International Journal on Advanced Science Engineering Information Technology

Vol.8 (2018) No. 4-2
ISSN: 2088-5334

A Review of Multimodal Interaction Technique in Augmented Reality Environment

Siti Soleha Muhammad Nizam[#,1], Rimaniza Zainal Abidin[#,2], Nurhazarifah Che Hashim[#,3], Meng Chun Lam[#,4], Haslina Arshad[#,5], Nazatul Aini Abd Majid[#,6]

[#]*Mixed Reality and Pervasive Computing Lab, Centre of Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
E-mail: [1]sitisoleha1610@gmail.com, [2]nizazainal90@gmail.com, [3]nurhazarifahchehashim@yahoo.com, [4]lammc@ukm.edu.my, [5]haslinarshad@ukm.edu.my, [6]nazatulaini@ukm.edu.my

Abstract— **Augmented Reality (AR) has proposed several types of interaction techniques such as 3D interactions, natural interactions, tangible interactions, spatial awareness interactions and multimodal interactions. Usually, interaction technique in AR involve unimodal interaction technique that only allows user to interact with AR content by using one modality such as gesture, speech, click, etc. Meanwhile, the combination of more than one modality is called multimodal. Multimodal can contribute to human and computer interaction more efficient and will enhance better user experience. This is because, there are a lot of issues have been found when user use unimodal interaction technique in AR environment such as fat fingers. Recent research has shown that multimodal interface (MMI) has been explored in AR environment and has been applied in various domain. This paper presents an empirical study of some of the key aspects and issues in multimodal interaction augmented reality, touching on the interaction technique and system framework. We reviewed the question of what are the interaction techniques that have been used to perform a multimodal interaction in AR environment and what are the integrated components applied in multimodal interaction AR frameworks. These two questions were used to be analysed in order to find the trends in multimodal field as a main contribution of this paper. We found that gesture, speech and touch are frequently used to manipulate virtual object. Most of the integrated component in MMI AR framework discussed only on the concept of the framework components or the information centred design between the components. Finally, we conclude this paper by providing ideas for future work involving this field.**
Keywords— **review; multimodal interaction technique; augmented reality**

I. INTRODUCTION

Augmented Reality (AR) has become one of the emerging technologies and gaining attention among society globally in line with the importance of technology in today's daily life. AR is generally defined as a technology that incorporates 2D or 3D virtual objects into three dimensional real environments [1]. This definition has been provided through a visualization of the Reality-Virtuality Continuum as in Fig. 1 where AR is a general idea of Mixed Reality. AR has three main features which are (1) a combination of virtual and real-world elements, (2) drawn in real-time interactively and (3) registered in 3D environment [2].

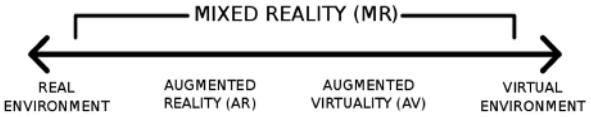

Fig. 1 Reality-Virtuality Continuum [2]

There are several interaction techniques in AR that are often discussed. Interaction techniques in the AR environment focused on how users interact with virtual objects that appear in the AR environment [3]. There are several types of AR interaction techniques, such as 3D interactions, natural interactions, tangible interactions, spatial awareness interactions and multimodal interactions. An interface relies upon the number and assortment of information inputs and outputs, which are information or communication channels that enable users to associate with a computer. Every independent single channel is called modality. A system that is based on only one modality is called unimodal [4]. There are commonly three categories involving unimodal inputs such as visual-based, sensor-based and audio-based. These three modalities inputs are often associated in a unimodal environment. Interaction based on the behaviour of human computer probably is the most common area in Human Computer Interaction (HCI) research [5]. Given the scope of the applications and the various problems and approaches, researchers tried to address the different aspects of human reactions and ability

[6] that could be perceived as a visual signal. Examples of visual-based modality include facial expression analysis (emotion recognition), body movement tracking, gesture recognition and gaze detection (eyes movement tracking). Audio input is also often used in the HCI environment as it is one of the important interactions to convey information [7]. Among the inputs involving audio-based interaction are speech recognition, speaker recognition, auditory emotion analysis and musical interaction. For sensor-based interactions, inputs include pen-based, keyboard, mouse, joystick, touch, motion tracking sensor and digitizer, haptic sensor, pressure sensor and taste / smell sensor. Touch interaction technique includes single and multi-touch interaction, as most of touch screen devices allow user to interact with more than one touch input [8]. This paper discusses the multimodal interaction in AR environment. The combination of more than one modality is called multimodal [9]. Different with unimodal, multimodal requires a combination of multiple architecture to combine two or more input orders, naturally and efficiently. Recent researches have shown that multimodal interaction (MMI) allows very natural interaction by letting a person to use two or more input channels at the same time especially in augmented reality and virtual reality [10] environment . For example, combining speech input with pen gestures creates an intuitive command and control application. In AR environments, multimodal is considered a solution to enhance interaction between physical and virtual entities. Besides, AR supports interaction in real world and virtual world at the same time. Hence, multimodal interaction is an ideal interaction technique for AR applications. This study was conducted to answer 2 main questions: (1) What are the interaction techniques, types of AR application and domain for MMI AR and (2) What are the components integrated into MMI AR model/framework.

