

Integration of Augmented Reality and Voice Recognition in Learning English for Children

Dimas Wahyu Wibowo^{*1}, Ika Kusumaning Putri²

Politeknik Negeri Malang, Jalan Soekarno Hatta No. 9 Malang, (0341) 404424

*E-mail : dimas.w@polinema.ac.id ^{*1}, ikakputri@polinema.ac.id ²*

**Corresponding author*

Leni Saputri³

Politeknik Negeri Malang, Jalan Soekarno Hatta No. 9 Malang, (0341) 404424

E-mail : lenisaputri98@gmail.com ³

Abstract - Application development by combining two technologies, namely Augmented Reality and Voice Recognition, can make learning media regarding object recognition at home interact directly with 3D virtual objects. The technology can also help the pronunciation or pronunciation of sentences in English. Natural Language Processing (NLP) is used to understand human language so that machines can understand and process it. This ability supports Voice Recognition to have intelligence and interact like humans. wit.ai is an open-source NLP platform that can support speech-to-text application development. The merging of the two technologies in this development using the wit.ai platform. With the wit.ai platform that is used to understand voice commands and perform tasks as needed for applications regarding object recognition at home, users will be able to interact with objects at home through the given voice commands. In the Black Box testing, each functionality got the results that all the features had functioned properly. User Acceptance Test was also carried out and the average test results were 95.77% and 93.26% on a Likert scale with test results on 13 respondents aged 6-9 years old who have tried the application and 14 respondents as observers when 13 respondents aged 6-9 years tried the application. These results show that the application can be accepted and used as a tool in learning media.

Keywords - English, Learning Media, Augmented Reality, Voice Recognition

1. INTRODUCTION

Application development by combining two technologies, namely Augmented Reality and Voice Recognition, can make learning media regarding object recognition at home interact directly with 3D virtual objects. This application can also help the pronunciation of sentences in English. Augmented Reality (AR) is a combination of 2D or 3D digital objects / virtual objects that developed as close as possible to the original into the real world (real-time) [1]. Augmented Reality technology uses the environment in the real world by adding new information inside by using computers, webcams, smartphones, and special glasses such as Google Glass [2].

Voice recognition is a voice identification process based on spoken words by converting signals, which are captured by audio devices [3]. It can also recognize human commands and then translate them into data that can be processed by a computer, such as

operating a device, performing commands, or being able to write without typing via a keyboard [4]. Natural Language Processing (NLP) is used to understand human language so that machines can understand and process it. This ability supports Voice Recognition to have intelligence and interact like humans.

Google cloud platform, IBM Watson and wit.ai are NLP used to build speech-to-text applications that have different advantages. The merging of the two technologies in this development can be done using the wit.ai platform. Wit.ai is an NLP service to extract important structured information from a sentence [5]. In addition, wit.ai is also an open-source platform that can be used free of charge. Wit.ai allows developers to add a few lines of code to instantly build voice recognition and voice control into apps. Wit.ai will also check the user's pronunciation with the voice command that has been made. With the wit.ai platform to understand voice commands and perform tasks as needed for applications regarding object recognition at home, users can interact with objects at home through the given voice commands.

The application requires markers and speech from the user to be processed. It will detect the marker that will be tracked by the phone's camera and display the Augmented Reality content. When Augmented Reality content has been detected, the user provides input speech sentences. The application will detect the user's voice by using the API. Then the audio will be extracted into text for merging Augmented Reality content and Voice Recognition to produce output in the form of animation of Augmented Reality content. The result is the animation of Augmented Reality content and also text containing speech from the user. Figure 1 describes the application flow.

