

A Design and Implementation of Speech Recognition and Synthetic Application for Hearing-Impairment

Woo-Lin Kim*, Hye-Won Ham*, Sang-Un Yun**, Won Joo Lee*

*Student, Dept. of Computer Science & Engineering, Inha Technical College, Incheon, Korea

*Student, Dept. of Computer Science & Engineering, Inha Technical College, Incheon, Korea

**Graduate Student, Dept. of Applied Artificial Intelligence, Hanyang University, Ansan, Korea

*Professor, Dept. of Computer Science & Engineering, Inha Technical College, Incheon Korea

[Abstract]

In this paper, we design and implement an Android mobile application that helps hearing impaired people communicate based on STT(Speech-to-Text) and TTS(Text-to-Speech) APIs and accelerometer sensor of a smartphone. This application provides the ability to record what the hearing-Impairment person's interlocutor is saying with a microphone, convert it to text using the STT API, and display it to the hearing-Impairment person. In addition, when a hearing-impaired person inputs a text using the TTS API, it is converted into voice and told to the interlocutor. When a hearing-impaired person shakes their smartphone, an accelerometer based background service function is provided to run the application. The application implemented in this paper provides a function that allows hearing impaired people to communicate easily with other people when communicating with others without using sign language as a video call.

▶ **Key words:** Google Cloud Platform, STT(Speech-to-Text), TTS(Text-to-Speech), Accelerometer sensor

[요 약]

본 논문에서는 STT(Speech-to-Text), TTS(Text-to-Speech) API와 가속도 센서 기반의 청각 장애인의 의사소통을 도와주는 안드로이드 모바일 애플리케이션을 설계하고 구현한다. 이 애플리케이션은 청각 장애인의 대화 상대가 말하는 것을 마이크로 녹음하고 STT API를 이용하여 텍스트로 변환하여 청각 장애인에게 보여주는 기능을 제공한다. 또한, TTS API를 이용하여 청각 장애인이 문자를 입력하면 음성으로 변환하여 대화 상대에게 들려준다. 청각 장애인이 스마트폰을 흔들면 이 애플리케이션이 실행하도록 가속도 센서 기반의 백그라운드 서비스 기능을 제공한다. 본 논문에서 구현한 애플리케이션은 청각 장애인들이 다른 사람과 의사소통을 할 때 영상통화로 수화를 이용하지 않고 쉽게 대화할 수 있는 기능을 제공한다.

▶ **주제어:** 구글 클라우드 플랫폼, STT(Speech-to-Text), TTS(Text-to-Speech), 가속도 센서

-
- First Author: Woo-Lin Kim, Corresponding Author: Won Joo Lee
 - *Woo-Lin Kim (rladnfls0500@gmail.com), Dept. of Computer Science & Engineering, Inha Technical College
 - *Hye-Won Ham (hyewon6588@naver.com), Dept. of Computer Science & Engineering, Inha Technical College
 - **Sang-Un Yun (dbstkdnds4@naver.com), Dept. of Applied Artificial Intelligence, Hanyang University
 - *Won Joo Lee (wonjoo2@gmail.com), Dept. of Computer Science & Engineering, Inha Technical College
 - Received: 2021. 11. 10, Revised: 2021. 12. 06, Accepted: 2021. 12. 08.

I. Introduction

국립국어원의 자료에 따르면 청각 장애인의 79.3%는 언어장애를 동반하기 때문에 상대방과 대화에 어려움을 나타내고 있다. 또한, 청각 장애인의 29%는 수화를 못 하거나 간단하고 쉬운 것만 가능하다. 그림 1과 같이 청각 장애인의 대다수가 청각 장애뿐만 아니라 자신의 의사를 제대로 전달하지 못하는 언어장애를 동반한 중복장애를 가지고 있다.

