movement.

Figure-7 shows the time taken to complete the task for each participant. Panning has the greatest deficit which is from 20 seconds to 61 seconds. This is because different mental memory on individual which decide their time taken to recall the movement of the dialogue. The SUS scored by participants is shown in Figure-8, ranging from lowest 40 to highest 100. From the Figure-7 and Figure-8, it was found that the participants spend longer time in panning gave the score below 70 and only five participants gave the score 70 or above (Participant 3, 8, 10, 13, and 20). It shown that the dialogue with more mental memory required will reduced the usability of the MapNI. It is suggested that a lower mental memory dialogue will help to improve the usability of the multimodal interaction. However, the average SUS score for overall tasks is 63.1. This score reached the high marginal of the SUS score acceptability stated by Bangor, Kortum, and Miller (2008). An acceptability score is above 70.

**Table-8.** Time completion results by task.

| Task | Mean (seconds) | Standard deviation |
|---|---|---|
| Find route | 19.5 | 9.16 |
| Zoom | 24.4 | 15.18 |
| Pan | 40.6 | 14.81 |

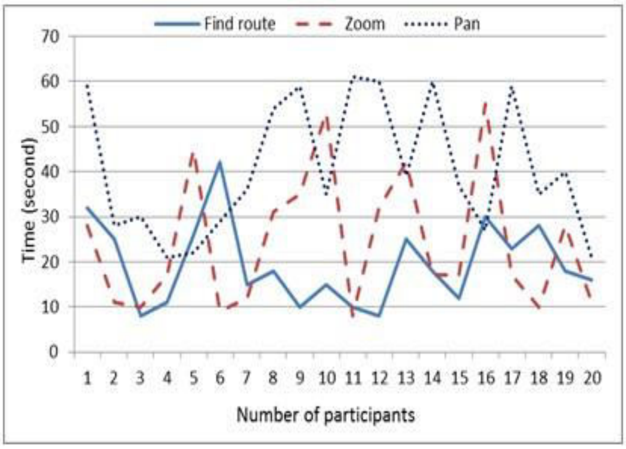


**Figure-7.** Time taken to complete each task by participants (n=20).

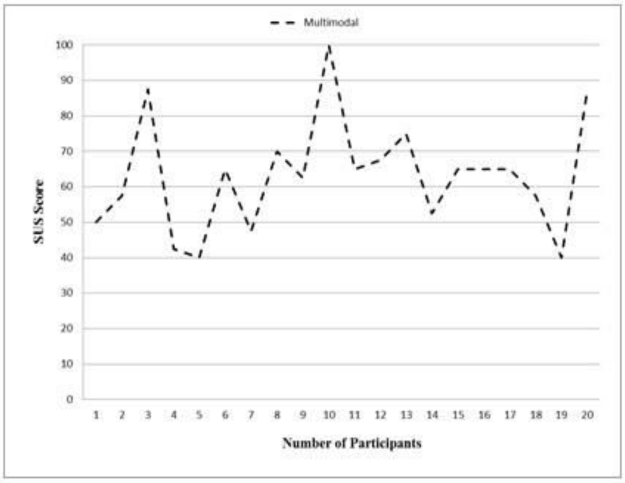


**Figure-8.** SUS scored by participants (n=20).

## CONCLUSIONS

This study began with a discussion of the implementation issues. Technology details that are of interest to this study were presented. A discussion of the architecture and components of MapNI prototype itself was given, including an explanation of user, multi-modality, and interaction manger and view components in relation with the design of the prototype. After the development of MapNI, there are needed for evaluation on their usability. The results show that there is a room to improve the MapNI, based the high marginal rating on the SUS scored.

**ACKNOWLEDGEMENTS**

This study was sponsored by the Research University Grant (Vote No. Q.J130000.2501.09H09), Universiti Teknologi Malaysia.

**REFERENCES**

[1] Bangor, A., Kortum, P. T., & Miller, J. T. 2008. An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction, 24(6), 574-594.

[2] Bolt, R. A. 1980. "Put-that-there": Voice and gesture at the graphics interface. Paper presented at the Proceedings of the 7th annual conference on Computer graphics and interactive techniques, Seattle, Washington, USA.

[3] Cohen, P., McGee, D., & Clow, J. 2000. The efficiency of multimodal interaction for a map-based task. Paper presented at the Proceedings of the sixth conference on Applied natural language processing, Seattle, Washington.

[4] Doyle, J., Bertolotto, M., & Wilson, D. 2008. A Survey of Multimodal Interfaces for Mobile Mapping Applications. In L. Meng, A. Zipf & S. Winter (Eds.), Map-based Mobile Services (pp. 146-167): Springer Berlin Heidelberg.

[5] Fang, Y., Chai, X., Xu, L., & Wang, K. 2009. Hand tracking and application in map navigation. Paper presented at the Proceedings of the First International Conference on Internet Multimedia Computing and Service, Kunming, Yunnan, China.

[6] Grandhi, S. A., Joue, G., & Mittelberg, I. 2011. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[7] Hansen, T. E., Hourcade, J. P., Virbel, M., Patali, S., & Serra, T. 2009. PyMT: a post-WIMP multi-touch user interface toolkit. Paper presented at the Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces.