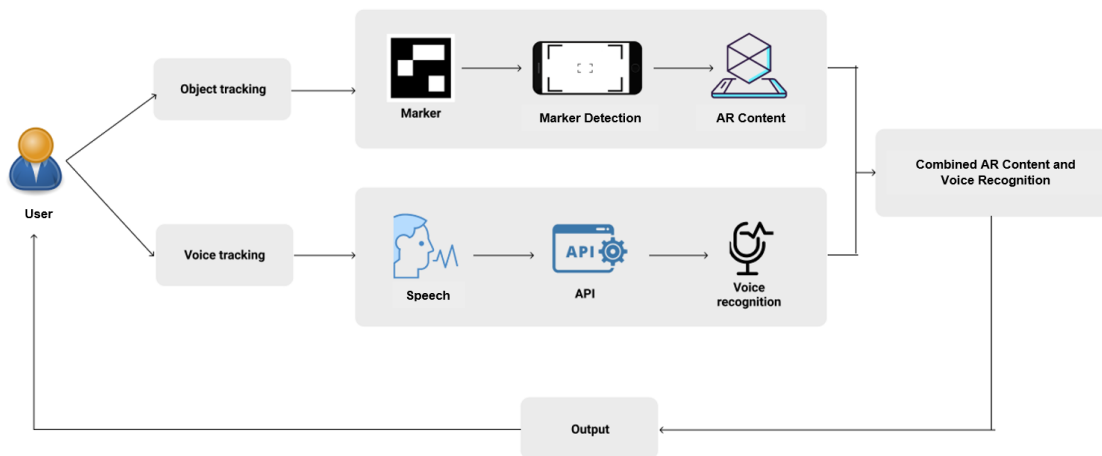


Figure 1. Application Flow

2. RESEARCH METHOD

The development methodology used in Augmented Reality and Voice Recognition with the wit.ai platform regarding object recognition at home use the Multimedia Development Life Cycle (MDLC) development method. MDLC has six stages: concept, design, collecting material, assembly, testing, and distribution. Figure 2 shows the MDLC methodology.

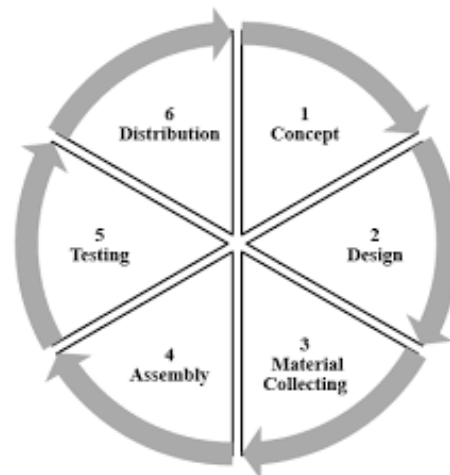


Figure 2. MDLC (Multimedia Development Life Cycle)

Requirement Analysis

Augmented Reality and Voice Recognition with the wit.ai platform regarding object recognition at home are applications that can make learning media interact directly with 3D virtual objects. Learning that is only in the form of an explanation or description orally makes the child only able to describe the explanation without seeing directly the shape or event described. With an android-based mobile application that applies Augmented Reality and Voice Recognition technology with the help of the wit.ai platform on learning media, children can interact directly with 3D virtual objects.


The process of combining Augmented Reality and Voice Recognition technology in the application to be made, namely, the application requires markers and speech from the user to be processed in the application. The application will detect the marker tracked by the phone's camera and display the Augmented Reality content. When Augmented Reality content in the form of 3D objects has appeared, the user provides speech input in the form of sentences that will later be provided. The application will detect the user's voice then the application will extract the audio and there will be a change from speech to text.

The use of wit.ai as an open-source platform that has been implemented in Augmented Reality (AR) can make 3D objects change when users pronounce sentences correctly in English. The application will display 3D object animation or color changes to 3D objects and text sentences if the 3D object changes as instructed. If it is wrong, the application still shows the 3D object with no animation of changing the color of the 3D object, and the text "Request unknown, please ask a different way" will appear.

Marker Implementation

The markers made are four markers that correspond to the AR-Voice play menu, namely bedroom markers, bathroom markers, livingroom markers, kitchen markers. Marker descriptions are shown in Table 1.

