Marker-less Augmented Reality Interactive Furniture Layouting (Voice Recognition and Narrator)

Software Requirements Specification

Project ID: 16-105

Author: Alawatta A K S Y

IT13030254

Sri Lanka Institute of Information Technology

Special Honors Degree of Bachelor of Science in Information Technology

Submission Date : 01-04-2016

Project Title: Marker-less Augmented Reality Interactive Furniture Layouting (Mark less Furniture)

Project ID: 16-105

Name of the Author: Alawatta A K S Y

Supervisor: Mr. Udara Samaratunge

Signature of Supervisor: .....

Date of submission: .....

Declaration

I hereby declare that the project work entitled “Markless Furniture” (Voice Recognition and Narrator) submitted to the Sri Lanka Institute of Information Technology, is a record of original work done by our group under the guidance of Mr. Udara Samaratunge (Supervisor) and Mr. Tharindu Perera (Co- Supervisor), and this project work is submitted in the fulfillment for the award of the Bachelor of Science (Special Honors) in Information technology Specialization in Software Engineering. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree or diploma. The diagrams, research results and all other documented components were developed by myself and I have cited clearly any references I have made.

Name : Alawatta A K S Y Signature :

Table of Contents

List of Table

List of Figures

1 Introduction

1.1 Purpose

This document intended to illustrate the requirement for the project “Mark less Furniture”. The software describes this document is voice recognition and narrator in augmented audio format mapping with the augmented reality environment for customers to customize their object in an augmented way. The document gives the detailed description of the functional and non-functional requirement for the purpose system. It explains the high-level architecture, purpose system features, User Interfaces, User requirements, what the system purpose is, and how the system react with external factors.

On the whole, this document is written with the purpose of providing a complete description on the behavior of voice recognition and narrator in augmented audio format mapping with the augmented reality environment to be developed will be provided in this document for better understanding about system included features, sub-modules, how those sub-modules works with each other, and the relationship to customer, developers. The intended audience of this document is the project supervisor and research team.

This document will also benefit the former researchers, developers who will be engaged in developing and maintaining the application in the future and any person, who is interested in implementing this kind of application.

1.2 Scope

This document covers the requirements for the first release of “Mark less Furniture”. The release will focus on providing a better solution to real world problems in purchasing furniture for both seller and customer. This application covers basics that customer considered before buying the furniture from the store. Basically by using this application it maps the customer environment to virtual reality, then it give the chance for applying furniture into the virtual customer’s environment before purchasing, so that customer can get more experience about how the furniture fit with actual world scenario, and also the application provide the suggestion regarding the actual environment so that customer can see the suggested furniture according to their environment parameters, customer can trigger the application by using voice commands. The customer can apply changes via voice command. And augmented audio voice description uses for gain better experience in the augmented world to the customer.

1.2.1 Speech (voice) Recognition

User input is given as voice command and should be detected via a speech recognition module. Speech (voice) is a continuous audio stream mixed with dynamically changed states [1]. This is a most challenging part, there are no clearly distinguished parts in an audio recording in the waveform. Words are understood in android (the smallest unit of a waveform) but it is low in true percentage [2].

The common way of speech recognition in android (e.g.: Google Now) is conversational mode. It is taking a waveform, split it by spaces try to identify word such as calendar, alarm. Then apply conversation or open the application. If it couldn't find any matching word convert all to the search query in a google search. [2]

In here we take a waveform, and transfer it to text format, then utterance parsing module process into formalized command, and finally search for actionable objects and manipulate highest marching object /command

1.2.2 Text To voice and Audio Augmented Reality

The input is given as text format by this component it will convert to the speech format. There have been several technologies which have been implemented by basically using AIML based on XML. Google API in Android. The speech covers most instance of making a device for making blind people assistance.

Augmented Audio Reality refers technology that ambled artificially generated audio into user's real environment. According to previously done research on AAR, there is three different geographically definition for head related coordination system [3]

- Origin
- Horizontal plane
- Median plane

In our first iteration, we are going to research about origin based one format perception of distance.

Determine the distance is the challenging part. Distance perspective variance on different perspective/ factors. An important distance factor for sources is the ratio between direct and reverberant sound energy. [4]

For Augmented Audio Reality, we have to convert android TTS APIs matching with android AR.

In our first iteration, we are providing a basic AAR with TTS system to perform augmented reality in mobile technology. For AAR, we have to measure distance according to image factor. The visibility level of the image, we assign according to a scale in origin point factor. These sections are handled by AR assistance module and Narrator will be handled by TTS Agent in Android Platform. According to the scale, we can divide the augmented area and apply augmented audio TTS.