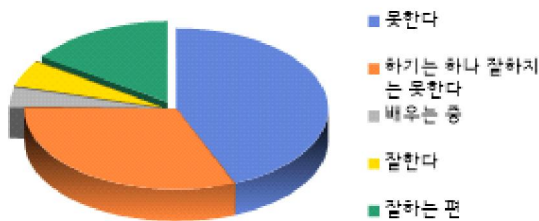


Fig. 1. Speech Skills for the hearing-impaired people[1]

청각 장애인들을 도와주는 도구들은 다양하지만 대부분 ‘듣기’에 집중되어 있다. 보건복지부의 장애인 등록현황에 따르면 전국에 등록된 청각 장애인의 수는 377,094명으로 추산되어 대한민국 전체 인구의 0.7%를 차지하고 있다[2]. 청각 장애인은 주로 상대방과 수화로 대화하기 때문에 전화 통신은 의사소통의 매우 큰 장애 요소이다. 택배 기사나 관공서에서 전화가 걸려 오면, 청각 장애인들은 주변 사람에게 매번 대신 받아달라고 부탁하거나 통화 거절 문자메시지를 보내야 하는 불편함을 겪고 있다. 청각장애인의 사회 참여 활성화를 위해 그들이 통화시 겪는 어려움을 해소하는 것은 필수적이다.

현재 서비스중인 청각장애인용 애플리케이션은 많지만 보통 상대방의 말을 텍스트로 변환하여 보여주는 STT 기능에 집중되어 있다. 따라서 언어 장애를 동반하고 있는 청각 장애인들이 갑작스러운 전화에 즉각적으로 통화하는 것은 쉽지 않다. 이러한 불편함을 줄이기 위해 본 논문에서는 기존의 청각장애인용 애플리케이션의 단점을 보완하여 TTS 와 STT 기능을 가진 애플리케이션을 개발하고자 한다. 이 애플리케이션은 음성 인식을 통해 상대방이 말하는 내용을 문자로 디스플레이에 출력하는 STT 기반의 기능을 제공하고, 음성 합성(을 이용하여 문자로 입력한 말을 음성으로 변환하여 상대방에게 자신의 의사를 전달하는 TTS 기반의 기능을 제공한다.

본 논문의 구성은 2장에서는 청각 장애인을 위한 기존 애플리케이션의 장단점을 분석한다. 3장에서는 청각 장애인을 위한 음성 인식 및 합성 애플리케이션을 설계하고, 4장에서 음성 인식 및 합성 애플리케이션을 구현한다. 그리고 마지막으로 5장에서 결론을 맺는다.

II. Preliminaries

1. Related works

1.1 Rogervoice

청각장애인용 자막 전화 통화 애플리케이션 Rogervoice는 그림 2와 같다[3].

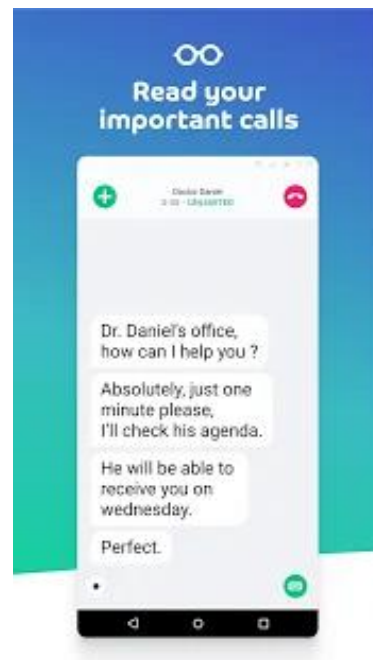


Fig. 2. STT of Rogervoice App

이 애플리케이션은 STT 기능으로 상대방의 통화 내용을 화면에 출력하고, 통화 목록 및 통화 내용을 기록하는 기능을 제공한다. 하지만 청각 장애인이 입력한 텍스트를 음성으로 변환해주는 TTS 기능은 지원하지 않고 있다. 언어 장애인들이 통화 내용을 파악할 수는 있지만 통화상에서 곧바로 대답하기 힘들다는 단점이 있다. 이는 통화 상대방이 청각 장애인임을 모르는 상태에서 난처한 상황으로 이어질 수 있다. TTS 기능이 추가된다면 통화상에서 청각 장애인들이 사운드로 대답할 수 있도록 하여 통화할 때 겪는 불편함을 해소하는 데에 도움을 줄 수 있을 것이다.

1.2 SpeakLiz

청각장애인과 비장애인이 대화를 원활할 수 있는 기능을 제공하는 SpeakLiz 애플리케이션은 그림 3과 같다[4].