[8] Jacobson, R. D., & Sam, K. 2006. Multimodal Web-GIS: Augmenting Map Navigation and Spatial Data Visualization with Voice Control. AutoCarto 2006.

[9] Johnston, M., & Bangalore, S. 2004. MATCHKiosk: a multimodal interactive city guide. Paper presented at the Proceedings of the ACL 2004 on Interactive poster and demonstration sessions, Barcelona, Spain.

[10] Jokinen, K. 2008. User Interaction in Mobile Navigation Applications. In L. Meng, A. Zipf & S. Winter (Eds.), Map-based Mobile Services (pp. 168-197): Springer Berlin Heidelberg.

[11] Kamel Boulos, M., Blanchard, B., Walker, C., Montero, J., Tripathy, A., & Gutierrez-Osuna, R. 2011. Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation. International Journal of Health Geographics, 10(1), 45.

[12] Karpov, A., Carbini, S., Ronzhin, A., & Viallet, J. E. 2008. Comparison of two different similar speech and gestures multimodal interfaces. Paper presented at the Signal Processing Conference, 2008 16th European.

[13] Karray, F., Alemzadeh, M., Saleh, J. A., & Arab, M. N. 2008. Human-Computer Interaction: Overview on State of the Art. International Journal on Smart Sensing and Intelligent Systems, 1(1), 137-159.

[14] Kim, S.-G., Kim, J.-W., Bae, K.-T., & Lee, C.-W. 2006. Multi-touch Interaction for Table-Top Display. In Z. Pan, A. Cheok, M. Haller, R. H. Lau, H. Saito & R. Liang (Eds.), Advances in Artificial Reality and Tele-Existence (Vol. 4282, pp. 1273-1282): Springer Berlin Heidelberg.

[15] Krahnstoever, N., Kettebekov, S., Yeasin, M., & Sharma, R. 2002. A real-time framework for natural multimodal interaction with large screen displays. Paper presented at the Multimodal Interfaces, 2002.

[16] Krum, D. M., Omoteso, O., Ribarsky, W., Starner, T., & Hodges, L. F. 2002. Speech and gesture multimodal control of a whole Earth 3D visualization environment. Paper presented at the Proceedings of the symposium on Data Visualisation 2002, Barcelona, Spain.

[17] Leap API: Hand class. 2015. Retrieved 11 August, 2015, from https://developer.leapmotion.com/documentation/csharp/api/Leap.Hand.html.

[18] Leap API: Pointable class. 2015. Retrieved 11 August, 2015, from https://developer.leapmotion.com/documentation/csharp/api/Leap.Pointable.html.

[19] Leap API: SwipeGesture class. 2015. Retrieved 11 August, 2015, from https://developer.leapmotion.com/documentation/csharp/api/Leap.SwipeGesture.html.

[20] Leap Motion API Overview. 2015. Retrieved 10 August 2015, from

https://developer.leapmotion.com/documentation/csharp/devguide/Leap_Overview.html.

[21] McClymont, J., Shuralyov, D., & Stuerzlinger, W. 2011. Comparison of 3D navigation interfaces. Paper presented at the Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2011.

[22] Nakatsu, R., Nicholson, J., & Tosa, N. 1999. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. Paper presented at the Proceedings of the seventh ACM international conference on Multimedia (Part 1), Orlando, Florida, USA.

[23] Neale, D. C., & Carroll, J. M. 1997. The role of metaphors in user interface design. Handbook of human-computer interaction, 2, 441-462.

[24] Oviatt, S. 1997. Multimodal interactive maps: designing for human performance. Hum.-Comput. Interact., 12(1), 93-129.

[25] Oviatt, S. 2000. Multimodal system processing in mobile environments. Paper presented at the Proceedings of the 13th annual ACM symposium on User interface software and technology.

[26] Oviatt, S. 2003. Multimodal interfaces. In A. J. Julie & S. Andrew (Eds.), The human-computer interaction handbook (pp. 286-304): L. Erlbaum Associates Inc.

[27] Pang, Y. Y., & Ismail, N. A. 2012. A survey of hand gesture dialogue modeling for map navigation. Computer Engineering and Applications Journal, 1(2), 57-62.

[28] Pang, Y. Y., & Ismail, N. A. 2015. Users' preferences for map navigation gestures. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(1), 77 - 83.

[29] Pang, Y. Y., Ismail, N. A., & Gilbert, P. L. S. 2010. A Real Time Vision-Based Hand Gesture Interaction. Paper presented at the Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International, Kota Kinabalu, Malaysia.

[30] Pavlovic, V., Sharma, R., & Huang, T. S. 1997. Visual interpretation of hand gestures for human-computer interaction: A review. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7), 677-695.

[31] Tzovaras, D. 2008. Multimodal User Interfaces: From Signals to Interaction: Springer Science & Business Media.

[32] Villaroman, N., Rowe, D., & Swan, B. 2011. Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor. Paper presented at the Proceedings of the 2011 conference on Information technology education.

[33] Wilson, A. D. 2004. TouchLight: an imaging touch screen and display for gesture-based interaction. Paper presented at the Proceedings of the 6th international conference on Multimodal interfaces.