1.3 Definitions, Acronyms, and Abbreviations

Acronyms

API	Application Programmable Interface
SRS	Software Requirement Specification
AR	Augmented Reality
TTS	Text to Speech
SDK	Software Development Kit
RAM	Random Access Memory
VGA	Video graphic adapter
AAR	Audio augmented reality
Android	Mobile Operating System
OS	Operating System
UI	User Interface
UX	User Experience

Table 1: Acronyms

1.4 Overview

1.4.1 Main Goal of the Software

The goal of voice recognition and augmented audio is to deliver a great support to the consumer to customize the furniture in a very efficient way via the less consumer involved method and have a great augmented reality experience by providing voice description in the augmented audio format in most famous Android mobile platform. Like another application on Augmented Reality furniture assistance software is more complex to use and customize, so that those are less famous in the world. In here we are researching for provide best customizing and efficient friendly verbal approach for the better UX.

1.4.2 Tasks

- Implement a speech recognition API to translate the speech into text format. The speech recognition engine can train focusing a specific person's voice on increasing its accuracy.

- By analyzing input text supplied by speech recognition engine and the process agent can understand the user request

- Agent checks for available commands and method for executing the relevant command that according to provided text details.

- Once generate the output ,provide it in proper managed way to user

- Text to voice module gathers the description from web service according to the selected object

- The TTS and AR agent identify the object perspective in mobile camera and module identify the distance of the camera

- AR assistance module provides scale according to the model and divides augmented areas according to the scale of image perspective

- TTS module converts the text into verbal format and according to AR assistance's parameters provide augmented audio experience

- Finally, provide the output of audio and play using pre-configure voice.

This document contains information about the "Mark less Furniture" and how the application will be developed. In the next chapter, it will describe the overall description of the product in product perspective with all software, hardware, user and communication interfaces, memory constraints, operations site adoption requirements, product functions and User characteristics.

1.4.3 Organizing of the SRS

In **section 1**, the purpose of the preparing the document is explained, the aspects of the application the document tends to cover , the scope of the software ,Glossary and an overview of the software , it's main goals, tasks and users are described .

Section 2 describes the Overall description of software is explained in a non-technical manner, It includes product perspective, product functions, user characteristics, constraints, assumptions, dependencies of requirements.

Section 3 describes the technical requirements of the software and contains functional requirements, non-functional requirements and design constraints. It also includes the class diagram of the system.

Section 4 describes contains supporting information for the readers of this document.

2 Overall Descriptions

2.1 Product perspective

Project “Mark less Furniture” can be related several category of software, mainly it based on augmented reality, furniture mapping, voice recognition and image processing. By combining this component and its sub-components of software and by studying the existing product available under that category we hope to implement a software solution for the customer.

There have been several voice-controlled personal assistants such as “Google Now”, “Microsoft Cortana” and “Apple’s Siri”. In that application are user-friendly, expect Google Now others are intelligent, but these applications have been created in such a way to that users can interactive with limited functionalities predefine according to the device.
“Google Now” is a dialog system Android, that support execute the function of most google product. For example, users can send a message by saying “Send a Message to 123456789” then it prompts the message application by asking one by one such as “who do you want to send”, “what is the message”, and similarly with other commands.

For Narrator/ Text to speech, there is several type of software have been deployed namely screen reader, magnifiers. But these methods are time-consuming and read what is on the screen. In our product, we have to implement a solution for reading web service’s data to narrator format.

AAR the 3D voice implementation with Narrator is not commonly found in the real world. As in Audio Augmented Reality, it basically applies to real augmented reality world with AR devices such as augmented reality stereo headphones. In here we implementing a system for augmented audio reality to apply with normal devices such as mobile phone and normal headphone.

In project “Markless Furniture” the sub-components of high-level sub-diagram for apply AAR with limited functionalities, Text to speech for web service’s outcome and

voice recognition system without conversational mode we need to follow different approach solution. For those design approach clearly described in flowing high-level diagrams in figure 1 and 2

