

청각장애인을 위한 음성인식 기반 메시지 전송 시스템

김성진¹ · 조경우¹ · 오창현^{1*}

Speech Recognition based Message Transmission System for the Hearing Impaired Persons

Sung-jin Kim¹ · Kyoung-woo Cho¹ · Chang-heon Oh^{1*}

¹Department of Electrical, Electronics & Communication Engineering, Korea University of Technology and Education(KOREATECH), Cheonan 31253, Korea

요 약

음성인식 서비스는 청각장애인에게 화자의 음성을 텍스트로 변환하여 시각화함으로써 의사소통의 보조적인 수단으로 사용되고 있다. 하지만 강의실 및 회의실과 같은 개방된 환경에서는 다수의 청각장애인에게 음성인식 서비스를 제공하기 힘들다. 이를 위해 주변 환경에 따라 음성 인식 서비스를 효율적으로 제공하기 위한 방법이 필요하다. 본 논문에서는 화자의 음성을 인식하여 변환된 텍스트를 다수의 청각장애인에게 메시지로 전달하는 시스템을 제안한다. 제안하는 시스템은 다수의 사용자에게 동시에 메시지를 전달하기 위해 MQTT 프로토콜을 사용한다. MQTT 프로토콜의 QoS level 설정에 따른 제안 시스템의 서비스 지연을 확인하기 위해 중단 간 지연을 측정하였다. 측정 결과 가장 신뢰성이 높은 QoS level 2와 0간의 지연이 111ms로 대화 인식에 큰 영향을 끼치지 않음을 확인하였다.

ABSTRACT

The speech recognition service is used as an ancillary means of communication by converting and visualizing the speaker's voice into text to the hearing impaired persons. However, in open environments such as classrooms and conference rooms it is difficult to provide speech recognition service to many hearing impaired persons. For this, a method is needed to efficiently provide it according to the surrounding environment. In this paper, we propose a system that recognizes the speaker's voice and transmits the converted text to many hearing impaired persons as messages. The proposed system uses the MQTT protocol to deliver messages to many users at the same time. The end-to-end delay was measured to confirm the service delay of the proposed system according to the QoS level setting of the MQTT protocol. As a result of the measurement, the delay between the most reliable QoS level 2 and 0 is 111ms, confirming that it does not have a great influence on conversation recognition.

키워드 : 음성인식, 메시지 전송 시스템, Speech-to-Text, MQTT(Message Queue Telemetry Transport), 청각장애인

Keyword : speech recognition, message transmission system, speech-to-text, MQTT(Message Queue Telemetry Transport), hearing-impaired persons

Received 24 August 2018, Revised 13 September 2018, Accepted 24 September 2018

* Corresponding Author Chang-heon Oh(E-mail:choh@koreatech.ac.kr, Tel:+82-41-560-1187)

Department of Electrical, Electronics & Communication Engineering, Korea University of Technology and Education, Cheonan, 31253, Korea

Open Access <http://doi.org/10.6109/jkiice.2018.22.12.1604>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

청각장애인을 위해 보청기 및 인공와우와 같은 장치, 의사소통 보조 서비스 등의 연구가 진행되고 있다 [1-5]. 이러한 서비스 중 음성인식 서비스의 경우 화자의 음성을 텍스트로 시각화하여 청각장애인의 의사소통 보조 수단으로 사용할 수 있다. 음성인식 서비스는 대부분 스마트폰을 활용하여 스마트폰에 내장된 마이크로 화자의 음성을 인식한다. 하지만 강의실 및 회의실과 같은 개방된 환경에서 다수의 청각장애인에게 음성 인식 서비스를 제공하기는 어렵다 [6-9]. 청각장애인에게 음성 인식 서비스를 제공하기 위해서는 주변 환경을 고려하고, 효율적으로 청각장애인들에게 음성 인식 결과를 전달하는 방법에 대한 연구가 필요하다.

