The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 16, No. 5, pp.111-116, Oct. 31, 2016. pISSN 2289-0238, eISSN 2289-0246

http://dx.doi.org/10.7236/JIIBC.2016.16.5.111
JIIBC 2016-5-17

청각 장애인을 위한 모바일 통화 시스템 설계 및 구현

Design and Implementation of Mobile Communication System for Hearing- impaired Person

윤동희*, 김영웅**

Dong-Hee Yun*, Young-Ung Kim**

요 약 미래창조과학부의 정보격차 실태조사에 따르면 장애인의 스마트폰 보유율은 일반인의 1/3 수준에 머물러 있어 장애인의 정보접근성은 비장애인에 비해 현저히 떨어지는 실정이다. 본 논문은 청각장애인의 모바일 음성 통화를 보다 편리하게 사용할 수 있도록 도와주는 어플리케이션인 CallHelper를 개발하였다. CallHelper는 전화가 오면 자동으로 구동되어 상대방의 전화음성을 텍스트로 실시간 번역하여 모바일 화면에 출력하고, 상대방의 음성에서 감정을 추론하여 이모티콘으로 시각화해서 표시하며, 음성과 번역된 텍스트, 감정을 동시에 저장하여 추후 번역결과를 확인해 볼 수 있도록 하는 어플리케이션이다.

Abstract According to the Ministry of Science, ICT and Future Planning's survey of information gap, smartphone retention rate of disabled people stayed in one-third of non-disabled people, the situation is significantly less access to information for people with disabilities than non-disabled people. In this paper, we develop an application, CallHelper, that helps to be more convenient to use mobile voice calls to the auditory disabled people. CallHelper runs automatically when a call comes in, translates caller's voice to text output on the mobile screen, and displays the emotion reasoning from the caller's voice to visualize emoticons. It also saves voice, translated text, and emotion data that can be played back.

Key Words : CallHelper, Speech-to-Text, Emotion Recognition, Deep Learning

Ⅰ. 서 론

우리나라의 정보화는 세계적 수준에 있지만 정보화 취약 계층에 대한 사회적 배려는 부족한 실정이다. 정보화 사회에서 장애인에 대한 정보통신의 접근성이 보장되지 않는다면 장애인들은 더욱 더 사회로부터 소외받을 수밖에 없다.

청각장애인의 경우 주된 언어소통 수단은 수화이어서 음성을 통한 전화통화가 어려운 상황이다. 이를 해소하기 위해 다각적인 방안이 제시되고 있다. 손말이음센터에서는 청각장애인들을 대신해서 상대방과의 문자나 영상통화를 통한 수화를 중계해주는 서비스를 제공하고 있는데, 대표적인 서비스로는 청각장애인이 문자를 입력하면 중계사가 음성으로 상대방에게 전달하는 문자중계서비스, 청각장애인이 영상전화를 통해 수화를 전달하면 중계사가 음성으로 상대방에게 전달하는 영상중계서비스가 있다.

이 서비스의 문제점으로는 개인의 사사로운 통화정보

*준회원, 한성대학교 컴퓨터공학과
**정회원, 한성대학교 컴퓨터공학부(교신저자)
접수일자 : 2016년 7월 26일, 수정완료 : 2016년 9월 8일
게재확정일자 : 2016년 10월 7일

Received: 26 July, 2016 / Revised: 8 September, 2016 / Accepted: 7 October, 2016
**Corresponding Author: yukim@hansung.ac.kr
Dept. of Computer Engineering, Hansung University, Korea

가 제 3자(중계사)에게 노출되어 사생활침해 측면에서 이용의 한계가 존재하며, 중계를 통한 서비스임을 고려할 때 실시간 통화의 의미가 퇴색될 수 있다.

또한, 이동통신 사업자들은 청각장애인의 휴대폰 영상통화를 지원하기 위해 청각장애인을 위한 영상통화요금제를 출시하여 배려하고 있지만, 수화를 사용하는 청각장애인의 특성상 통화시간이 일반인들보다 현저히 길어 여전히 요금에 부담을 가지며, 양손을 사용해야 하는 수화의 특성상 영상통화 방법은 그 한계가 있다[1].