This paper was organized as follows: In section 2, documents selection and result statistic are discussed as material of this paper. Review in multimodal AR is also discussed and explored in Section 2. Next, Section 3 discussed the findings of this paper. Finally, conclusions and future work are presented in Section 4. This review paper will give an overview about multimodal interaction that contributes in recent AR technology.

II. The Material and Method

To begin the search, queries were multimodal interaction in augmented reality. Based on Fig. 2, the initial result of query search contains 72 documents from SCOPUS database. Then, the result was filtered by the document type, field, language, and year. The filtration ended up with 32 documents. After title and abstract filtering, 32 documents have reduced to 21. Then, full text reading process excluded 7 documents, resulting in 14 documents as the final document to be analysed.

From SCOPUS database, 14 articles related to multimodal interaction technique in AR have been selected and reviewed. All the papers were published from 2015 until 2018. Table 1 shows information that has been extracted from the reviewed paper such as title, interaction technique, description, framework/model, type of AR and domain. This table will be further discussed by answer following research questions:

Q1: What are the interaction techniques, types of AR application and domain for MMI AR?

Q2: What are the components integrated into MMI AR model/framework?

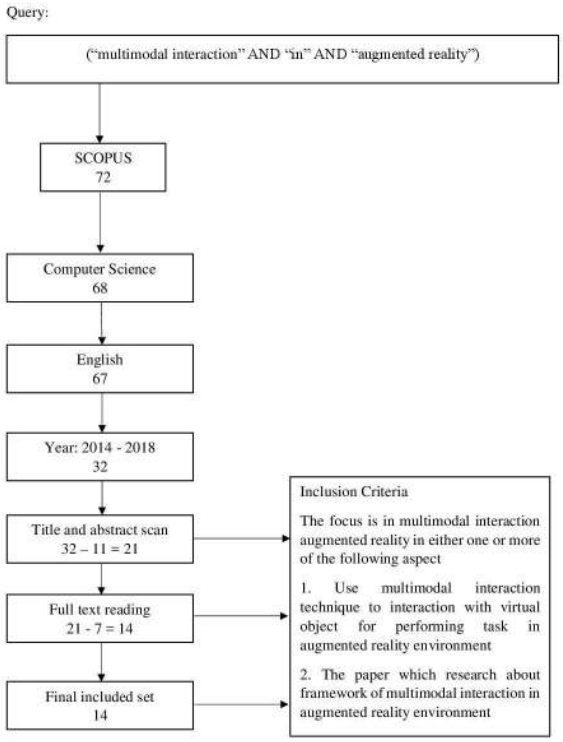

Fig. 2 Document Selection for Review Process

The authors answered the questions separately in the following subsections.

A. *Answer to Q1: What are the interaction techniques, types of AR application and domain for MMI AR.*

To describe the input modality used in previous AR environment. The 14 papers are summarised based on the following characteristics: interaction technique, type of AR systems and application domain.

1) Interaction Technique

Interaction technique plays an important role in AR environment which offers the control of the virtual objects, including selection and manipulation functions of the virtual object such as color, shape and position [11]. In AR, multimodal interaction is one of the interaction types which combines two or more input modalities [12].

The use of multimodal interaction for AR environment has recently become widespread as can been seen in the works. For instance, three main interaction modalities focused by [13] included gesture recognition, facial recognition and speech recognition. Gesture recognition allowed user to issue commands which may be as simple as a 3D pointer or as complex as the virtual copy of the hand itself. Extracting only the desired 3D hand motion data such as fingertip orientations, finger positions and global hand pose could provide simple interfaces with real-time operation speed. The face recognition process for systems has three stages which includes detection, recognition and tracking of face. Speech recognition focused on how user

interacts with the systems using voice commands. There were two interactions involved in the web application in [14] which included touch that allowed users to click any interactive components using mouse and gesture which help user to move the furniture/object by pointing, grabbing and releasing the object. The applications were targeted to solve two problems in choosing furniture online which were size and colour. Multimodal fusion was used to combine speech and gesture interaction technique [15]. They found that there are three levels of fusion which are feature, intermediate and hybrid.

An AR game application developed by [16] tried to explore about two input modalities, the face recognition and touch. The AR application required users to shoot the zombie on the head to earn points. Face recognition included several components which were face detection, face normalization, face recognition and 3D elements. The virtual zombie will be displayed based on the matching of detected human face and the specified set of zombie data. The fusion engine will identify whether the human being detected is a zombie or not by using face recognition. Touch input was the interaction used to fire a gun to attack the zombie that appear on screen. Next, a system developed by [17] was to complete daily task through Head Mounted Display (HMD). Two types of interaction involved which were the adaptive interaction and touch. [18] had define adaptive interfaces as a system that adapts its display with current user needs and the ability to monitor user's task, system's tasks and current situation. For mobile adaptive interface, the device adapts behaviour based on the variation of the interaction context such as user, environment and device itself. User-driven adaptation approach has been used to enable users to flexibly access their work from different devices then map the retrieved inputs to interact with HMD. For example, when user is on mobile, they can easily use the input from other device and interaction event to support mobility. The touch interaction functioned when user click on the screen.