본 논문에서는 청각장애인을 위한 음성인식 기반 메시지 전송 시스템을 제안한다. 제안하는 시스템은 강의실 및 회의실에서 화자와 가까운 마이크에서 인식된 음성을 인식하여 텍스트로 변환한다. 변환된 텍스트는 청각장애인에게 메시지로 전달된다. 이 때 강의실 및 회의실과 같은 개방된 환경에서 다수의 청각장애인이 동시에 텍스트를 전달받을 수 있도록 MQTT(Message Queue Telemetry Transport) 프로토콜을 이용하여 메시지를 전송한다. MQTT는 publish/subscribe 모델 구조를 통해 다수의 장치에서 동일한 토픽을 구독하는 경우 동시에 메시지를 수신할 수 있다 [10]. 메시지 전송의 신뢰성을 보장하기 위해 QoS level 0, 1, 2의 설정이 가능하며, 높은 level이 적용될수록 메시지 전달의 신뢰성이 높다. 하지만 높은 QoS level의 경우 신뢰성을 보장하기 위해 4-way handshake 방식을 사용하여 메시지 전달간의 지연이 발생할 수 있어 QoS level에 따른 중단 간 지연 측정을 통해 제안 시스템의 MQTT 적용을 검토한다.

본 논문의 구성은 다음과 같다. 2장에서는 청각장애인을 위한 의사소통 보조 서비스 기존 연구들과 MQTT 프로토콜을 설명한다. 3장에서는 제안하는 음성인식 기반 메시지 전송 시스템에 대해 설명한다. 4장에서는 제안하는 시스템을 실험을 통해 평가하고, 5장에서 결론을 맺는다.

II. 관련 연구

2.1. 청각장애인의 의사소통을 위한 연구

T. Aujeszyk의 연구에서는 수화 인식을 위한 프레임워크를 제안하였다. 제안한 프레임워크는 두 가지 기능을 수행한다. 첫 번째는 아랍어의 수화 인식을 실시간 통신으로 지원한다. 두 번째는 아바타 애니메이션을 통해 청각장애인에게 실시간으로 피드백을 제공한다 [1]. E. W. Healy의 연구에서는 보청기 및 인공와우와 같은 장치에서 주변 소음을 완화하기 위한 방법을 제안하였다. 제안하는 방법은 단일 마이크 기반 IBM(Ideal Binary Mask)를 사용하여 음성 신호를 분리한다. 분리된 음성 신호는 잡음을 추출하기 위해 DNN(Deep Neural Network)를 이용하여 학습 및 훈련하였다 [2]. R. Akmeliawati의 연구에서는 청각장애인을 위한 자동 수화 통역 시스템을 제안하였다. 제안하는 시스템은 비전 기반 접근 방식을 사용하여 손 모양, 움직임, 위치로부터 다양한 특징 벡터를 추출한다. 인식 단계에서 해당 특징 벡터를 HMM(Hidden Markov Model)를 통해 학습한 뒤, 학습된 모델을 통해 실시간 수화 인식 시스템을 구현하였다 [3]. D. Watanabe의 연구에서는 청각장애인의 의사소통을 보조하는 스마트 안경을 제안하였다. 제안한 안경은 4개의 마이크로 음원의 방향 및 각도를 확인하여 잡음이 많은 환경에서 음성인식의 정확도를 향상시켰다 [4]. A. Chern의 연구에서는 스마트폰 보청기인 ‘SmartHear’를 제안하였다. ‘SmartHear’은 주변 소음 및 잔향을 줄이기 위해 청취자의 귀에 직접 화자의 음성을 스마트폰의 블루투스 및 Wi-Fi(Wireless-Fidelity)를 통해 전송한다. 또한 시각적 보조를 위해 전달받은 음성을 텍스트로 변환하는 기능을 구현하였다 [6]. C. K. A. Reddy의 연구에서는 보청기로 전달되는 주변 잡음을 줄이기 위한 방법을 제안하였다. 제안하는 방법은 보청기 장치를 대신하여 스마트폰에서 잡음 및 음성 왜곡에 대한 복잡한 연산 처리를 수행한다 [9].