Fig. 3. STT of SpeakLiz app

이 애플리케이션은 STT 기능을 이용하여 상대방의 대화 내용을 화면에 출력하고, 청각 장애인이 원하는 경우 주변 소리를 인식하여 어떤 소리인지 알려준다. 또한, 이 애플리케이션은 STT 기능과 TTS 기능을 모두 제공하며 웨어러블 기기를 사용하는 경우 센서와 카메라를 이용하여 수화를 인식하는 기능도 제공한다. 하지만, 이 기능은 별도의 웨어러블 기기를 보유한 사람만 이용 가능하며 ASL(American Sign Language)만 인식 가능하다. 해당 기능을 활용하여 국내에서 사용하는 수화도 인식이 가능하다면 국내의 청각 장애인도 이 애플리케이션을 편리하게 사용할 수 있다.

1.3 Talk to deaf

휴대폰으로 사용자의 음성을 텍스트로 자동 번역할 수 있는 애플리케이션은 그림 4와 같다[5]. 응답하는 청각 장애인의 경우는 키보드로 텍스트를 입력할 수 있으며, 원하는 대로 글꼴을 확대 및 축소할 수 있고 이력을 삭제할 수도 있다. 하지만 청각 장애인이 대화 상대와 반드시 같은 장소에 함께 있어야 한다는 한계점이 존재하고 사실상 음성의 텍스트 변환만 가능하므로 사용자가 텍스트로 입력하는 것과 별 차이가 없어 매우 불편하다는 단점이 있다.

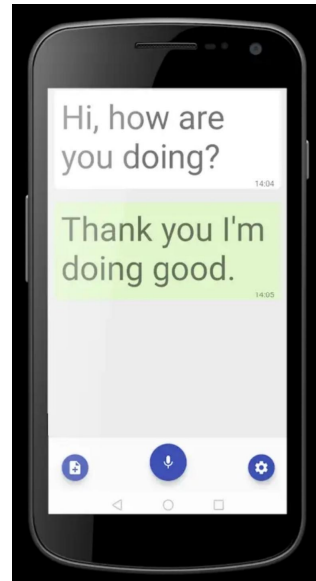


Fig. 4. Talk to deaf

III. Design of Application for Hearing-Impairment

본 논문에서는 Google 클라우드 플랫폼의 Speech API를 활용하여 청각 장애인들을 위한 안드로이드 애플리케이션을 설계 한다. 이 애플리케이션은 그림 5와 같이 사용자 인터페이스를 구현한다.

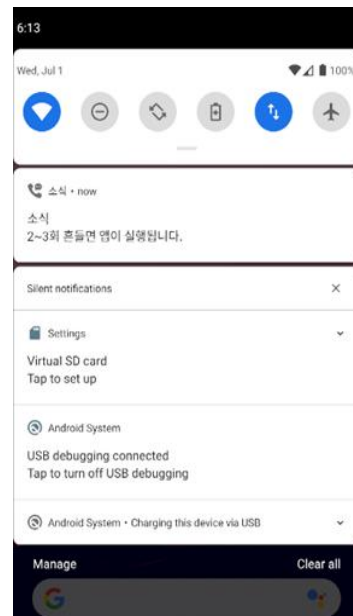


Fig. 5. User Interface

애플리케이션을 실행하면 STT를 위한 서비스를 수행하고 마이크로 음성을 입력받아 실시간으로 하단의 textView에 STT 내용을 출력한다. 그리고 문장의 끝을 확인하면 recyclerView[5]의 좌측 상단부터 저장된다. 하단의 editText를 통해 대화 내용을 입력하고 변환 버튼을 클릭하면 TTS 기능을 제공하고, recyclerView의 우측에 저장되어 출력된다. 상단 상태바를 확인하면 현재 가속도 센서가 동작 중인 서비스 알림을 확인할 수 있다. 청각 장애인들이 애플리케이션을 쉽게 사용할 수 있도록 가속도 센서를 활용하여 스마트폰을 2~3회 흔들었을 때 해당 애플리케이션을 실행할 수 있도록 구현한다. 이때, 안드로이드 API 26레벨 이후부터 백그라운드 서비스는 자원 정책에 의해 종료될 수 있으므로[6] 그림5의 Notification의 형태로 포그라운드 서비스로 구현한다.