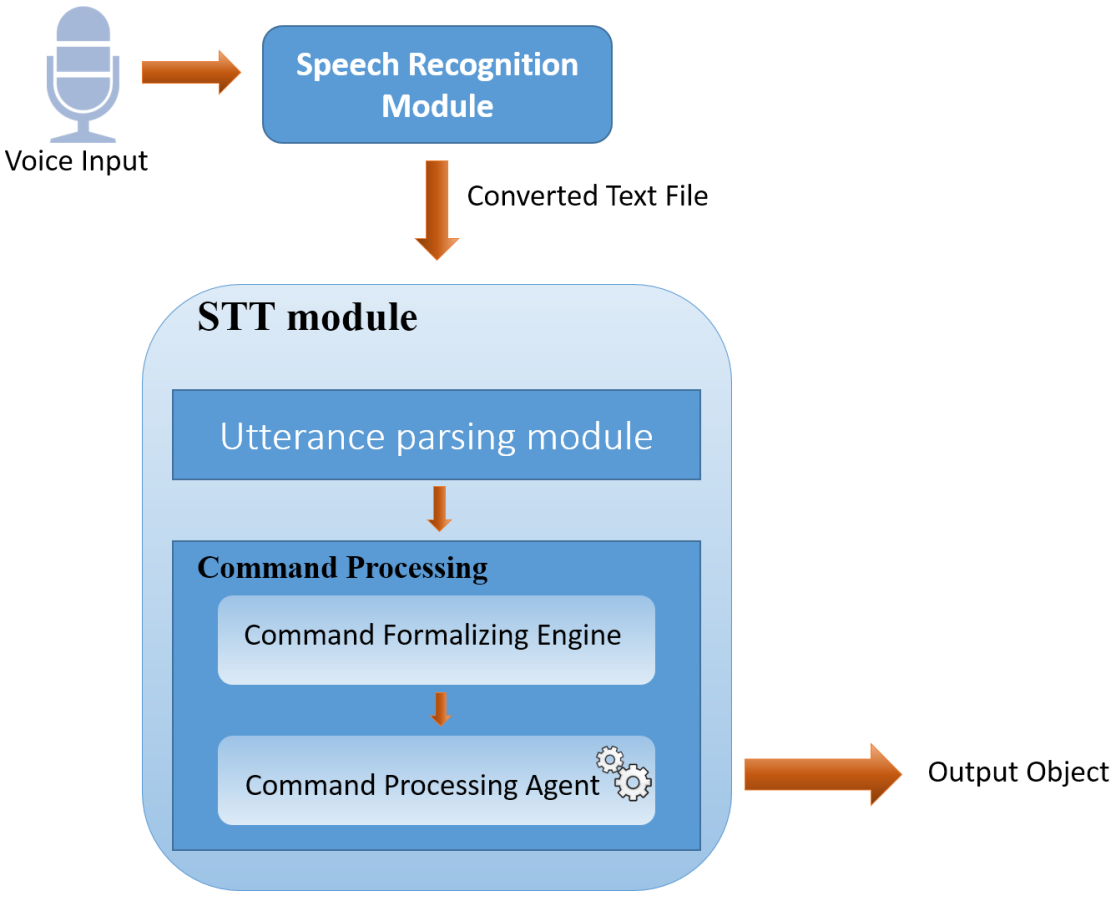

Figure 1: High-Level Diagram (Voice Recognition)