Table 1. Markers Description

No	Target Image	Description
1		Bedroom
2		Bathroom
3		Kitchen
4		Livingroom

Making commands for voice input is done using the wit.ai platform. By creating nine entities that will be used for instruction when inputting voice and to train bots: change, clock, close, color, drain, fill, off, on, open. Table 2 shows the entities used for voice commands.

Table 2. Entities for voice commands

Entities	Change	Clock	Close	Colour	Drain	Fill	Off	On	Open
Command	<ul style="list-style-type: none"> •Change the frying pan •Change the bowl •Change the plate •Change the glass •Change the sofa •Change the picture •Change the carpet 	<ul style="list-style-type: none"> •One o'Clock •Two o'Clock 	<ul style="list-style-type: none"> •Close the toothpaste •Close the toilet •Close the wardrobe •Close the desk •Close the drawer •Close the refrigerator 	<ul style="list-style-type: none"> •Change color to blue •Change color to yellow 	Drain the bathtub	Fill the bathtub	<ul style="list-style-type: none"> •Turn off the fan •Turn off the fan 	<ul style="list-style-type: none"> •Turn on the fan •Turn on the fan 	<ul style="list-style-type: none"> •Open the toothpaste •Open the toilet •Open the wardrobe •Open the desk •Open the drawer •Open the refrigerator

Each entity created in training uses utterances by creating utterances and sorting them into different entities to train them to convert speech into text.

System Implementation

The implementation of this system shows the process of combining Augmented Reality and Voice Recognition with the wit.ai platform for object recognition at home. Based on the results of the analysis and design, a Wit.ai platform is needed to combine Augmented Reality and Voice Recognition. Wit.ai is an NLP service to extract important structured information from a sentence [5]. The application uses markers to display 3D objects. After the 3D objects appear, the user provides voice input. Then the application will save the user's voice command and send it to the wit.ai platform using the API. After the voice input is has been received by the wit.ai platform, wit.ai will send back the result text if the voice input given by the user is correct. There will be changes to the 3D object displayed by the marker as shown in Figure 3.

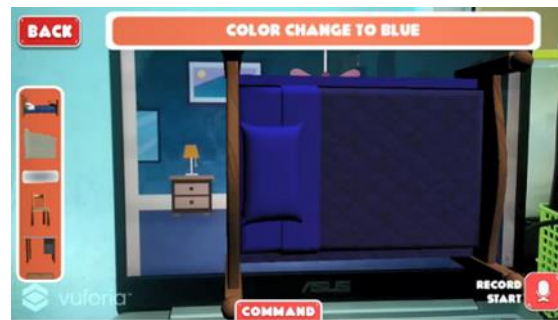


Figure 3. The example of 3D objects displayed by the marker

By adding entities and periodically training the bot in different ways, it can identify different sentences and use APIs to connect wit.ai to the app. Training each entity periodically with different text can produce a greater accuracy value. Merging in this application uses the API provided by wit.ai.

This API is used to connect the script so that it can retrieve data from wit.ai and check the voice input provided by the user and after checking, the results will be displayed. Actions that appear according to the called entities and conversion data from wit.ai then fetch the data and send commands back and interact with the user. The source code below is the source code for a 3D object that can change color. The 3D object that changes is the object detected by the marker. After checking, wit.ai sends the correct voice input and the 3D object change process will occur.

3. RESULTS AND DISCUSSION

Testing Augmented Reality and Voice Recognition applications with the wit.ai platform regarding object recognition at home using Black Box testing and UAT (User Acceptance Test) testing. This test aims to make the final result of the application implemented as needed.

Black box testing is used to determine whether the application is made by the requirements that have been designed. Testing is done by running the application and checking whether all the features in the application are running well. The black box testing process is more directed at the application functionality with the requirements that have been designed. Based on the tests, the results obtained from each functionality that all features have been functioning properly.