IV. Implementation of Application for Hearing-Impairment

본 논문에서는 안드로이드 스튜디오를 이용한 Java 기반의 안드로이드 애플리케이션을 구현한다. Android Studio[7]에서 생성한 프로젝트 구조는 그림 6과 같다. 그림 6의 프로젝트에서 생성한 각각의 파일에 선언한 클래스와 각 클래스의 기능은 표 1과 같다.

Table 1. Function of Class

| Class | Function |
|------------------------|----------------------------------|
| MainActivity | Implementation of app |
| MessageDialog Fragment | Setting dialog object property |
| MyService | Setting sensor object property |
| SpeechService | Setting STT object property |
| VoiceRecorder | Setting Recorder object property |

MainActivity 클래스는 메인 액티비티를 실행한다. 그리고 android의 speech.tts를 이용하여 TTS 기능을 구현한다. MessageDialogFragment 클래스에서는 메시지를 화면에 출력해준다. MyService 클래스에서는 가속도센서를 이용하여 핸드폰이 흔들리는 것을 감지하여 애플리케이션을 실행하는 기능을 제공한다. SpeechService 클래스에서는 Google 클라우드 플랫폼의 API를 활용하여 STT 기능을 구현한다. VoiceRecorder 클래스에서는 사용자의 음성을 녹음하는 기능을 제공한다.

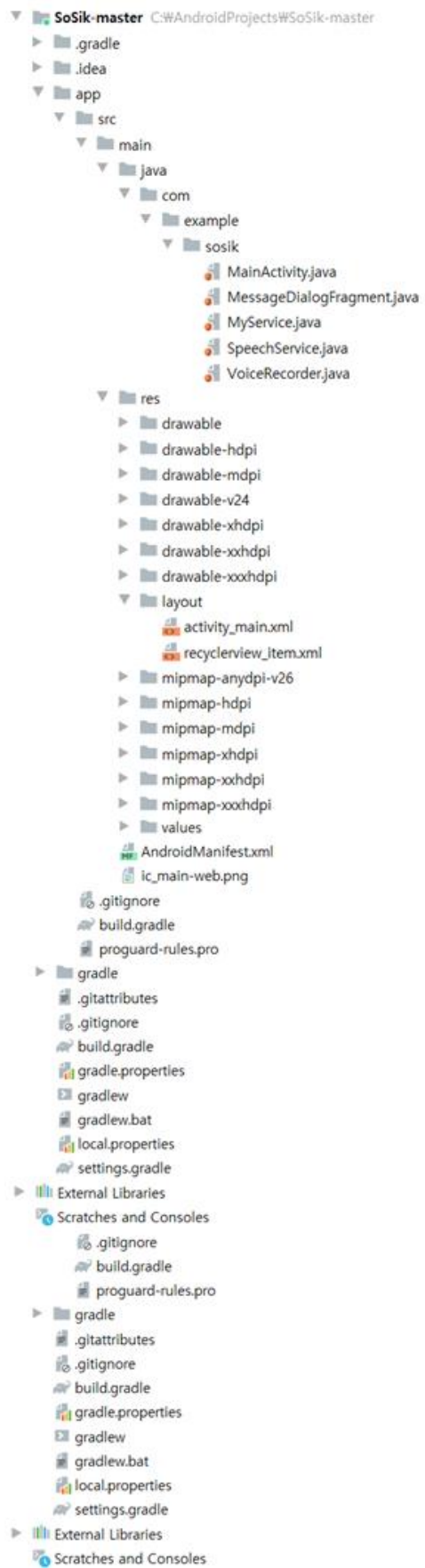


Fig. 6. Android Studio project architecture

MessageDialog 클래스는 대화 상자에 메시지를 표시한다. AppCompatActivity 클래스를 상속받는다. 해당 클래스에서 구현한 사용자 정의 메소드는 표 2와 같다.