이를 해소하기 위해 본 논문은 청각장애인의 모바일 음성 통화를 보다 편리하게 사용할 수 있도록 도와주는 어플리케이션인 CallHelper를 개발하였다. CallHelper의 기능은 다음과 같다.

① 전화 통화 시 자동 실행 : 전화를 걸거나 받을 때 자동으로 어플리케이션이 실행되도록 개발하여 사용자가 전화 통화 전이나 후에 추가적인 조작을 할 필요가 없도록 한다.

② 상대방 음성 실시간 번역 : 상대방의 음성을 텍스트로 실시간 번역하여 어플리케이션 화면에 표시하도록 하여 청각장애인이 상대방의 내용을 파악할 수 있다.

③ 상대방 감정 실시간 추론 : 상대방의 음성에서 감정을 추론하여 이모티콘으로 시각화하여 화면에 표시하여 상대방의 현재 감정을 추측할 수 있다.

④ 음성과 텍스트 동시 저장 : 상대방의 음성과 번역된 텍스트, 감정을 동시에 저장하여 재생할 수 있도록 하여 추후 청각장애인이 일반인에게 부탁해 번역결과를 확인해 볼 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 CallHelper 시스템 구조에 대해 기술하고, 제 3장에서는 CallHelper 시스템 기능에 대해 기술한다. 제 4장에서는 CallHelper의 구현환경과 성능평가에 대해 기술하고, 끝으로 제 5장에서 결론 및 향후 연구과제에 대해 기술한다.

II. CallHelper 시스템 구조

1. 어플리케이션 구조

그림 1은 CallHelper 어플리케이션 구조도를 보여준다.

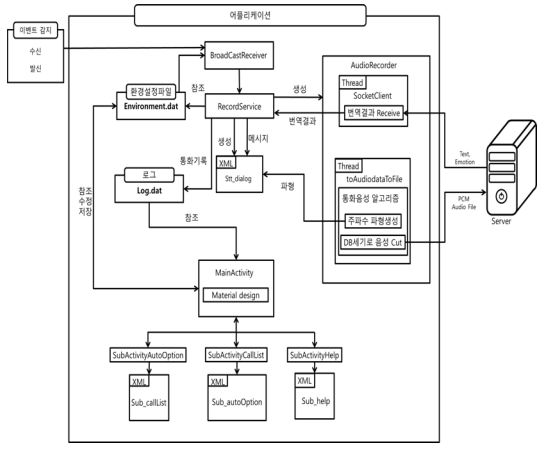


그림 1. CallHelper 어플리케이션 구조도
Fig. 1. CallHelper Application Architecture

어플이 탑재된 스마트폰에서 전화수신이 발생하면 BroadCastReceiver 모듈에서 이를 감지해 어플을 구동시킨다. 어플이 구동되면 음성을 텍스트로 번역하는 과정을 총괄하는 RecordService 클래스가 생성되어 Stt_dialog XML을 이용해 번역결과를 사용자 화면에 보여준다. RecordService 클래스는 AudioRecorder 클래스를 생성시켜 통화내역을 녹음하고, AudioRecord와 toAudiodataToFile Thread를 이용해 실시간 통화내용을 가져와 작업을 진행한다. 실시간 통화내용은 데시벨 강도를 분석하는 내부 알고리즘을 거쳐 알맞은 단위로 잘라 서버에 전송되며, 동시에 RealDoubleFFT 라이브러리를 이용해 해당 음성데이터 주파수 파형을 생성한다. 생성된 파형은 Handler를 통해 즉시 출력화면으로 이동되며 음성데이터에 따라 실시간으로 갱신된다.

서버에 전송하는 음성데이터의 단위는 16bit PCM 파일이며 서버는 이 음성데이터를 받아 번역결과와 감정분석결과를 텍스트 형태로 반환한다. 변역결과는 SocketClient Thread에서 수신함과 동시에 Handler를 이용하여 출력화면으로 전송한다. RecordService는 이 작업들을 관리하며 번역결과를 Log.dat에 저장하여 통화기록을 생성한다. MainActivity는 Material design을 통해 사용자에게 옵션을 선택하게 한다. 정의된 옵션정보는 Environment.dat 환경파일에 저장되며 이 환경파일은 어플 실행 시 자동으로 참조한다.