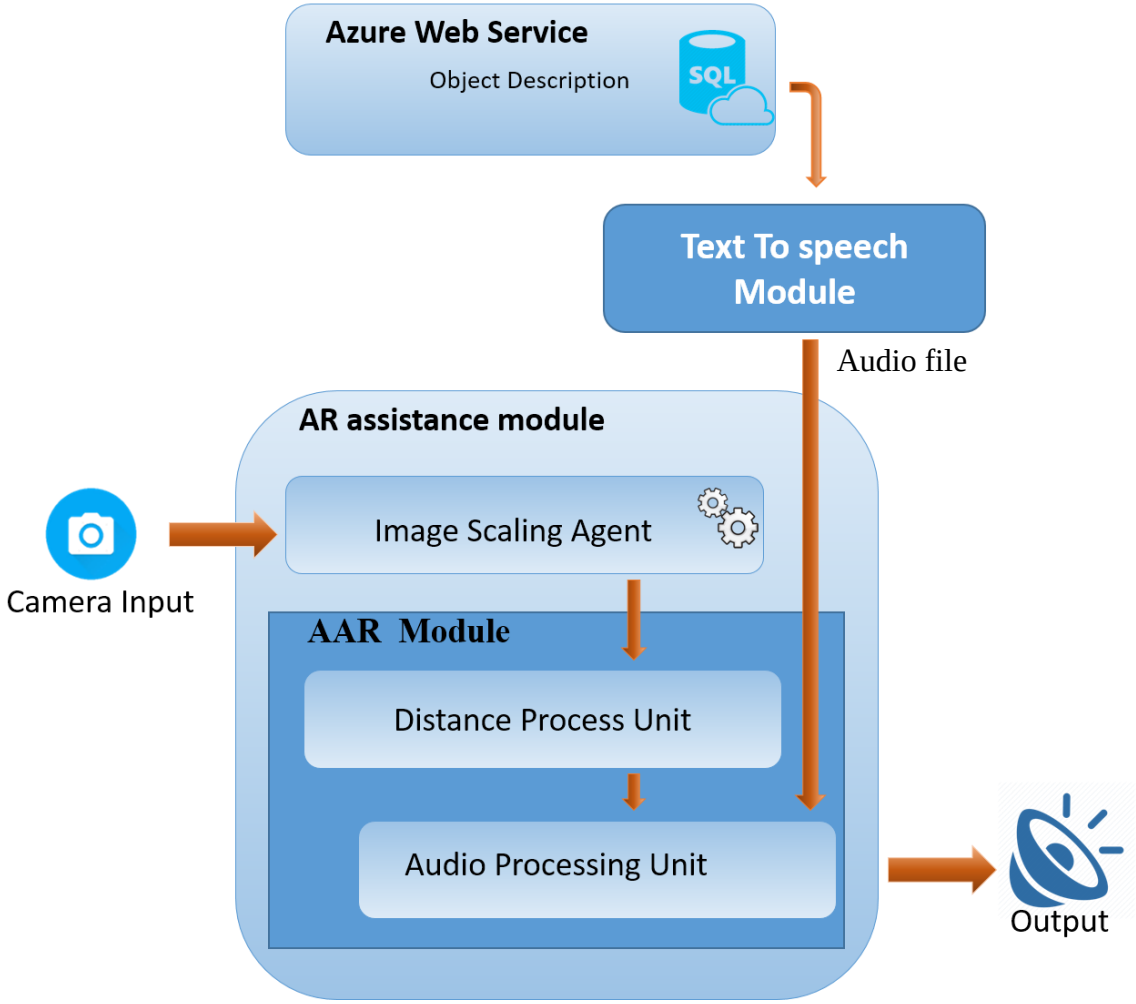

Figure 2: High-level Diagram (AAR+TTS)

2.1.1 System interfaces

Voice recognition and AAR is a component within the Mobile Application “Markless Furniture”. The component functions as the foundation for the augment reality functionalities and will be utilizing the mobile device’s camera for recognize the environment and the mic for recognize the user’s voice.

And devices that support to access the following android API

- Android voice recognition API
- Android Augmented Reality API

For this component to function the mobile device should work on Android OS (4.3 Jelly Bean or newer).

2.1.2 User Interfaces

Figure 3: Main User interface

Figure 4: User Interface Processing the command

2.1.3 Hardware interfaces

An android OS running mobile device consisting of,

- Octet-core Processor
- Microphone
- Speakers for TTS output
- Mobile back camera more than 5MP
- Mobile Device's sensors
    - Accelerometer

2.1.4 Software interfaces

TTS engine – for provide read out loud service

Optical Character Recognition (OCR Platform) – for recognize optical characters

2.1.5 Communication interfaces

The user's smart phone should support high speed data communication methods such as 3.5G (HSPDA+), 4G (LTE) or Wi-Fi connection.

2.1.6 Memory constraints.

"Markless Furniture" is expected to use no more than 1GB of Ram and 1GB of external storage.

2.1.7 Operations

- User should enable the internet access of the mobile device.
- Download the furniture they wish to place using the serial No. obtained from the dealer.
- Aim the mobile camera at the location he/she wish to place the 3D furniture.
- Click on Microphone icon
- Say the object ID that want to edit
- Say the execution command and details that want to modify
    - Change Color to red
- Say "Description" to listen the Description
- Move the camera near and far for better hearing the description

2.1.8 Site adaptation requirements

Since the user can be of any nationality the user interface must be created for English language.