UAT (User Acceptance Test) testing is a testing process by users to produce documents that can be used as evidence that the application developed is acceptable or not by the user. The UAT (User Acceptance Test) test is carried out using a Likert scale. The Likert scale is used to measure a person's attitudes, opinions, and perceptions about social events or phenomena, in this study using a questionnaire to respondents to determine the level of assessment of the usefulness of the application that has been made (Waliyuddin et al., 2019). UAT is done by asking several questions to application users. Its users consist of elementary school-level children, their parents, and teachers. These children are the target users of this application while their parents or teachers provide answers based on their observations when children use the application.

The test uses the UAT as a measure of success in developing applications. Test results data on 13 respondents, children aged 6-9 years who have tried this application and 14 respondents with 1 teacher and 13 parents of students who observe when children play with the application, can be seen in Table 3 and Table 4.

Table 3. User Acceptance Test Results for children aged 6-9 years old

No	Aspects of Assessment for Students (Children aged 6-9 years)	Score	Percentage (%)
1	Is the appearance of this application attractive?	63	96.92%
2	Is the color display and appearance in this application good-looking and appropriate?	57	87.69%
3	Is this learning media easy to operate?	63	96.92%
4	Is the object recognition of objects at home in the application interesting?	62	95.38%
5	Are the information and material presented easy to understand?	63	96.92%
6	Can this learning media add insight into the introduction of objects at home in English?	64	98.46%
7	Is the menu easily accessible?	64	98.46%
8	Is the material menu easily accessible?	62	95.38%
9	Is the AR-Voice menu easily accessible?	52	80%
AVERAGE			95.77%

Table 4. User Acceptance Test Results for Teachers and Parents

No	Aspects of Assessment for Teachers and Parents	Score	Percentage (%)
1	Is the appearance of the object recognition application at home with AR and Voice Recognition attractive and easy to understand?	58	89.23%
2	Is the color display and appearance in this application good looking and appropriate?	56	86.15%
3	Are the objects at home displayed attractive and helpful in learning to recognize objects at home in English?	64	98.46%
4	Is the information and material presented complete?	61	93.84%
5	Does this learning media add insight into the introduction of objects at home in English?	63	96.92%
6	Can object recognition applications at home using AR and Voice Recognition attract students' interest in learning to recognize objects at home in English?	61	93.84%
7	Do you think this application is easy to use?	59	90.76%
8	Have you been helped by having an object recognition application at home using AR and Voice Recognition?	63	96.92%
9	Do you think this application is suitable for use as a learning medium?	63	96.92%
AVERAGE			93.26%

From the test data in Table 3 and Table 4 above using the UAT test, the total scores and percentages are obtained. The total score is obtained from the calculation results:

$$\text{Score} = \text{total (points per answer} \times \text{weight)} \quad (1)$$

To get the percentage value of each aspect of the test obtained by the calculation:

$$\text{Percentage} = \frac{(\text{Score} / \text{Number of Respondents})}{5} \times 100 \quad (2)$$

Based on the percentage result data from each aspect of the assessment, to get the final score, an average of the percentage results of each aspect of the assessment is carried out. The results of the UAT test carried out on 13 respondents who had finished trying this application and 14 respondents as observers for respondents who tried the application, obtained the results with an average interpretation of 95.77% and 93.26% on a Likert scale, which means this application can be received by the user strongly agree.

4. CONCLUSION

Based on the results of this research can be concluded that this development can make 3D objects change according to the sound input given. The combination of the two

technologies has succeeded in creating 3D objects that are displayed when the camera detects a changing marker simply by providing voice input by the given command, although it will be difficult when we provide voice input in English. This is because the wit.ai platform not only accepts user voice input and extracts audio, wit.ai will also check whether the voice input provided is by the command and also the correct pronunciation in English. A combination of the two technologies in the application can be further developed to make it easier when detecting voice input given by the user and can also give points or scores when the user pronounces the sentence correctly, not only the 3D object that changes.