2. 서버 구조

그림 2는 CallHelper 서버 구조도를 나타낸다.

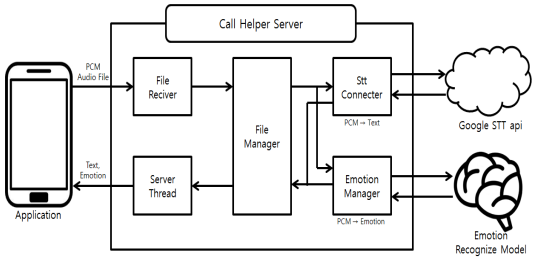


그림 2. CallHelper 서버 구조도
Fig. 2. CallHelper Server Architecture

어플리케이션에서 PCM 오디오 데이터를 socket 통신을 이용해 수신하며 파일정보들을 포함한 InputFile 클래스를 생성해 FileManager로 보낸다. FileManager에서는 PCM 데이터를 SttConnecter를 이용해 Google STT API[2] Server로 전송하고 텍스트로 번역된 결과를 json[3] 방식으로 받아온 후 최종 결과를 InputFile클래스에 저장한다. 이와 동시에 같은 PCM 데이터를 EmotionManager를 이용해 감정추론 딥러닝 모델에 전송하고 추론된 감정을 받아와 저장한다. 이 때 네 개의 감정 중 대표 값에 해당되는 감정을 저장하고, 특별히 대표 값이 없으면 "감정없음"으로 저장한다. 마지막으로 다시 Socket통신을 이용해 저장된 텍스트와 감정정보를 어플리케이션에 전송한다.

3. 감정추론 딥러닝 구조

그림 3은 통화자의 음성신호를 분석하여 감정을 추론하는 감정추론 딥러닝 구조를 나타낸다.

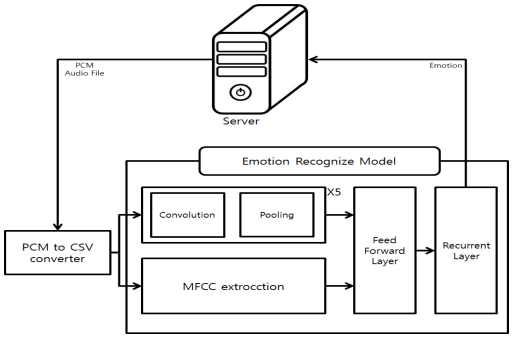


그림 3. 감정추론 딥러닝 구조도
Fig. 3. Emotion Recognition Architecture

감정을 추론하기 위한 딥러닝 모델은 합성곱 신경망(Convolutional Neural Networks: CNN)[4]과 순환 신경망(Recurrent Neural Network: RNN)[5] 모델을 혼합하여 생성하며, 주파수에서 음성의 특징을 잡아내기 위해 사람의 음성 인지 양상이 로그 스케일과 비슷한 멜 스케일을 따른다는 특성을 반영한 MFCC(Mel Frequency Cepstral Coefficients)[6] 추출법을 이용하였다. 수신한 음성 데이터를 일정 단위로 잘라 CNN 모델을 통해 추상화시키고 MFCC를 이용해 추출한 특징과 합쳐 RNN 모델을 통해 감정을 추론한다. 감정 추론은 대표적인 네 가지 감정(화남, 놀람, 기쁨, 슬픔)이 각각 확률로 추론하고, 이 값 중 제일 높은 값이 나머지 값의 합보다 더 크면 그 감정을 최종감정으로 추론하며, 아니면, 무감정으로 추론한다.