2.2 Product functions

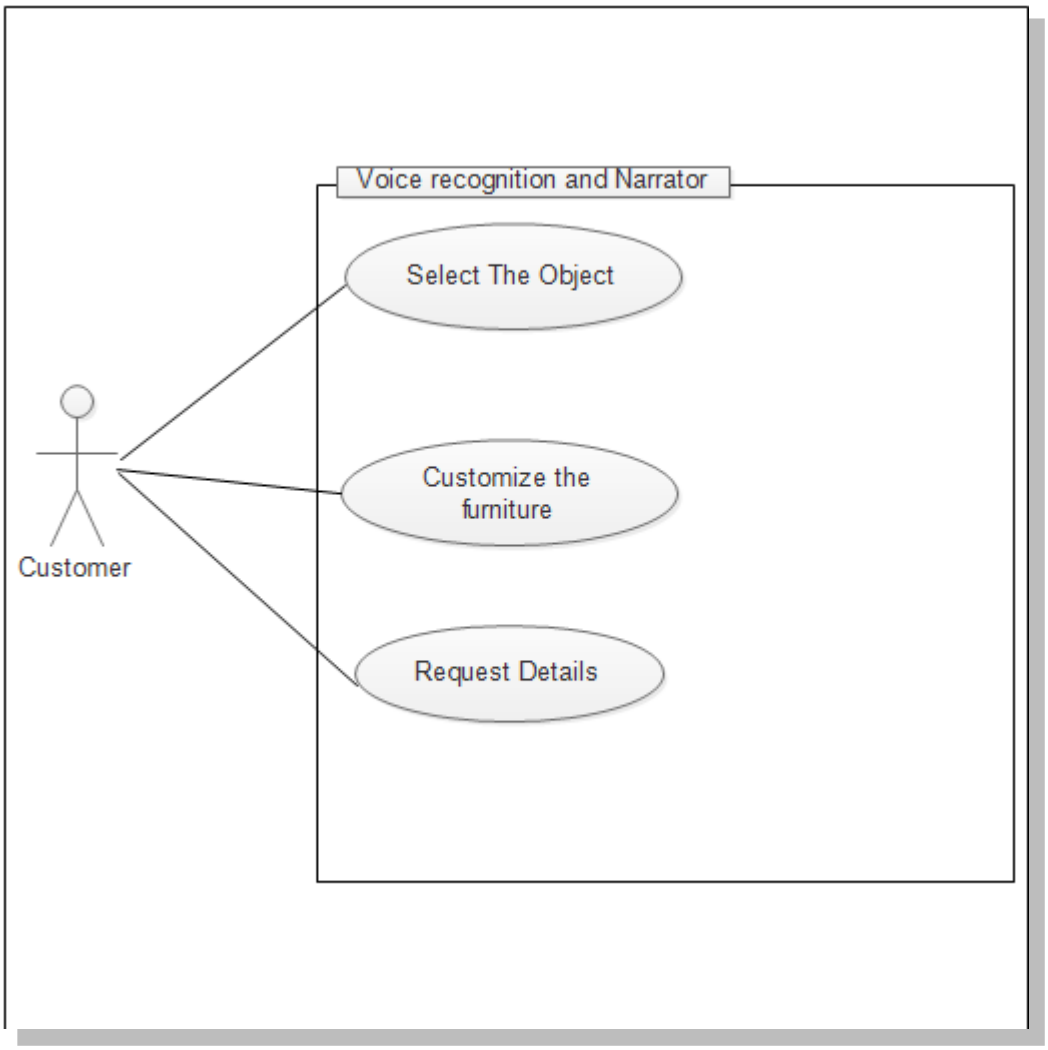

Figure 5: Use case Diagram for Voice Recognition and Narrator

Use case scenarios

Use case 01	Select the 3D object
Primary Actor(s)	User (Customer)
Pre-Condition	Mobile device is connected to the internet The object is placed in the environment Properly configure the Android Voice Recognition API
Post-Condition	
Main Success Scenario	1.0 Use case starts with the object placed in the environment 2.0 User clicks on the microphone button on the left corner 3.0 User input voice input to select the object 4.0 System validates the object id 5.0 The use case ends with locking the object and wait for further instructions
Extensions	4.a Disconnected for internet, system displays a message to reconnect. 4.b Invalid object id, system displays an error. 5.b The dialog show processing dialog

Table 2: Use-case Scenario

Use case 02	Customize furniture
Primary Actor(s)	User (Customer)
Pre-Condition	3D object is placed in the environment The selected furniture has customizable features (colour, texture, design)
Post-Condition	The furniture is customized
Main Success Scenario	*1.0* User input customization command in voice *2.0* System checks for available customize commands and variables
Extensions	*2.a* Disconnected for internet, system displays a message to reconnect**.**

Table 3: Use case Scenario

Use case 03	Request Details
Primary Actor(s)	User (Customer)
Pre-Condition	Mobile device is connected to the internet The object is placed in the environment Properly configure the Android Voice Recognition and TTS API
Post-Condition	Give the voice description via voice
Main Success Scenario	1.0 Object is selected properly 2.0 User asks for details 3.0 System validates the object id and get the relevant details 4.0 System convert description to the voice
Extensions	3.a , 4.a Disconnected for internet, system displays a message to reconnect.

