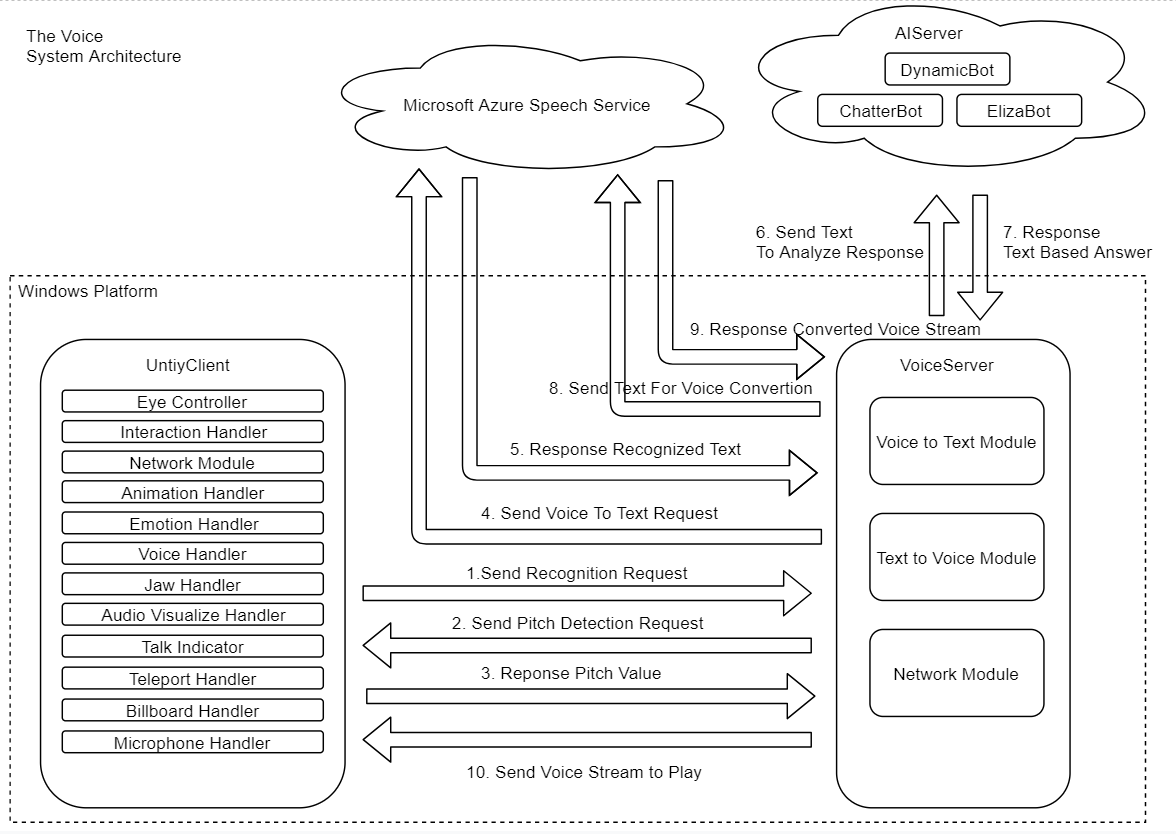
# Introduction

This is an immersive dialogue experience based on technologies of machine learning, virtual reality, natural interaction, speech recognition and voice synthesis, etc.

In a beautiful park, you can chat with your friend at zero distance. Your friend can express their feelings for your speech. Your friend knows everything, no matter what question you ask, she will answer it correctly. This is a dialogue experience based on dialogue topic data training, real-time web search, and advanced techniques such as natural language synthesis.

# System Architecture



# Problems We Solved

1. For the solutions of speech recognition, we did some research on this, finally select Microsoft Azure Cognitive Service because of better performance, but problem is it’s not easy to use with Unity3d, so we find a solution that is create a VoiceServer which is individual process running on windows platform to connect Azure Service, that make extra work but worth to do it.
2. For the facial animation to perform different emotions, normally model should have blendshapes for emotion animating and blending, but the resources we got is not very professional, only have bones on face, so we can only control bones to perform different emotion, that cause a little bit unnatural and low amount of emotions, so we only implement happy and sad face.
3. We use microphone sampling to get player’s voice pitch to detect how much it should perform on certain emotion, but it’s not work like we expected, because we considered the pitch level and keyword together to determine results, for example, “I am happy” with 100 pitch will perform a 100% happy face, but “I am not happy” with 100 pitch still perform 100% happy face, because our backend not able to know the real meaning of texts, I believe this function take more times.
4. We me a problem on training conversation data set, because the data is limited, so during the conversation, player always get same responses, for this reason, we develop a module called “Dynamic bot”, to implement a runtime search function when player send any texts, if player said “Who is Michael Jackson”, “Dynamic bot” will search on internet immediately and get one of abstract from results.

# Client Structure Design

**Interactive Handler**

This module contains function that handle all interact behaviors, such as move forward, turning, move backward, range detect, microphone distance detect, etc.

**Emotion Handler**Controls the face bones on model, convert text-based command to facial animation.

**Animation Handler**

This is the controller of model’s animator, have different kind of animation state, and simple state machine to implement full body and upper body animations. In animator, there are different animation layers to implement blending.

**Voice Handler**

This module receive voice stream and other command from VoiceServer by local network, parse them and play voice sound in unity as 3d sound.

**Audio Visualization**

This module able to detect player’s voice highest pitch during taking, then send to backend to implement emotions with different weights.

# AIServer Design

# Voice Server Design

Voice Server have 4 modules:

**Speech-To-Text**

**Text-To-Speech**

There are all connected to MS Azure service.

**Backend Connection**

This module will send converted text content to our backend to analyzing, then get responses with correct answer and other parameters.

**Network Module**

This module will send messages between unity project and this process as a local network service.