Table 2. MessageDialogFragment() method

| Method | Function |
|------------------|--|
| newInstance() | creates a new instance of MessageDialogFragment and return it. |
| onCreateDialog() | Create a AlertDialog instance and return it. |

newInstance() 메소드는 MessageDialogFragment의 새로운 객체를 생성한다. MessageDialogFragment와 Bundle의 새로운 객체를 생성한 뒤 Bundle에 문자열을 삽입한 후 MessageDialogFragment 객체를 반환한다. onCreateDialog() 메소드는 AlertDialog 객체를 생성하여 리턴한다. AlertDialog가 사라질 때는 녹음 권한을 부여한다.

MyService 클래스에서 구현한 사용자 정의 메소드는 표 3과 같다.

Table 3. MyService method

| Method | Function |
|-----------------------|---|
| onStartCommand() | restart on the termination of the application |
| onSensorChanged() | motion detection |
| notificationChannel() | setting notification |
| startForeground() | run in the foreground |
| showIntent() | setting vrbation |

MyService 클래스에서 onSensorChange() 메소드는 shake_threshold 변수에 흔들리는 속도를 저장하고, shake_duration 함수에는 동작시간, shake_time에 횟수를 저장하여 1초동안 2번 800의 속도로 흔들면 핸드폰이 흔들리는 것으로 간주하여 움직임을 감지하는 기능을 제공한다. notificationChannel() 메소드에서는 NotificationManager 객체를 생성하여 알림 형태를 정한다. startForeground() 메소드는 알림 바에 포그라운드 형태로 애플리케이션이 계속 실행되도록 구현한다. onStartCommand() 메소드는 서비스 종료시 재실행하고 서비스가 종료되지 않았을 때는 각 객체를 생성하고 메소드를 실행하여 알림 형태로 포그라운드에 있는 것으로 간주 되도록 구현한다. SpeechService 클래스는 백그라운드의 작업을 처리하기 위해 Service 컴포넌트를 상속받는다. 이 클래스는 Google Speech API[9]와 오픈소스[10]를

이용하여 구현한다. Google Speech-to-Text API는 Google의 딥 러닝 신경망 알고리즘을 적용한 AI 기술로 지원되는 API를 사용하여 자동으로 음성을 인식하여 텍스트로 변환하는 기능을 제공한다.

VoiceRecorder 클래스에서 구현한 사용자 정의 메소드는 다음과 같다. 오픈소스를 활용하여 개발하였다. 해당 클래스에서 구현한 사용자 정의 메소드는 표 4와 같다.

Table 4. VoiceRecorder() method

| Method | Function |
|---------------------|---|
| start() | stop recording if it's currently ongoing |
| stop() | stop recording audio |
| dismiss() | dismisses the currently ongoing utterance |
| createAudioRecord() | creates a new AudioRecord. |
| isHearingVoice() | check if there is a voice |

start() 메소드는 음성 녹음을 시작하고, stop() 메소드는 음성 녹음을 중지한다. createAudioRecord() 메소드는 이후에 임시적으로 음성 데이터를 보관할 buffer[8]를 만들고 새로운 audioRecord 객체를 생성하는 메소드이다. isHearingVoice() 메소드는 사용자가 현재 말하고 있는 중인 지를 임의로 설정한 진동수를 기준으로 버퍼에 임시적으로 넣은 음성과 비교 판단하여 말하는 중일 경우 true를, 말하는 중이 아니라고 판단되면 false를 반환한다. 최종적으로 실행 결과 화면은 그림 7과 같다.