Table 4: Use case scenario

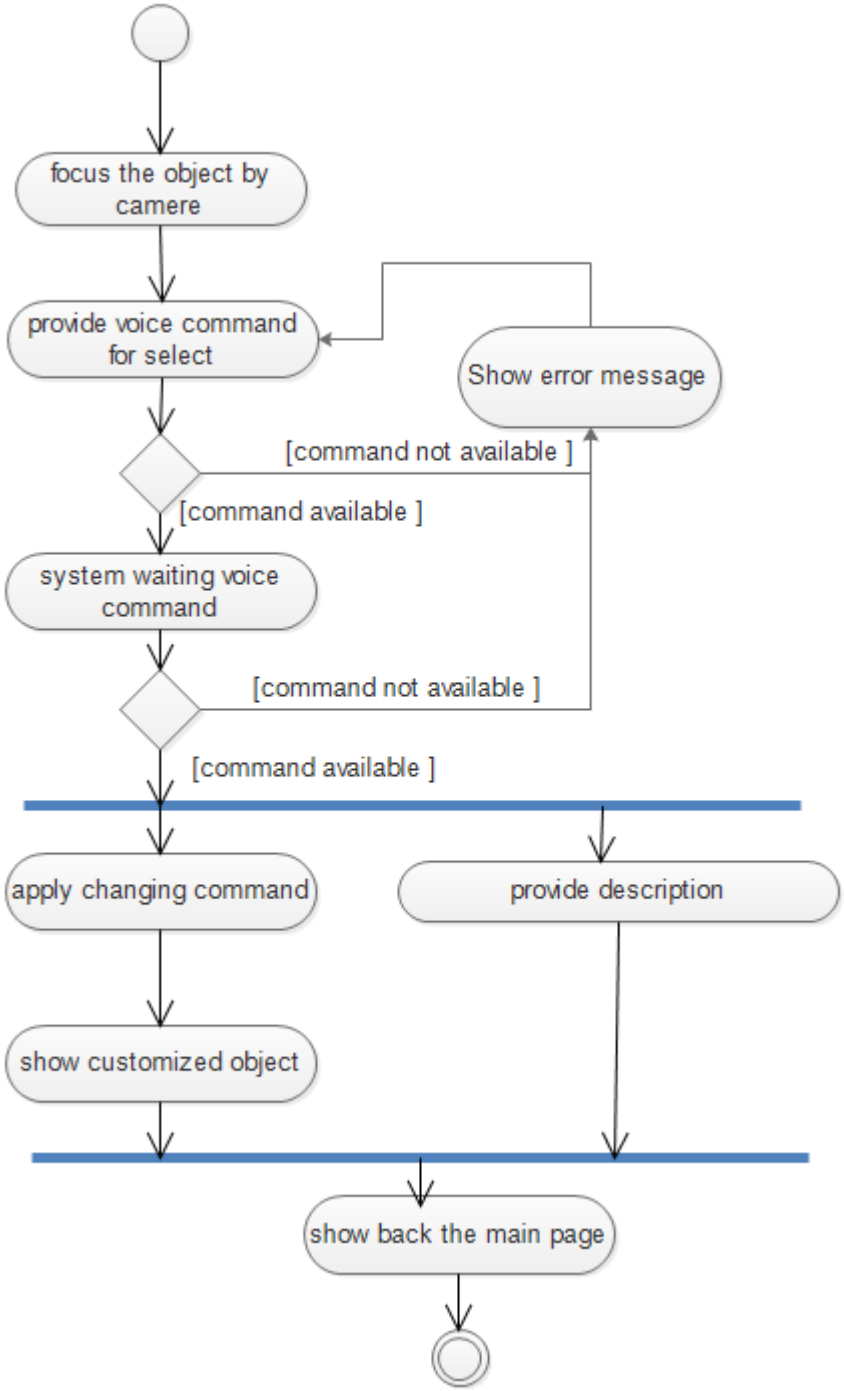

Figure 6: Activity diagram

2.3 User characteristics

The system is mainly focus for 2 type of users

- Furniture Dealer, who wish to utilize the application to advertise their products.
- Customers, who wish to examine furniture before purchasing.

This application is mainly developed for any customer with compatible smart phones who wish to purchase furniture and like to examine the products prior to purchasing. The application will retrieve information of 3D furniture which the furniture dealers will upload to the database and provide a serial number for each product in the database. Customers can retrieve this serial number from the furniture dealer and utilize it in the application to download the 3D furniture and augment.

2.4 Constraints

The propose system consist of a mobile application and web application. The constraint for each are considered separately. This document focuses on explaining the mobile application constraints.

- The 3D furniture model downloading is depend on internet connectivity.
- Perform well with mobile sensors , microphone, and camera
- The mapped object identification and apply command performed automatically without any customer intervention.
- The application only allow download limited amount of furniture and apply narrator to maintain the performance of the application.
- A recognition rate of 80% is expected but this may drop to a lower level if a low quality microphone is used.

2.5 Assumptions and dependencies

- The user is assumed to have a sufficient knowledge at using mobile phone and also a sufficient knowledge in using common internet services

- User is assumed to have a knowledge of English, verbal and reading, since the software is configured to recognize commands given in English.

- The software is dependent on the quality of the microphone used in recognizing voice input.

- Best quality of camera used in identify objects

2.6 Apportioning of requirements

The section 1.5 of this document provides the overview of the proposed system and the Section 2 provides the detailed overall description on the system and the requirements. The section 3 contains detailed requirements that should be followed while implementing the system.

Project Markless Furniture will confine to both functional and non-functional requirements mentioned in section 2 and 3.

The methodology of implementation may slightly differ but will ultimately tally with the purpose and objectives mentioned in the document.

3 Specific requirements

3.1 External interface requirements

3.1.1 User interfaces

Figure 7: Main User interface

Name of the item	Add button
Description of purpose	Display a button forget the operation of voice inputs
Source of input or destination of output	Touch
Valid range, accuracy and/or tolerance	N/A
Units of measure	No
Timing	No
Relationships to other inputs/outputs	No
Screen formats/organization	Top right
Window formats/organization	right of screen
Data formats	N/A

Table 5: User interface details for figure 1

The voice recognition and AAR sub module are using assistive technology. These processes are running on background in the main features.

The AR module uses software for better experience, so that scaling methodology uses.

Users may use the software's voice recognition part as a digital virtual assistant completing those tasks need to access internet and control mobile using voice command

Before start using software user have to complete all pre configurations on their mobile environment.

