Speech to Art – NUGU can Art

말하는대로

Kwak SangHeon
College of Engineering
Hanyang University
Dept. of Information System
Incheon, Korea
ekzmwos02@naver.com

Haixu Nan
College of Engineering
Hanyang University
Dept. of Information System
Guangzhou, China
omga2333@gmail.com

Goo Zheng Ze
College of Engineering
Dept. of Information System
Malaysia
zhengze00@hotmail.com

Ailin LI
College of Engineering
Hanyang University
Dept. of Information System
China
a975281621@gmail.com

Abstract— **Our team aimed to create a picture that users wanted to use NUGU to create a speech and display it on the screen, especially in the future, and we believe that this small program will bring a good experience to ordinary people, especially those who have difficulty getting out of the house.**

Role Assignments

Roles	Name	Task description and etc.
User/Customer	*Goo ZhengZe*	The student is expected to play an optimal role in using the speaker and function because he has a room for himself, and there is no language problem at all because he can speak both Korean and English. And I think he's able to pinpoint the problem because he's dealt with these similar things before.
Software developer	*Haixu Nan*	He thinks about the software system in general. And analyzes, categorizes the related software and examines the software that can be used to implement the application. he will investigate about what development tool we will use to implement.
Software developer	*Ailin LI*	She thinks about the software system in general. And analyzes, categorizes the related software and examines the software that can be used to implement the application. he will investigate about what development tool we will use to implement.
Development manager	*Kwak SangHeon*	development manager should confirm if the fuctions of software full fill users need. when they do not satisfy users, he gathers some information from users and customers. Then, he advises software developer to update it and mediate both of them.

I. INTRODUCTION

Motivation

After visiting the LG Signature Kitchen Suite Nonhyeon Showroom before, I realized a great deal about the future home saw in there. That's because the future is closer to us than we think, and technological advances have come to us much more familiarly than we think. Anyway, after experiencing this experience, I've been thinking about what features I can add to the artificial intelligence secretary, so that users can experience a better fresh experience. A few days later, on my way home from school, I was surfing the Internet on my smartphone, and I saw a post posted by a user, which showed how advanced AI is now and how it is. I got a second realization there, and it's also similar to the first experience I had.

At first, I simply thought, "If you speak through a NUGU speaker, it'll be fun to create a picture file based on that sentence!" But as I went through the process of embodying the idea, I thought about how to use it in a more practical way.

First of all, people, especially those who are sensitive to external factors, control the optimal temperature, humidity, and light that suit them to make their home environment the way they want it to be. Obviously, this adjustment can make up the best environment, but I think it's a little bit lacking. No matter how beautiful a painting or a landscape is, if you keep looking at it, it's the selfish mind of a person that feels bored. But is there a better alternative than that if you can always create a different landscape when you want to, say, and create it? Due to the limitations of technological advancement, output displays are limited, but there are still products that can

be displayed on textiles, and they are also designed to enable wireless communication. In other words, in the future, if you don't want to see the scenery outside, you just have to hit a flat blackout curtain and say a few words to the artificial intelligence speaker.

"Ari, draw a picture for me.”

"Yes, what kind of picture do you want?"

"The sunset scenery."

So we combined the NUGU speaker developed by SKT with the AI picture artificial intelligence, DALL-E, and when we speak Korean to the NUGU speaker, we translated the sentence into English through Papago and input it into DALL-E, and then analyzed the sentence to draw a picture and send it out. This way of speaking is much more useful and convenient, especially for people who are not comfortable with their bodies or are not familiar with increasingly complex devices.

Problem statement (client's needs)

- Environmental changes caused by lighting also affect people's emotions, so environmental changes caused by pictures displayed on the display also affect people.[1]

- There are many children of low age who are still unable to draw what they have imagined, and there are many people who want their own creations although they are difficult to draw due to lack of skills.

Research on any related software

A. NUGU

- NUGU is a voice recognition device equipped with an artificial intelligence platform based on big data, and it can understand and execute the context of conversations with customers, and its voice recognition function is also very good. In addition, SKT has its own NUGU play builder program, so it is an appropriate option for our team to proceed with the project because it is convenient to make related programs.

B. DALL-E 2

- The program is a Text to Picture tool created by an artificial intelligence company called OpenAI, co-

founded by Elon Musk and Sam Altman on December 11, 2015, with Dall-e 1 released in January 2021 and Dale 2 released in 2022. Among the artificial intelligence-related software that has been released so far, it is evaluated as the most powerful software except for certain parts such as animation style. In addition, some texts such as grades, violence, political content, and real celebrities are restricted, so users may feel a little sorry for not getting the picture they want, but families with children can also use it with confidence because they will be able to avoid unpleasant situations due to unwanted strange pictures. These features are thought to be suitable for voice recognition devices that are available to everyone.