하나의 time-step에서 time-domain의 정보와 frequency-domain의 정보를 얻어 처리하는데, time-domain은 단순히 해당 time-step에서의 raw data를 사용할 수도 있지만 본 모델에서는 raw data에 convolution-pooling 연산을 통해 비선형성을 더한 후의 결과를 취한다. frequency-domain은 MFCC를 구해 얻을 수 있다. time-domain과 frequency-domain에서의 정보를 FeedForward Layer와 Recurrent Layer로 보내서 감정을 추론한다. Recurrent Layer는 이전 time-step의 정보를 기억할 수 있기 때문에 위 모델은 어떠한 길이의 음성 데이터도 처리할 수 있다.

Ⅲ. CallHelper 기능

이 장에서는 CallHelper에서 제공하는 기능들을 스마트폰 화면을 중심으로 기술한다. 그림 4는 CallHelper에서 제공하는 주요 기능들을 보여준다.

그림 4-가의 사용자 설정 기능은 다음과 같다.

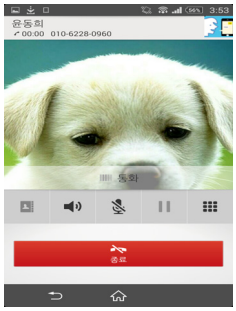
- 자동실행: 통화시 CallHelper 서비스를 자동 실행 유무 설정
- 감정분석: 번역시 감정을 표현해서 분석유무 설정
- 통화내용저장: 통화내용을 저장유무 설정
- 배경화면: 배경화면 설정

각각의 옵션들은 환경설정파일에 저장되며 이 파일은 어플이 동작할 때 자동으로 참조된다.

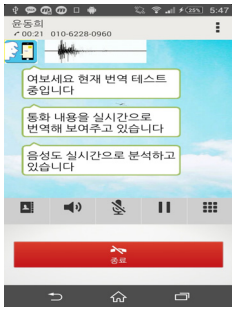
그림 4-나의 전화 송수신 기능은 그림 가의 사용자 설정화면에서 설정한 자동 실행기능 On/Off에 따라 어플리케이션이 자동으로 실행된다. On/Off 여부는 전화 이벤트가 발생 시 자동으로 환경설정파일을 참조하여 알아낸

다. 오른쪽 위에 작은 Call Helper 어플리케이션 로고 아이콘이 뜨게 되며 사용자는 이를 눌러 번역결과 창을 키우거나 줄일 수 있다.

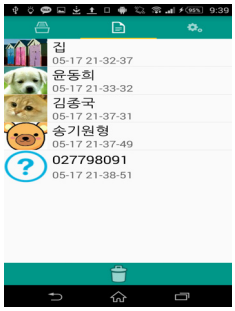
가. 사용자 설정화면



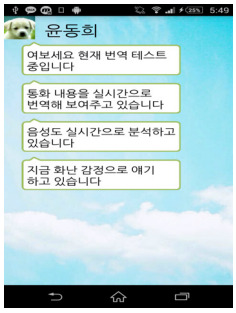
나. 전화 송수신화면



다. 번역 결과화면



라. 통화내용 기록화면



마. 통화내용 상세정보

그림 4. CallHelper 기능
Fig. 4. CallHelper Functions

그림 4-다의 통화중 변역결과 및 감정결과 화면은 음성을 텍스트로 번역한 결과를 보여주는 창이다. 제일 상단에는 Call Helper 어플리케이션 로고와 음성데이터 파형을 나타내는 창이 위치한다. 로고를 다시 누르게 되면 작은 아이콘상태로 창을 축소시키는 기능을 하며 파형을 그려주는 창은 상대방이 현재 말하고 있는지 아닌지를 판단할 수 있게 도와주는 기능을 제공한다. 번역결과는 말풍선에 텍스트로 시각화하여 나타나게 되며 감정분석 결과가 있는 경우 말풍선의 오른쪽에 감정분석결과가 이모티콘 형태로 나타나게 된다.