An interactive system that used mobile device to capture indoor scene of CAD-like 3D models was proposed by [19]. They used speech and touch interaction to perform the interaction with the system. By using touch interaction, a sketching paradigm was used to determine the scale, position, type and orientation of a 3D furniture model (e.g. beds, dressers, tables, chairs etc) that the user wishes to place. In order to correctly align with the live view, the 3D model will be automatically oriented, scaled and positioned in the 3D room model. Besides, voice commands can optionally issue by the user at any point. Voice commands allow for easy and unambiguous selection of objects from the database. They have found that the retrieval problem can be dramatically simplifies by using voice recognition and enables more performant than sketch-based retrieval.

[20] used the combination of gestures speech interaction in a real-time interactive system domain. The role of the gesture is to point and grab the virtual object and place it to workspace. For speech interaction, the processing of the first speech token was required to check if there is an executable action that is associated with the semantic concept grab. For example, the adjective "green" is identified to be an instance of the semantic type color and "grab" is detected to denote the gesture interaction in their system. Types of input modalities applied in Location-based Augmented Reality (LBAR) system have been reviewed in [21]. They are combining multimodal interaction (user natural interaction) and adaptive interaction (current environment and device state data) to perform as an input to LBAR system. They have discussed several factors such as mobility factor, mobile context factor and user preference factor which determine which modalities to be considered appropriate for LBAR system. For multimodal interaction input, they suggest speech, motion and touch as explicit interaction. Meanwhile, state of device and environment factor are the suitable implicit interaction which can be applied in LBAR.

An AR game application has been developed in [22] purposed to explore interaction modalities in AR environment. This prototype allowed user to interact with virtual dog by using two common interaction techniques which were voice input and gesture. This application is a mobile based application where leap motion was used as standalone depth-sensing camera to detect gesture interaction input while Google Cloud Speech API was used to enable speech interaction modality. Symbolic gesture was used in the application because the virtual dog designed to respond to the symbolic gestures from users. The virtual dog will perform corresponding actions based on the gesture and speech orders from user. This scenario is the same as how people interact with their pets in real life. Based on the result from the experiment setup of this application, it shows that the combination of speech and gesture in this application enhanced user experience. A system that allowed user to control a robot manipulator by interacting with 3D model in mixed reality environment has proposed in [23]. This system combined tangible and gesture to interact with 3D model. Tangible interaction technique needs physical object to interact with digital object. In this system, physical object that been used is a robot. User can control a real robot to manipulate digital cube that serves as a target for virtual robot. Meanwhile, gesture interaction technique used as a movement (jogging) commands to the virtual robot. An application that provides digital information to help tourist has designed in [24]. User can use gesture, motion and touch interaction technique to interact with the digital information. For instance, for gesture and motion interaction technique, user can point at POI to select the place. For gesture interaction technique, user can select the POI by pointing it using their finger. Meanwhile, for motion interaction technique, user can select the POI by moving their device at the POI to select them. Furthermore, this application also provides touch interaction technique as user can interact with the augmented digital information through the phone screen.

A prototype that explains the physiological structure of the human body has developed by [25]. User can explore the bones, nerve, muscles and vein structures of human body parts in x-ray illustration. This system used tangible, gesture and speech to interact with virtual contents. Physical cube was used in this system as the physical object to interact with digital object. Manipulation functions that have provided with tangible interaction techniques are move, rotate and pick menu. User can move and rotate the virtual human body by performing the exact same action with the physical cube.

TABLE I
REVIEW OF MULTIMODAL INTERACTION TECHNIQUE IN AUGMENTED REALITY ENVIRONMENT

Name (Year)	Interaction technique	Description	Framework/ Model	Type of AR	Domain
Vision-Based Technique and Issues for Multimodal Interaction in AR (2015) [13]	Face, Gesture, Speech	Discussion about related issues on multimodal interaction in AR. The paper was concluded with the future direction for multimodal interaction in AR. The multimodal interaction was discussed based on a few topics which include input and output modalities, multimodal fusion that will integrate all the interaction involves and lastly, discussion on multimodal in AR. Three main interactions focused in this paper includes Gesture Recognition, Facial Recognition and Speech Recognition.	✓	See-through based, Mobile based	Education, Entertainment, Medical, Art.
Model-based Design of Multimodal Interaction for AR Web Applications (2015) [14]	Click, Gesture	Web system was developed that allowed customers to buy furniture online based on the criteria they want. Click: Clicking interaction using mouse Gesture: Moving the furniture/object. Pointing the object Grab and release the object	✓	PC-based	Architecture
Multimodal Fusion: Gesture and Speech Input in AR Environment (2015) [15]	Speech, Gesture	The paper discussed about previous work on multimodal input in AR. The guideline about multimodal fusion was discussed in this paper. Explanation about multimodal fusion level in AR: First fusion: Feature level Second fusion: Intermediate decision fusion Third fusion: Hybrid fusion (mixture of two modalities).	✕	✕	Education, Entertainment, Medical, Art, Business, Architecture.
ARZombie: A Mobile Augmented Reality Game with Multimodal Interaction (2015) [16]	Face recognition, Touch	This paper focused on the development of AR game that integrates multimodal interaction to enhance better gaming experience where virtual zombie will be display on screen based on the detected face (human) recognized on specific class of the zombie. Face recognition: The game engine will identify whether the human detected is a zombie or not by using face recognition. Touch: The interaction uses to attack the zombie that appear on screen.	✕	Mobile based	Game
Input Forager: A User-Driven Interaction Adaptation Approach for Head Worn Displays (2016) [17]	Adaptive, Touch	The system was developed to help facilitate the work and completing daily task through HWD. Adaptive: Borrowing embedded inputs from mobile and wearable devices Allocation/mapping of AR interaction-events to borrowed input methods. Touch: Clicking interaction on the screen.	✕	See-through based, Mobile based	Business