2.2. MQTT

MQTT는 1999년 IBM사에서 설계한 TCP/IP 기반 양방향 메시지 통신 규약이다. MQTT는 토픽 기반 publish/subscribe 모델을 채택하여 메시지 전송의 실시간성과 많은 노드의 연결을 보장한다. 이러한 특징으로 Facebook Messenger에서 활용되고 있으며, 장치의 원격제어, 실

시간 모니터링 등의 서비스에 적용하기 위한 연구가 진행 중이다 [10-12]. 그림 1은 publish/subscribe 모델이다.

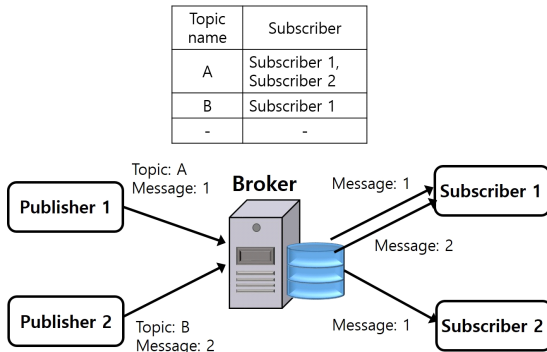


Fig. 1 Publish/Subscribe Model

Publish/Subscribe 모델은 publisher, subscriber, broker가 있다. Broker는 subscriber가 구독한 토픽을 저장한다. Publisher에서 토픽으로 메시지를 전송하는 경우 해당 토픽을 구독한 subscriber에게 메시지를 전달한다. 또한, MQTT는 3 가지의 QoS(Quality of Service) level을 지원하여 메시지 전달의 신뢰성을 보장한다. QoS level 0은 메시지를 한 번만 전달하여 신뢰성이 낮다. QoS level 1은 메시지가 전달되는 것을 응답 메시지로 확인한다. 하지만 응답 메시지가 유실되면 메시지가 중복으로 전달될 수 있다. QoS level 2는 4-way handshake 방식으로 메시지를 명확하게 전달한다.

III. 청각장애인을 위한 음성인식 기반 메시지 전송 시스템

3.1. 청각장애인을 위한 음성인식 기반 메시지 전송 시스템

청각장애인을 위한 음성 기반 메시지 전송 시스템의 서비스 시나리오는 그림 2와 같다. 화자는 마이크를 통해 음성을 전달한다. 전달된 음성은 speech-to-text를 통해 텍스트로 변환된다. 변환된 텍스트는 MQTT를 통해 해당 토픽을 구독한 청각장애인의 스마트 디바이스로 전달된다. 이를 통해 청각장애인은 화자의 음성을 텍스트 형태로 변환된 메시지로 스마트 디바이스를 통해 확인 가능하다.

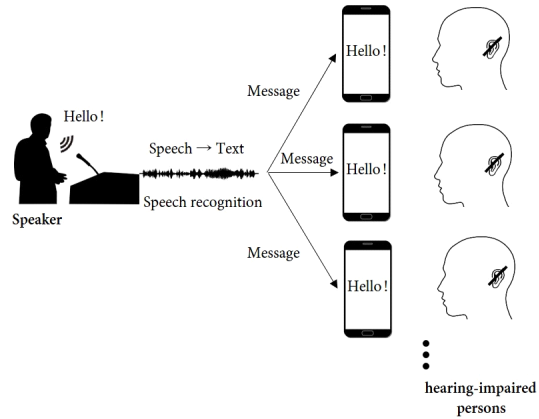


Fig. 2 Speech Recognition based Message Transmission Scenario for the Hearing Impaired Persons

3.2. 시스템 설계 및 구현

그림 3은 제안하는 시스템에 대한 구성도이다. 음성 인식 시스템은 raspberry pi3을 통해 구현하였으며, 화자의 음성을 전달받기 위해 마이크를 부착하였다. 음성인식을 위해 google cloud 서비스 중 하나인 google speech-to-text를 사용한다. Google speech-to-text는 신경망 모델을 통해 오디오를 텍스트로 변환한다. 또한, 고급신호처리 및 잡음 제거 등의 오디오 전처리 과정을 수행하여 음성 인식률을 높인다 [13]. Broker는 mosquitto 1.5.1을 사용하여 구현하였으며, MQTT 기능 구현은 paho 라이브러리를 통해 구현하였다 [14],[15].