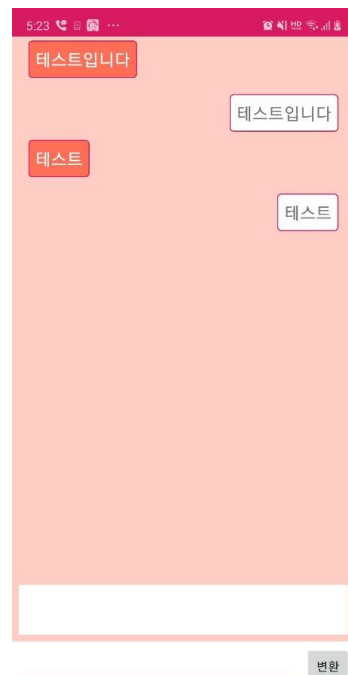


Fig. 7. Application Execution

V. Conclusions

본 논문에서는 Google 클라우드 플랫폼의 Speech API를 이용하여 청각 장애인들의 의사소통을 도와주는 모바일 애플리케이션을 설계하고 구현한다. 이 애플리케이션은 사용자가 말을 했을 때 음성을 인식하고 텍스트로 변환하여 화면에 출력한다. 또한 사용자가 문자를 입력 창에 타이핑한 후 변환 버튼을 클릭하면 이 애플리케이션은 텍스트를 음성으로 변환하여 입력받은 텍스트를 읽어준다. 이때 결과값은 recycler view를 이용해 채팅창 형식으로 표시하고, 음성으로 인식하여 텍스트로 반환한 데이터는 화면 왼쪽에 표시하고, 문자 입력을 받은 내용은 화면 오른쪽에 표시한다. 또한 애플리케이션의 실행을 편리하도록 가속도 센서를 이용해 휴대전화를 2~3번 흔들면 애플리케이션이 바로 실행될 수 있도록 구현하였다.

본 논문에서 개발한 애플리케이션은 청각 장애인들이 다른 사람과 의사소통을 할 때 영상통화로 수화를 이용하지 않고도 쉽게 대화할 수 있는 기능을 제공한다. 이 애플리케이션을 통하여 청각 장애인들이 전화 통화 중에 느끼는 불편함을 해소함으로써 청각 장애인들의 사회 참여 활성화에 기여할 것이다.

REFERENCES

- [1] Jae-il Kwon, "Study of Usage in Language of a hearing-impaired people" The National Institute of the Korean Language, Vol. 4, No. 3, pp. 40-57, Mar. 2009.
- [2] <http://www.mohw.go.kr/react/jb/sjb030301vw.jsp>
- [3] <https://rogervoice.com/en/>
- [4] <https://www.talovstudio.com/speakliz>
- [5] <https://developer.android.com/jetpack/androidx/releases/recyclerview>
- [6] Android developers, "Background Execution Limits" <https://developer.android.com/about/versions/oreo/background>
- [7] https://developer.android.com/studio?gclid=CjwKCAiAnvj9BRA4EiwAuUMDf2lTS3NGbD6v5EgidTOKz_2tt2wgJiSJRLaqu7TY4tjnEFyJfYnJhoCUJkQAvD_BwE&gclid=aw.ds
- [8] <https://developer.android.com/reference/java/nio/Buffer>
- [9] <https://cloud.google.com/speech-to-text>
- [10] <https://github.com/GoogleCloudPlatform/android-docs-samples/tree/master/speech/Speech/app/src/main/java/com/google/cloud/android/speech>

Authors



Woo-Lin Kim received the Associate Degree in Computer Science from Inha Technical College, Korea, in 2020. She received B.S. degree in Computer Science Engineering from Inha Technical College, Korea, in

2021. She is interested in Web programming, Mobile programming, and Machine Learning.



Hye-Won Ham received the Associate Degree in Computer Science from Inha Technical College, Korea, in 2020. She received B.S. degree in Computer Science Engineering from Inha Technical College,

Korea, in 2021. She is interested in Mobile programming, Internet of Things, and Physical computing.



Sang-Un Yun received the Associate Degree in Computer Science from Inha Technical College, Korea, in 2020. He received B.S. degree in Computer Science Engineering from Inha Technical College, Korea, in

2021. He is currently a graduate student in the Department of Applied Artificial Intelligence, Major in Bio Artificial Intelligence, Hanyang University, Korea. He is interested in Linux system, Cloud computing, and Internet of things, artificial intelligence.



Won Joo Lee received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Hanyang University, Korea, in 1989, 1991 and 2004, respectively. Dr. Lee joined the faculty of the Department of

Computer Science at Inha Technical College, Incheon, Korea, in 2008, where he has served as the Director of the Department of Computer Science. He is currently a Professor in the Department of Computer Science, Inha Technical College. He has also served as the Vice-president of The Korean Society of Computer Information. He is interested in parallel computing, internet and mobile computing, and cloud computing, data science, artificial intelligence.