<table>
<tr>
<td>In Situ CAD Capture (2016) [19]</td>
<td>Speech, Touch</td>
<td>A mobile system for placing virtual 3D model of furniture in the scene by using 2 types of interaction:
Speech: user can optionally issue voice commands at any point by tapping the listen button.
Touch: User draws 2D line to get the desired furniture in database and display on the screen.</td>
<td>✕</td>
<td>Mobile based</td>
<td>Architecture</td>
</tr>
<tr>
<td>Semantics-based Software Techniques for Maintainable Multimodal Input Processing in Real-time Interactive Systems (2016) [20]</td>
<td>Speech, Gesture</td>
<td>In this paper, they use combination of gestures and speech interaction for real-time interactive system
1. Gesture: to point and grab the virtual object
2. Speech: to check if there is an executable action that is associated with semantic concept grab. For example, “green” is identified to be an instance of the semantic type Color and “grab” is detected to denote the gesture interaction in their system. the adjective</td>
<td>✕</td>
<td>Mobile-based</td>
<td>✕</td>
</tr>
<tr>
<td>A Framework of Adaptive Multimodal Input for Location-Based Augmented Reality Application (2017) [21]</td>
<td>Speech, Motion, Touch, Adaptive</td>
<td>A Location-based Augmented Reality application which combines multimodal and adaptive interaction.
Multimodal Interaction Input:
Speech – Provide input query for specific location
Motion – To point the device at specific direction and view the POI
Touch – View location information

Adaptive Interaction Input:
Device: the state of user’s device
Environment: the environment factor which will affect the number of displayed POI.</td>
<td>✓</td>
<td>Mobile based</td>
<td>Tourism</td>
</tr>
<tr>
<td>Multimodal Interaction in Augmented Reality (2017) [22]</td>
<td>Speech, Gesture</td>
<td>An AR game where user can interact with virtual dog by using gesture and speech.

<table>
<tr><th>Tasks</th><th>Speech Orders</th><th>Gesture Orders</th></tr>
<tr><td>Make the dog stand up</td><td>“Sit down”</td><td>Push down hand</td></tr>
<tr><td>Make the dog sit</td><td>“Stand-up”</td><td>Pull up hand</td></tr>
<tr><td>Make the dog bark</td><td>“Bark”</td><td>Draw a circle</td></tr>
<tr><td>Make the dog stop barking</td><td>“Stop”</td><td>Re-draw a circle</td></tr>
</table>
</td>
<td>✕</td>
<td>Mobile based</td>
<td>Game</td>
</tr>
<tr>
<td>Towards Multimodal Interactions: Robot Jogging in Mixed Reality (2017) [23]</td>
<td>Tangible, Gesture</td>
<td>A system that allows user to control a robot manipulator by interacting with 3D model in mixed reality environment.

Tangible Interaction technique (real robot):
Move a digital cube. The cube is a target for virtual robot.
Gesture interaction technique:
Move (jog) the virtual robot.</td>
<td>✓</td>
<td>See-through based</td>
<td>Robotic</td>
</tr>
</table>

Combining Intelligent Recommendation and Mixed Reality in Itineraries for Urban Exploration (2017) [24]	Gesture, Motion, Touch	An application that will help tourist to get information in AR environment. Gesture interaction technique: Pointing: Select POI. Touch interaction technique: Click button: Confirm the selection. Motion interaction technique: Move device to POI.	✕	Mobile based	Tourism
Mobile AR Illustrations that entertain and inform: Design and Implementation issues with the Hololens (2017) [25]	Speech, Gesture, Tangible	3D-Human on a Box is a prototype that explains the physiological structure of the human body. Tangible interaction technique (Cubic): Move the virtual human body. Rotate the virtual human body. Pick menu to display information. Speech interaction technique: Voice command: Open and close menus for further interactions. Gesture interaction technique: Tap virtual body: Open menus for further interactions. Tap virtual body: Close menus for further interactions.	✕	See-through based	Education
Let's Cook: An AR System Towards Developing Cooking Skills for Children with Cognitive Impairments (2018) [26]	Tangible, Click	Let's Cook is an AR serious game to educate cognitive impairments children to prepare simple meals. Tangible interaction technique (Card): Pick recipe material or utensil Click interaction technique (Pointing device): Combine recipe material. Setting appliance. Plug or unplug device. Turn on and off device.	✕	Projection based	Game
AR in Maintenance: An information centred design framework (2018) [27]	Gesture and 2D and 3D visualisation	AR in maintenance application has proposed several types of interaction to analyse maintainer's performance. 3D and 2D Visualisation: 3D colour model was visualized to detect the equipment's problem during diagnose process and 3D to guide maintainers to repair the equipment. 2D visualisation: Text was visualized to present the data of maintainers performance during the analyse process. Gesture: Gestures will be tracked to assess maintenance's performance.	✓	Mobile based	Maintenance