REFERENCES

- [1] Franciska, M. B., Setyawan, M. B., & Zulkarnain, I. A. (2018). Rancang Bangun Media Pembelajaran Bahasa Inggris Berbasis Android Menggunakan Teknologi Augmented Reality Untuk Sekolah Dasar (Studi Kasus Mi Ma'Arif Patihan Kidul). *Komputek*, 2(2), 48.
- [2] Syahidi, A. A., Tolle, H., Supianto, A. A., & Arai, K. (2019). *BandoAR: Real-Time Text Based Detection System Using Augmented Reality for Media Translator Banjar Language to Indonesian with Smartphone*. 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences, ICETAS 2018, 1–6.
- [3] Anggraini, N., Kurniawan, A., Wardhani, L. K., & Hakiem, N. (2018). Speech Recognition Application for the Speech Impaired using the Android-based Google Cloud Speech API. 16(6).
- [4] Aouam, D., Benbelkacem, S., Zenati, N., Zakaria, S., & Meftah, Z. (2018). *Voice-based Augmented Reality Interactive System for Car's Components Assembly*. Proceedings - PAIS 2018: International Conference on Pattern Analysis and Intelligent Systems, 1–5.
- [5] Wijaya, S., & Wicaksana, A. (2019). JACOB voice chatbot application using wit.Ai for providing information in UMN. *International Journal of Engineering and Advanced Technology*, 8(6 Special Issue 3), 105–109.
- [6] Agustini, M., Yufiarti, & Wuryani. (2020). Development of learning media based on android games for children with attention deficit hyperactivity disorder. *International Journal of Interactive Mobile Technologies*, 14(6), 205–213.
- [7] Ahmad, A., Hadiansa, A., Hidayatullah, R., Informatika, J. T., Informatika, J. M., & Informatika, J. T. (2018). *Len t e r a d u m a i*, . 9, 42–46.
- [8] Akrim, M. (2018). Media Learning in Digital Era. 231(Amca), 458–460.
- [9] Billinghamurst, M., Clark, A., & Lee, G. (2014). A Survey of Augmented Reality Foundations and Trends R in Human-Computer Interaction. *Human-Computer Interaction*, 8(3), 73–272. CS4HS 2016 The University of Queensland. (2016).
- [10] Faqih, M., & Kusumaningsih, A. (2018). PENERAPAN AUGMENTED REALITY PADA SERIOUS GAME EDUKASI PENYAKIT GIGI. 9(2), 1033–1042.
- [11] Hashim, N. C., Majid, N. A. A., Arshad, H., & Obeidy, W. K. (2018). User Satisfaction for an Augmented Reality Application to Support Productive Vocabulary Using Speech Recognition. *Advances in Multimedia*, 2018.
- [12] Megista Putri, D., & Rasmita, R. (2019). Upaya Meningkatkan Minat Belajar Bahasa Inggris Siswa Sekolah Dasar (Sd) Negeri 13 Padang Menggunakan Metode "Total Physical Response." *Jurnal Kepemimpinan Dan Pengurusan Sekolah*, 4(1), 11–18.
- [13] Musthafa, B. (2010). Teaching English to Young Learners in Indonesia : Essential Requirements. *Educationist*, IV(2), 120–125.
- [14] Saputra, D. S. (2017). Interactive Learning Dalam Pembelajaran Speaking Di Kelas V Sekolah Dasar. *Jurnal Cakrawala Pendas*, 3(1).
- [15] Simonetti, Alexandro; Paredes, J. (2016). Vuforia v1.5 SDK: Analysis and evaluation of

- capabilities. *Medicina Clinica*, 147(9), 393–396.
- [16] Waliyuddin, M. H., Sukamto, A. S., & Anra, H. (2019). Rancang Bangun Aplikasi Panorama Wisata Kota dalam Upaya Pengenalan Budaya dan Pariwisata Kota Pontianak. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 7(2), 113.