그림 4-라의 통화내용 기록화면은 현재 저장되어 있는 통화내용 기록을 사용자가 다시 확인할 수 있게 도와주는 탭이다. 설정옵션에서 통화내용 저장기능이 On되어 있는 경우 통화 중 발생한 번역내용과 통화음성이 저장되게 되는데 그 내용을 이곳에서 확인 할 수 있다. 리스트는 전화번호부에 저장된 상대방의 이미지와 전화번호, 통화가 발생한 날짜와 시간을 출력하여 보여주게 되고 각각의 리스트를 클릭하면 해당 통화에서 번역되었던 번역결과 화면을 보여준다.

그림 4-마의 통화내용 상세정보 화면은 그림 라의 통화내용 기록리스트를 누르게 되면 이동되는 화면이다. 해당 통화 중에 번역되었던 모든 번역결과들이 다시 말풍선에 찍혀 사용자에게 보여주며, 각각의 말풍선을 누르면 해당 음성을 다시 재생되게 된다. 사용자는 이 기능을 통해 조금 애매한 번역이 된 음성을 주변 지인에게 들려주어 번역결과를 확인할 수 있다.

IV. 구현환경 및 성능

그림 5는 CallHelper의 구현환경을 보여준다.

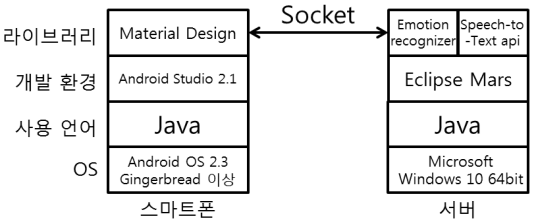


그림 5. 구현 환경
Fig. 5. Implementation Environments

CallHelper의 안드로이드 스마트폰용 어플리케이션은 Google사의 Android OS 2.3버전인 Gingerbread이상의 버전에서 구동된다. 개발언어로는 안드로이드용 Java언어를 사용하였고 Google에서 제공한 IntelliJ 기반의 Android 개발 툴 Android Studio[7] 2.3버전으로 개발을 진행하였다. 사용 라이브러리로는 Material Design 등이 있다.

CallHelper의 서버는 Microsoft사의 Windows 10 64bit버전에서 구동된다. 개발언어로는 Java를 사용하였

고, eclipse mars 4.2버전으로 구현하였다. 사용 라이브러리로는 직접 개발한 감정추론 딥러닝 라이브러리인 Emotion Recognizer와 Google에서 제공하는 STT API를 사용하였다. 그림 6은 음성을 텍스트로 변환 후 서버로 전송되는 정보를 보여준다. 전송받은 PCM파일을 Google STT API Server에 전송하여 JSON방식의 결과를 받아와 결과만 파싱하여 저장한다. 이와 동시에 같은 PCM파일과 감정정보 추론 모델을 이용해 음성 데이터의 감정을 추론한다. 이렇게 얻은 두 가지 정보를 다시 소켓통신을 이용해 Call Helper에게 전송한다. 위 작업을 통화 종료 시까지 반복한다.

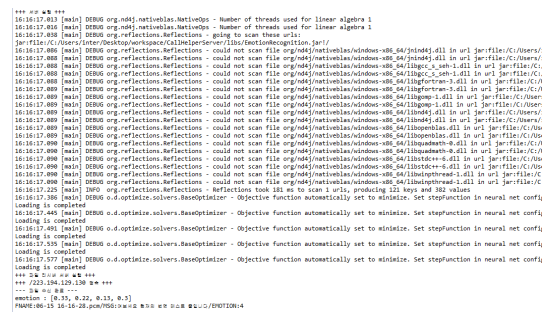


그림 6. 어플리케이션 동작 정보
Fig. 6. Information of Application operations

CallHelper의 성능으로는 크게 음성을 텍스트롤 변환하는 변환율과, 음성에서 감정을 인식하는 인식률로 구분된다.

표 1. 음성-텍스트 성능
Table 1. Performance of Speech-to-Text

단어수	성공률	오번역		번역불가
		추론가능	추론불가	
3단어 이하	70	15	13	2
4~6단어	63	19	17	1
7단어 이상	59	20	21	

표 1에서 보는 바와 같은 번역 성공률을 보이기 때문에 의미추론과 길이 모두를 만족하는 6단어 이하의 단위로 잘릴 수 있도록 구현하였다. 실제 소음이 있는 공간에서 CallHelper를 이용해 통화시 약 75%정도의 정확도로 번역하였다.