Speech and gesture interaction techniques were used to trigger menus that can be chosen using a physical cube. The user can get further information of the virtual human body parts that labelled accordingly by choosing the augmented menu. Both speech and gesture interaction techniques can be performed by the user to close the menu. A game that was designed to teach cognitive impairments children to prepare simple meals by following step-by-step approach in AR environment has been shown in the research by [26]. This system supported multimodal interaction techniques by combining tangible and click interaction technique to interact with virtual content. Physical object which is a card was used in this system to interact as tangible interaction technique. Student can pick recipe material or utensils by placing the card on the table. This system used pointing device as click interaction technique to allow student to click on the menu. An AR in maintenance application was proposed in [27]. Several types of interactions were suggested to analyse maintainer's performance. 2D and 3D visualisation were used based on the sensor data of the equipment tracked to help maintainers in diagnosing and repairing fault. For analysing maintainers skill performance, the system tracks the maintainer's gesture during the repairing process.

2) Type of AR System

Fig. 3 represents the type of AR mentioned in the 14 reviewed papers for this research. The trend of the type of AR that has been used from 2015 to 2018 is presented in this chart. Based on the figure, it clearly shows that mobile-based system is the trendiest type of AR with 57% that used by the researchers in this field. Followed by see-through based with 29% and both projection-based and PC-based have 7%. To summarise the usage of the types of AR system by the researcher, see-through based system were using HMD [17], [23], [25] and Google glass [13] to visualize the virtual object. Mobile-based system are used by [13], [17], [20]–[22], [27] and tablet [16], [19], [24] as a visualization device. For projection-based system, [26] used table top to display the virtual object and PC-based was used in [14].

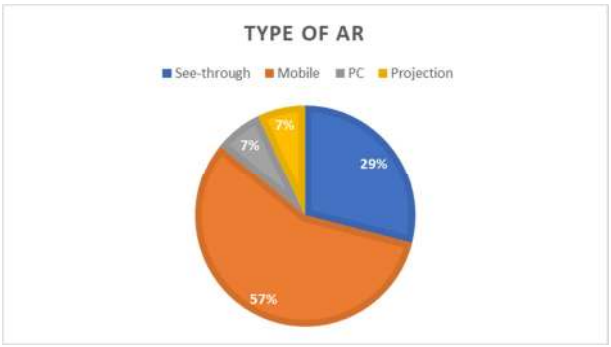

Fig. 3 Type of AR system used in MMI AR

3) Application Domain

Fig. 4 shows the application domain that has applied MMI in AR from 2015 until 2018. The most popular domains that have been explored are education [13], [15], [25], architecture [13], [15] and games [16], [22], [26]. Furthermore, application domain was averagely has been studied in MMI AR were entertainment, medical, art [13], [15], business [15], [17] and tourism [21], [24]. Meanwhile, maintenance [27] and robotic [23] are the least domain that have applied MMI in AR.

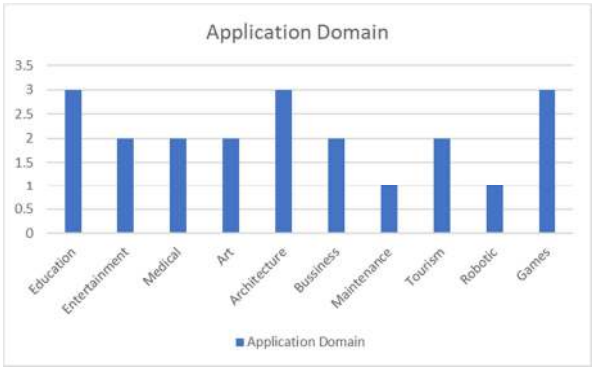

Fig. 4 Application Domain Applied in MMI AR

B. Answer to Q2: What are the components integrated into MMI AR model/framework?

Fig. 5 shows the behaviour model of the gesture-based interaction resource (IR) used to control the furniture shop [14]. The model consists of three main components which are (a) IR hand gesture that explaining the details behaviour description of one and two-handed gesture (b) Structure gesture that shows the concepts of gesture interaction and (c) the example of static gesture of the controlling hand. The model allows users to flexibly interact using one or two hands. One hand is for pointing and the other hand is used for controlling purpose using different gesture (posture: previous, next, release and select).