C. Stability.ai

-

II. REQUIREMENT ANALYSIS

A. translate

It will build and utilize a translation site using Papago api.

(1) When a user using the device says, "Ari, please draw a picture" in voice, the NUGU device that recognizes the sentence forms a sentence and translates the user's voice into English.

(2) The goal is to set all sentences to be translated into English, and to increase the recognition rate of the language by adding additional Chinese and Japanese languages in the future, so that other foreigners can use it comfortably. If I speak English, I will recognize and print English sentences as they are.

B. Dall-E 2

(1) Translate the sentence and use Dall-E to create a picture from the text. Because of the high amount of computation expected, instead of running it on the official website, we will build a separate website on the server using Dall-E's api.

C. App

Apps include logging in, a previously created picture (including the sentences in that picture), and entering sentences. This app helps to store the created pictures and create pictures through the application in difficult situations. And it can be set so that the user can choose not only the style of DALLE but also the style of other AI paintings.

(1) Sign in

- After you install and run the app, the login page appears. Ask the user to write down their previously registered email address and password, and if they do not have an account, click the Register button at the bottom to proceed to the subscription page.
- If the user has forgotten their ID or password, press the Find ID/Password button to go to the recovery page.
- If you use functions through NUGU without signing up, functions such as picture library and sentence input are limited.

(2) Registration
- On the registration page, users write down the ID and password to use, as well as the email to authenticate. If you have duplicate IDs, you cannot sign up, and your name will only be marked with a simple ID without a nickname.
- Authentication messages will be sent in a random six-digit number, and you must enter the number in the authentication box before you can sign up.

(3) Picture Library
- When you use the application after registering an account, pictures that are generated in real time through the account interworking with NUGU play are stored within the application, and you can see when the pictures were created and what sentences were created.
- And they're basically within the application, but they can be downloaded and shared, and you can replace the pictures on the NUGU-linked monitor in the Pictures library with whatever you want.

(4) Enter a sentence
- In situations where it is difficult to speak through the application, you can use it by typing sentences directly and pressing the create button. If you enter a sentence in Korean (or another foreign language) and press the Create button, the sentence goes through a translator once, so you can write a sentence even if you don't know English.

(5) AI Painting Selection
- The most basic style setting is Dall-E, but since there are many other AI paintings, we will increase the options so that users can change them to other AI styles they want. (Now just can do Dall-E2, Stable Diffusion.).

(6) App Translation
- Basically, the language in the app is Korean, but for the convenience of other foreigners, we will convert it to enable language change, and there are English, Chinese, and Japanese languages to be added.

III. DEVELOPMENT ENVIRONMENT

A. *Platform*
  (1) Android

- "NUGUcanArt" should be used on a simple and easy-to-access mobile platform, and we chose Android as the best environment for our development level, especially since Android is a globally popular platform.

B. *Programming language*
  (1) JAVA
  The biggest feature of JAVA is that it is an independent language of a platform. The program made of java works without any problems if only jvm is installed for the platform. Also, the stability is excellent in two aspects. First, since it is a popular language, there are many references and open sources, and based on this, many large projects have been carried out, so stability is guaranteed in many areas. And secondly, it does not allow pointer variables or memory direct access functions, and it does not allow multiple inheritance, so it is highly stable. JAVA is used as a major language when developing in Android studio. And in our team's development environment, I chose the language because I thought it would be more stable to proceed with the development in a familiar existing language than to try the development in a new language such as Kotlin.

C. *Software in use*

  (1) Firebase
  Firebase is a real-time data cloud service platform under Google, which aims to provide a real-time response data service for APP. The platform is suitable for various cross-platforms such as IOS, Android, and web front-end. The core goal is to help developers build applications more efficiently. Specifically designed for developers, it separates the back end and deploys it in the cloud; it provides many cross-platform excuses, which greatly reduces development costs and shortens product development cycles. The tools it provides cover most of the services that developers usually have to build themselves. For example, it can be used to easily implement the email authentication and phone authentication functions required for user registration, without the need for us to build it ourselves, which is very suitable for beginners like us.
  Firebase also has the ability to provide cloud storage. Provides a powerful, easy-to-use, and cost-effective object storage service for storing and sharing images with ease. The Firebase SDK for Cloud Storage provides Google's secure-quality file upload and download services for your Firebase apps, regardless of network quality.

Firebase has free and paid plans. But the advantage is that the free plan already covers most of the features of this platform, which is enough to develop our application. Paid plans are only required after a certain amount of storage space and enough access is reached. Choosing it can reduce our development costs.