표 2. 감정인식 성능
Table 2. Performance of Emotion Recognition

	화남	놀람	기쁨	슬픔
화남	75%		10%	15%
놀람	20%	50%	15%	15%
기쁨	58%	8%	21%	12%
슬픔	38%	2%	5%	55%
평균	54%			

음성으로부터의 감정 인식은 딥러닝 이전 Hidden Markov Model과 결합한 Gaussian Mixture Model[8]이 대표적인 방법이 있다. 하지만 딥러닝이 음성 인식 분야에서 돌파구가 된 이후로 음성으로부터의 감정 인식 분야는 큰 변화를 겪고 있다. 단순히 음성으로부터의 감정 인식은 매우 어려운 문제이며, 이 문제의 해결을 위해서 어떤 특징이 효과적인지조차 아직까지 명백하지 않다. 그러나 딥러닝은 다른 머신러닝 알고리즘과는 다르게 직접 특징을 추출하지 않아도 되는 장점이 있다.

표 2는 음성으로부터 감정을 추출한 성능을 보여준다. 전체적으로는 54% 감정추론 정확도를 보여주나, 이 중 화남 감정은 비교적 잘 추출하나, 기쁨 감정은 매우 저조하게 나타났다. 이는 음성정보만으로 감정을 추출하기에는 그 한계를 보여준다.

V. 결 론

본 논문은 청각장애인의 경우 주된 언어소통 수단은 수화이어서 음성을 통한 전화통화가 어려운 상황을 해소하기 위해 모바일 음성 통화를 보다 편리하게 사용할 수 있도록 도와주는 어플리케이션인 CallHelper를 개발하였다. CallHelper는 크게 상대방 전화음성을 실시간으로 텍스트로 변환하는 기능과, 전화음성으로부터 상대방의 감정을 실시간으로 추론하는 기능이 있다.

전화음성을 텍스트로 변환하는 성능은 소음이 어느 정도 있는 공간에서 평균 75%의 정확도를 보여 사용하기에 큰 무리는 없으나, 음성으로부터의 감정추론 성능은 평균 54%의 정확도를 보여 단순히 음성만으로 감정을 추론하기에는 한계를 보였다.

향후 연구과제로는 CallHelper의 성능 개선을 위해 번역의 성공률을 높이고 오번역을 줄이기 위한 연구와, 감정추론 정확도를 향상시키기 위해 추론 모델을 개선시키는 연구가 진행될 예정이다.

References

[1] S. S. Yang, S. C. Park, “Design and Implementation of CTI System for Hearing Impaired People in Mobile Environment”, The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 13, No. 6, pp.47-54, 2013.

[2] https://cloud.google.com/speech/

[3] http://www.json.org/json-ko.html

[4] A. Ossama, A. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu,“Convolutional neural networks for speech recognition”, IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), Vol. 22 No. 10, pp.1533-1545, 2014.

[5] A. Graves, N. Jaitly, “Towards End-to-End Speech Recognition with Recurrent Neural Networks”, Proceedings of the 31st International Conference on Machine Learning, 2014.

[6] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[7] https://developer.android.com/studio/index.html

[8] Z. Chen, T. Ellis, “A self-adaptive Gaussian mixture model”, Computer Vision and Image Understanding 122, pp.35-46, Elsevier, 2014.

저자 소개

윤 동 희(준회원)

- 2010년 ~ 현재 : 한성대학교 컴퓨터공학과

<주관심분야 : 모바일 시스템, 기계 학습>

김 영 웅(정회원)

- 1993년 : KAIST 전산학과 박사
- 1984년 ~ 1997년 : KT 통신망 연구소
- 1997년 ~ 현재 : 한성대학교 컴퓨터공학부 교수

<주관심분야 : 데이터모델링, 소프트웨어 신뢰도, 소프트웨어 설계>

※ 본 연구는 한성대학교 교내연구비 지원 과제임.