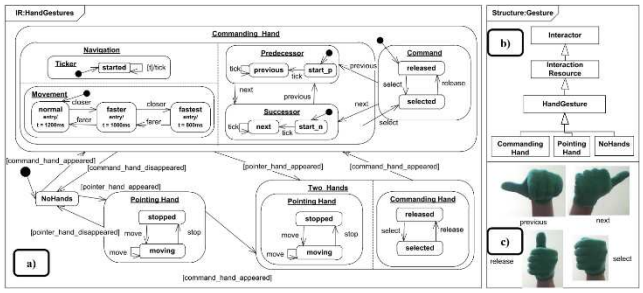

Fig. 5 The Behaviour Model of Gesture Interaction Use to Control Virtual Furniture [14]

Fig. 6 is a proposed framework that has produced by combining components from previous proposed frameworks in the field of AR, multimodal interface and adaptive interface [21]. This framework presents the concept of how input modalities and adaptive information will take place to serve location-based augmented reality application. User input modalities was categorized as explicit modalities where user uses several modalities such as motion gesture, speech and touch to interact with the system. While changes of environmental and mobile device were categorized as implicit input interaction. The data will be gathered from environment (e.g. day, night, level of temperature, noise level) and device state (e.g. battery level, time). All implicit and explicit interaction will be recognized by modalities recognizer and processor. Adaptive multimodal fusion and output fission module will

process and interpret the modalities by using a specific fusion technique and synthesize the data. In the AR module, the system will manage the AR view after the AR view controller and manager was initiated based on synthesized adaptive multimodal input.

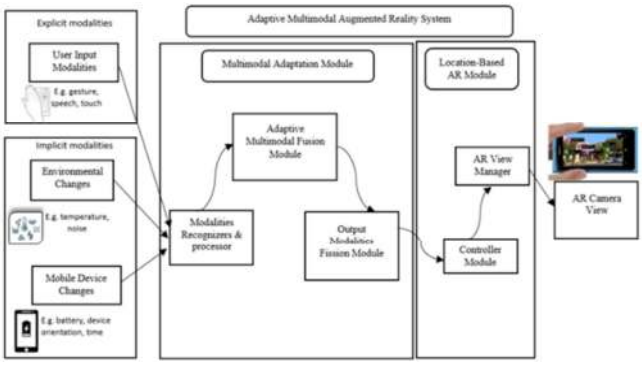

Fig. 6 Conceptual framework of adaptive multimodal interface for mobile AR [21]

The framework in Fig. 7 allows the user to interact with mixed reality 3D models that has been displayed with HMD to control a robot manipulator [23]. This framework consists of three main components which are Robot Operating System (ROS), Unity 3D and Robot. ROS and Unity 3D are connected by Rosbridge (WebSocket). The interaction in mixed reality environment were handled by Unity 3D. Robot will encode values through UDP communication in real time. The Robot will interpret the space coordinates in the HMD that has been published to ROS as topics and services.

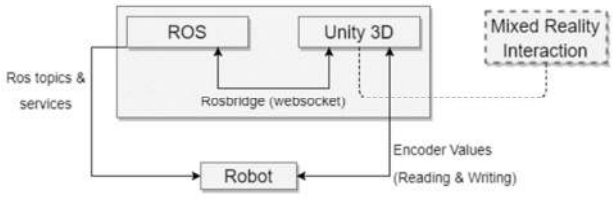

Fig. 7 Framework of the system [23]

In [27], they proposed an information framework for AR in maintenance. Fig 8 illustrates the framework that describes the relationship between maintenance information systems and maintenance environment. There are also the components of maintenance processes that amplify those processes using interaction abilities and AR visualisation. This is driven by the data needed to reach the process such as AR capabilities, information formats, and environmental data (respectively, dark blue, green and orange boxes). There are three maintenance processes included in the framework which are diagnosis, repair and analysis. For diagnosing, the framework can help maintainers to diagnosing faulty by using 3D coloured models based on the sensor data of the tracked equipment. Then, by using the same data can be used to check whether the diagnosis was performed correctly. For repairing, the steps of repairing can be easily explained by using 3D animations. In the case of analysing, the maintainer performance will be measured by tracking their gestures during the repairing process. Maintenance applications in AR environment with different levels of fascination and interaction between real and virtual world can be developed using this framework.

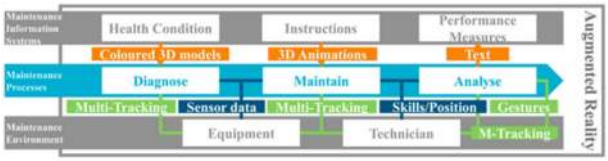

Fig. 8 Information Centred Design Framework [27]

III. Results and Discussion

1) *Interaction Technique*

Fig. 9 shows the statistic of modalities used in MMI AR based on selected paper that had been analysed. The modality that gets the most attention is gesture interaction technique. Meanwhile, speech interaction technique is the second popular modality used in MMI AR; Followed by other interaction technique which are face, touch, adaptive, motion, tangible and click. This is because, gesture interaction technique is one of the natural interaction technique as user can interact to perform task by moving their body. Speech interaction technique is also one of the natural interaction technique where user use their voice to interact by giving command or order. Hence, use of speech recognition in AR environment will enhance the usability of this technology [7]. Touch interaction technique also gets a lot of attention due to demand of mobile phone nowadays. Most of the systems reviewed have used touch interaction technique via phone screen. This is due to the current design of mobile phone which has big display screen and do not have phone keypad [28]. Most of the gesture interaction technique that has been used in the selected papers were PC-based, mobile based and see-through system. The devices such as Leap Motion and Kinect are often used as tracking device which are more compatible with that type of AR system. For tangible interaction technique, researchers were more focused on see-through and projection-based because it will allow user to interact with free hand movement.