3.1.2 Hardware interfaces

- Octet -core Processor (at least quad core)
  The process multitasking capabilities, in-order to Augmented Camera to function in background while other functions comes into action.

- Accelerometer
  To identify the orientation of Camera to the real-world

- Mobile Camera (at least 5MP 720p@30fps video quality)
  To get the better image and location of the object

- Speaker and Microphone
  For User input and out puts

3.1.3 Software interfaces

- Unity 3D

The main integrated development environment on which the mobile application and its respective components are developed.

- Microsoft Azure Cloud Server

This is used for Cloud storage of 3D objects and their respective information.

- Android TTS engine and Optical Character Recognition (OCR Platform)
  For voice narrator need to use android TTS engine and OCR platform. The engine should installed and running properly.

3.1.4 Communication interfaces

The data communication between mobile application and the web service is achieve through internet connections. Therefore data communication between centralized database of cloud and mobile device's local database is handled through the internet. The mobile device's connect can range from minimum of 3.5G connection to fast 4G (LTE) connection.

3.2 Classes/Objects

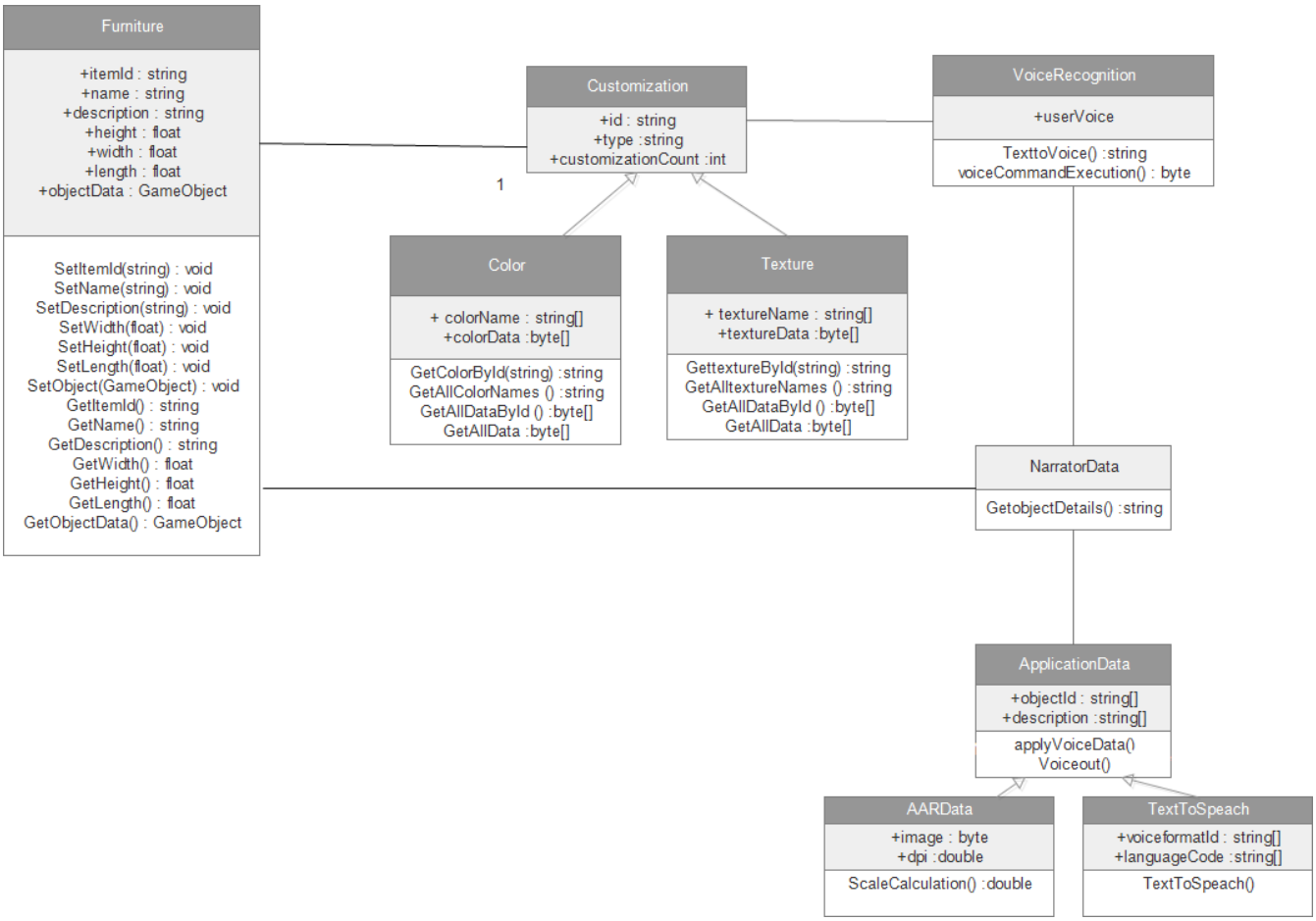

Figure 8: Class Diagram

3.3 Performance requirements

Since the application runs in the background, considerable resources will be allocated, and hence a modern mobile phone with significant amount of RAM (at least 1 GB) will be required for smooth usage of the software.