(2) Android studio

Android Studio is based on IntelliJ IDEA and is an official integrated development environment (IDE) for developing android apps. The android studio features IntelliJ's code editor and developer tools, and supports flexible gradient-based build systems, fast and functional emulators, integrated environments that can be developed for all android devices, code templates and GitHub integrations, extensive test tools and frameworks, tint tools, C++ and NDK support, and Google cloud platform. These features increase productivity when building the android app.

(3) NUGU PlayBuilder

NUGU Playbuilder connects to NUGU speaker to support SK Telecom's artificial intelligence speaker NUGU various services. Whose platform first identifies the user's intention to speak through speech recognition and natural language understanding. Then, it behaves and reacts appropriately in text-to-speech. NUGU Playbuilder is a GUI-based integrated development environment that provides the necessary technology for this process.

(4) GitHub

GitHub is a web service that supports git store hosting. A git is a distributed version management system and instruction to track changes in computer files and coordinate the operations of those files among multiple users. GitHub supports these feathers in the web format so that they can be easily viewed as a graphical interface. GitHub allows colleagues working on a project to share one workspace and set up an environment optimized for collaboration using the functions of git such as merge, commit, branch, etc. We will also share one repository to collaborate efficiently.

(5) Adobe XD

Adobe XD is one of the design tools supported by Adobe that is optimized for UI/GUI production. Before using Zeplin, we will use XD for prototype work. The XD has a sample screen that supports the SmartPhone, so it will be conveniently produced using the tool.

(6) Zeplin

Zeplin is one of the best tools for app designers and developers to collaborate. With its excellent compatibility with XD, designers can design the app and move it to Zeplin immediately to share it with everyone on the team. In addition, from the developer's point of view, Zeplin provides a guide on how to code after receiving the image, which increases productivity.

D. Task Distribution

TABLE II
TASK DISTRIBUTION

Name	Task description
Kwak SangHeon	Frontend, Play Builder
Haixu Nan	Backend
Goo Zheng Ze	Etc
Ailin LI	Etc

Task Distribution is shown in Table 2. Each member has his own main role, but since we do not know much about programming, we ask each other, share our thoughts and ideas to improve our project, and proceeds.

IV. SPECIFICATIONS

A. AI Speaker

(1) TTP(Text to Paint)

- The speaker sends a message when the user asks you to draw.

When the user sends a message to the speaker, "아리야, 그림 그려줘." the speaker who hears the m

essage runs NUGUCA and tells the user the following sentence: "네, 무엇을 그려드릴까요?" The user listens to the voice and says the sentence he wants, for example, "석양이 지는 모습" and the speaker converts the sentence into text with the TTS function and translates it through the translator. The sentence translated into English is then converted after a certain period of time through the set picture AI.

B. *Application*

(1) Creating and logging an account

The first thing users see when they open the app is the choice of logging in to their account or signing up for a new account. If the user chooses to join, the user will have to enter the email ID and password used by the user. First, information about whether the user's email is duplicated is sent to the database, and if there is no such email address in the database, it is indicated as available. Conversely, if there is an email address with the same name in the database, it is an email address in use. Afterwards, once all passwords are entered, the authentication number will be sent to the corresponding email address, and the registration will be completed by entering the authentication number.

If the user does not remember the password, the user must enter the password reset box and enter the email ID. This information is sent to the database, which sends an authentication number to the email address, and a window appears where you can enter a new password if the actual email is correct, and you can enter another password you want. And the password for the email ID will change. On the contrary, if you enter the authentication number and

find out that the email ID is missing, it will notify you that it is an email ID that needs to be registered.

(2) Main Menu
- When the user completes the login, a simple main menu with a total of four menus is displayed. The menus are "Picture Library", "Painting Creation", "Picture AI Change", and "Language Change", and at the bottom are logout and withdrawal items.

  ■ Picture Library
- When you enter the picture library, all the pictures you have created are classified by date and time, and by the order of creation according to what sentence you created and what AI you used. And the user can choose the picture they want and store it inside the device they use.

  ■ Painting creation
- This function is designed to be processed by entering sentences directly in the application for those who have difficulty in speaking. The basic processing structure is the same as the processing process after speech recognition. (Sentence input->Translation->Translated sentence input->Create a picture file by recognizing the corresponding sentence)
- At the bottom of the search bar, you can see what AI is currently in use.

  ■ Picture AI Change
- Dall E2, and Stable Diffusion show differences in each process and outcome, allowing the user to select each desired AI to create a picture, which will then be updated incrementally to add functionality.

  ■ Setting
- Settings allow you to change languages (Korean, Chinese, English) and log out.

C.

REFERENCES

[1] 박진희, 박영경.(2016).실내조명의 색과 조도에 따른
감정 상태의 관한 연구.한국색채학회 학술대회,(),67-68.