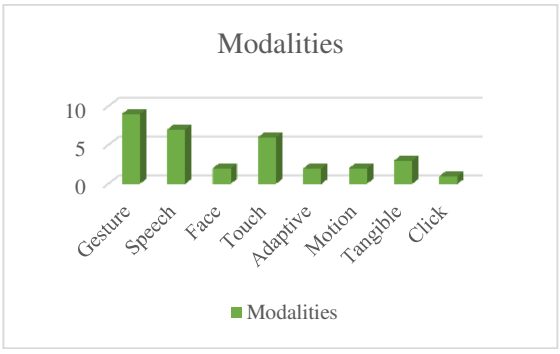

Fig. 9 Modalities used in MMI AR

2) *Integrated Component in MMI Framework*

The finding from the 14 reviewed papers, only 4 papers have proposed a framework of multimodal interaction in AR environment. We have analysed the selected paper and we have found that three of the papers discussed the overall

system's framework and only one paper discussed on input interaction in a specific manner. Two of the frameworks used multimodal natural interaction such as gesture and speech and another two papers both leveraging multimodal interaction and adaptive interaction for their system. Several insights of previous proposed framework are discussed in the following.

- How researchers were integrating the components in the framework are based on which type of AR they are focusing to perform a task. It is because mobile based, see-through based and PC based may have different fusion technique and modalities recognizer compared to mobile-based.
- From 2015 until 2018, the trend of the framework has changed were researchers started to focus on mobile based or see-through based application framework instead of PC-based application framework.
- Most of the reviewed frameworks were not discussing on how the interaction modalities have been fused in detail but only discussed on the concept of the framework components or the information centred design between the components.
- From the 4 reviewed frameworks, none of reviewed paper discussed the AR part in detail. For example, how AR component or module was adapted to the synthesized multimodal input.

IV. Conclusion

In multimodal AR technology, recent studies conducted mostly focused on a combination of interaction inputs which involves speech recognition and gestures. Multimodal inputs for AR are not so widely studied by researchers [29]. Adding variety of modality and communication channels can help improve accuracy and provide a better user experience [30].

In this paper, the reviews were done based on two research questions which are (1) What are the interaction techniques, types of AR application and domain for MMI AR and (2) What are the components integrated into MMI AR model/framework. For interaction technique, it can be concluded that most of the reviewed paper focusing on a combination of a few interaction techniques which are speech recognition, hand gesture and touch input. Other interaction techniques such as face, adaptive, motion, tangible and click are still rarely used in previous research. For type of AR, previous research mostly focused on mobile phone since mobile phone is the current trend that popular among the users. Multimodal interaction in AR also focused on a few domains which are education, architectures and games. The research for multimodal interaction in AR related to robotic still less studied by researchers even if it is the current trends nowadays. Only a few previous researches have been done related to framework or model that includes multimodal interaction in AR. However, most of the reviewed frameworks were not discussing on how the interaction modalities have been fused in detail.

Based on the review, further work is needed to improve existing techniques related to multimodal interaction in AR. New combination of interactions are needed for creating multimodal applications and interface especially in AR environment. Other than that, future research should be done for the type of AR related to wearable glasses since it will be an expected trend soon. Finally, the limitations related to the framework for multimodal interaction in AR, opens up the research to design and explore more so that application development can be thoroughly guided through a good framework.

Acknowledgment

This work is supported by UKM research grant, TD-2016-003.

References

[1] P. Milgram, F. K.-I. T. on I. and, and undefined 1994, "A taxonomy of mixed reality visual displays," *Search.Ieice.Org*, no. 12, pp. 1–15, 2003.

[2] R. T. Azuma, "A Survey of Augmented Reality," pp. 355–385, 1997.

[3] L. Ahmed, S. Hamdy, D. Hegazy, and T. El-Arif, "Interaction Techniques in Mobile Augmented Reality: State-of-the-art," *Int. Conf. Intell. Comput. Inf. Syst.*, pp. 424–433, 2015.

[4] K. Saroha, S. Sharma, and G. Bhatia, "Human Computer Interaction: An intellectual approach," *IJCSMS Int. J. Comput. Sci. Manag. Stud.*, vol. 11, no. 2, pp. 147–154, 2011.

[5] [5] P. Adkar, "Unimodal and Multimodal Human Computer Interaction:A Modern Overview," *Pratibha -International J. Comput. Sci. Inf. Engg*, vol. 2, pp. 2277–4408.

[6] R. S. Kraleva, "ChilDiBu – A Mobile Application for Bulgarian Children with Special Educational Needs," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 6, pp. 2085–2091, 2017.

[7] N. C. Hashim, N. Aini, A. Majid, H. Arshad, and W. K. Obeidy, "User satisfaction for an Augmented Reality application to support Productive Vocabulary using Speech Recognition."

[8] N. H. Hussain, T. S. M. Tengku Wook, S. F. Mat Noor, and H. Mohamed, "Children's interaction ability towards multi-touch gestures," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 875–881, 2016.