A good microphone with noise cancelling technology will ensure accurate user input which is another vital component in the software.

The distance calculator feature information will iteratively increase or decrease to maintain the optimal level necessary maintain the stability of augmented objects

3.4 Design constraints

Multi-Threading/Processing-The software requires multiple threads, i.e. a separate thread is needed to listen for user commands and another thread for the given commands to be executed, and this is done because if the user wants to stop execution of a command, the user can instruct the system to do so from another thread. This will be impossible to do so if the application ran on a single thread.

3.5 Software system attributes

3.5.1 Reliability

The system should be able to perform and maintain its function without the continuous use of internet. If the internet connection is lost during the data download, identification, the data in destination shall not be corrupted. In the context of voice recognition and object identification, the location of the object should be fixed and the description shows the uncompleted error message.

Test cases with reference to use-case diagram (figure 5)

- Select the object

Test case #	Description	Pre-conditions	Test Steps	Input	Expected output
01	Select 3D furniture successful	• Device is connect to the internet • Application is open • The object placed on the environment	1. Click on the microphone button in left upper Conner 2. Customer say the voice command with object no	Serial No : object1 Command: "Select object1"	• Application processing for the next object

Table 6 : Test case 1 for selecting object

Test case #	Description	Pre-conditions	Test Steps	Input	Expected output
02	Select 3D furniture unsuccessful	• Device is connect to the internet • Application is open • The object placed on the environment	3. Click on the microphone button in left upper Conner 4. Customer say the voice command with invalid object no	Serial No : item02 Command: "Select item02"	• Show the error message "The Item is not valid" • And prompt for another input

Table 7: Test case 2 for selecting object

- Customize furniture

Test case #	Description	Pre-conditions	Test Steps	Input	Expected output
04	Change of Colour of furniture	• One item is augemented on to the environment • The item has colour customization options • The object is selected for further instruction	1. Say the valid voice command to customizing with color	Command: "Change color to RED"	• The augmented item's colour changes to red

Table 8: Test case 1 for customizing object

Test case #	Description	Pre-conditions	Test Steps	Input	Expected output
05	Change of texture of furniture	• One item is augemented on to the environment • The item has colour customization options • The object is selected for further instruction	1. Say the invalid voice command to customizing with color	Command: "Change color to Black	• Showing error message "the requested color is not applicable for this product" • Showing for the next input

Table 9: Test case 2 for selecting object

- Get description

Test case #	Description	Pre-conditions	Test Steps	Input	Expected output
05	Change of texture of furniture	• One item is augmented on to the • The object is selected for further instruction	2. Say the valid voice command to get description	Command: "description"	• Readout the description of the object

Table 10: Test case 1 for description

Test case #	Description	Pre-conditions	Test Steps	Input	Expected output
05	Change of texture of furniture	• One item is augmented on to the environment • The item has colour customization options • The object is selected for further instruction	3. Say the invalid voice command to description	Command: "read out information"	• Showing error message "the command is invalid." • Showing for the next input

Table 11: Test case 2 for description

3.5.2 Availability

The mobile application should be available for access by any user. The application should not rely completely on internet connection and should only notify the user when it is necessary. The 3D furniture and their respective information which was downloaded should be available with-in the application throughout the runtime for easy access and only be remove once the application is closed.

3.5.3 Security

3D furniture models and their respective information will be maintained with-in the mobile device only while the application is active. When the application is close all downloaded details will be remove to keep confidentiality of information. Encryption methods should be used to store important data.

.

3.5.4 Maintainability

The system should be implementation should be done in a way to preserve the maintainability. The data-storage should be organized and information of performance issue which might occur should be logged and store with proper encryption for reference. The system be standalone to avoid any interruption from maintained done on the server-side. The mobile application should be able to upgrade with-ease.

3.6 Other requirements

Extensibility and Modifiability: System should be able to accept new features and customizations with ease.

Accessibility: The user interface should be easily accessible and understood by any use.

4 References

[1] "Information Extraction State of the art" [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=0FCBDF7214FAB1B049E008D4DBCD744F?doi=10.1.1.103.6834&rep=rep1&type=pdf [Accessed 28/03/2015]

[2] " JustSpeak: Enabling Universal Voice Control on Android" by Yu Zhong, T.V. Raman, Casey Burkhardt, Fadi Biadsy and Jeffrey P. Bigham; Google research publication 2004

[3] "Enabling technologies for audio augmented reality systems" by Hannes Gamper, Aalto University publication series ,doctoral dissertations 39/2014

[4] Augmented audio reality: Telepresence/VR hybrid acoustic environments by M Cohen, S Aoki, N Koizumi - Robot and Human …, 1993 - ieeexplore.ieee.org