[9] S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko, "An evaluation of an augmented reality multimodal interface using speech and paddle gestures," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4282 LNCS, pp. 272–283, 2006.

[10] L. M. Chun, H. Arshad, T. Piumsomboon, and M. Billinghurst, "A combination of static and stroke gesture with speech for multimodal interaction in a virtual environment," *Proc. - 5th Int. Conf. Electr. Eng. Informatics Bridg. Knowl. between Acad. Ind. Community, ICEEI 2015*, pp. 59–64, 2015.

[11] [11] S. S. Muhammad Nizam, M. C. Lam, H. Arshad, and N. A. Suwadi, "A Scoping Review on Tangible and Spatial Awareness Interaction Technique in Mobile Augmented Reality-Authoring Tool in Kitchen," vol. 2018, 2018.

[12] S. Jamali, M. F. Shiratuddin, and K. Wong, "An Overview of mobile-Augmented Reality in Higher Education," *Int. J. Recent Trends Eng. Technol.*, vol. 11, no. 1, 2014.

[13] A. W. Ismail, M. Billinghurst, and M. S. Sunar, "Vision-Based Technique and Issues for Multimodal Interaction in Augmented Reality," *Proc. 8th Int. Symp. Vis. Inf. Commun. Interact.*, pp. 75–82, 2015.

[14] S. Feuerstack, Á. C. M. de Oliveira, M. dos Santos Anjo, R. B. Araujo, and E. B. Pizzolato, "Model-based design of multimodal interaction for augmented reality web applications," *Proc. 20th Int. Conf. 3D Web Technol. - Web3D '15*, pp. 259–267, 2015.

[15] A. W. Ismail and M. S. Sunar, "Multimodal Fusion: Gesture and Speech Input in Augmented Reality Environment," vol. 331, pp. 245–254, 2015.

[16] D. Cordeiro, R. Jesus, and N. Correia, "ARZombie: A Mobile Augmented Reality Game with Multimodal Interaction," *Proc. 7th Int. Conf. Intell. Technol. Interact. Entertain.*, vol. 17, 2015.

[17] M. Al-Sada, F. Ishizawa, J. Tsurukawa, and T. Nakajima, "Input forager: A user-driven interaction adaptation approach for head

worn displays," *ACM Int. Conf. Proceeding Ser.*, pp. 115–122, 2016.
[18] L. Rothrock, R. Koubek, F. Fuchs, M. Haas, and G. Salvendy, "Review and reappraisal of adaptive interfaces: Toward biologically inspired paradigms," *Theor. Issues Ergon. Sci.*, vol. 3, no. 1, pp. 47–84, 2002.
[19] A. Sankar and S. M. Seitz, "In Situ CAD Capture," *Proc. 18th Int. Conf. Human-Computer Interact. with Mob. Devices Serv.*, pp. 233–243, 2016.
[20] L. Fischbach, Martin; Dennis, Wiebush; Marc Erich, "Semantics-based Software Techniques for Maintainable Multimodal Input Processing in Real-time Interactive Systems," pp. 623–627, 2016.
[21] R. Z. Abidin, H. Arshad, and S. A. A. Shukri, "A framework of adaptive multimodal input for location-based augmented reality application," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 2–11, pp. 97–103, 2017.
[22] Z. Chen, J. Li, and Y. Hua, "Multimodal Interaction in Augmented Reality," pp. 206–209, 2017.
[23] E. Sita and M. Studley, "Towards Multimodal Interactions : Robot Jogging in Mixed Reality," pp. 2–3, 2017.
[24] G. Jacucci *et al.*, "Combining intelligent recommendation and mixed reality in itineraries for urban exploration," *Int. Ser. Inf. Syst. Manag. Creat. eMedia*, vol. 2017, no. 2, pp. 18–23, 2017.
[25] C. Zimmer, M. Bertram, F. Büntig, D. Drochtert, and C. Geiger, "Mobile augmented reality illustrations that entertain and inform with the hololens," *SIGGRAPH Asia 2017 Mob. Graph. Interact. Appl. - SA '17*, pp. 1–1, 2017.
[26] F. Pratesi, A. Monreale, F. Giannotti, and D. Pedreschi, "Let's Cook: An Augmented Reality System Towards Developing Cooking Skills for Children with Cognitive Impairments Eleni," vol. 2, pp. 142–152, 2018.
[27] I. Fernández Del Amo, J. A. Erkoyuncu, R. Roy, and S. Wilding, "Augmented Reality in Maintenance: An information-centred design framework," *Procedia Manuf.*, vol. 19, no. 2017, pp. 148–155, 2018.
[28] H. Arshad, S. A. Chowdhury, L. M. Chun, B. Parhizkar, and W. K. Obeidy, "A freeze-object interaction technique for handheld augmented reality systems," *Multimed. Tools Appl.*, vol. 75, no. 10, pp. 5819–5839, 2016.
[29] M. Billinghurst, "Hands and speech in space: multimodal interaction with augmented reality interfaces," *Proc. 15th ACM Int. Conf. multimodal Interact.*, no. Mmi, pp. 379–380, 2013.
[30] M. Turk, "Multimodal interaction: A review," *Pattern Recognit. Lett.*, vol. 36, no. 1, pp. 189–195, 2014.