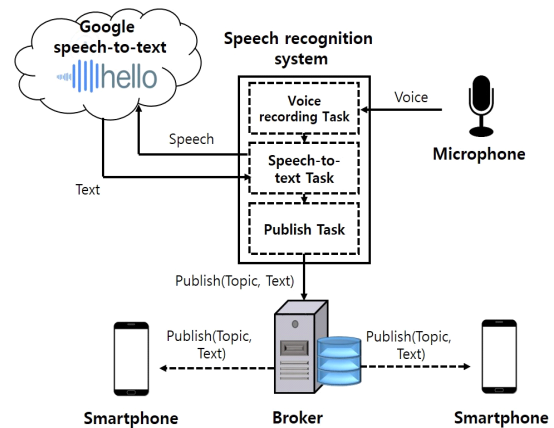


Fig. 3 System Configuration

음성인식 시스템은 voice recording task, speech-to-text task, publish task로 구성된다. 그림 4는 음성인식 시스템 동작에 대한 순서도이다. 화자는 음성인식을 통해 변환된 텍스트를 broker로 전달하기 위해 토픽의 이름을 설정한다. 마이크에 인식된 화자의 음성은 voice recording task를 통해 5초 간격의 wav파일로 저장된다. 저장된 wav 파일은 speech-to-text task을 통해 google speech-to-text로 전달된다. Google speech-to-text는 wav 파일에 저장된 음성을 인식하고, 텍스트로 변환된 메시지를 응답메시지로 전송한다. Publish task는 speech-to-text task로부터 받은 텍스트의 유무를 판단하고, 텍스트가 반환된 상태이면 broker로 전달한다.

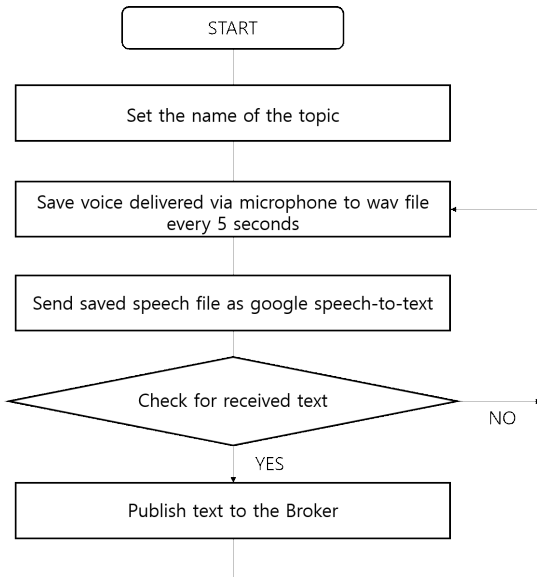


Fig. 4 Process Flow Chart of Speech Recognition System

Broker로부터 전달받은 메시지를 시각화하기 위해 안드로이드 애플리케이션을 구현하였다. 그림 5는 구현된 안드로이드 애플리케이션의 화면이다. 사용자는 음성인식 시스템에서 등록된 토픽을 구독하기 위해 입력 칸에 토픽이름을 작성 후 구독버튼을 클릭한다. 구독버튼을 클릭한 경우 broker로 작성한 토픽이름에 해당하는 토픽 구독요청을 전송한다. 구독요청이 완료되면 'Class_name'이 해당 토픽 이름으로 변경되며, 이를 통해 broker에 해당 토픽에 대한 메시지가 게시되면 사용자 애플리케이션으로 전달된다. 전달된 메시지는 메시

지 수신시간과 함께 텍스트뷰에 표시된다.



Fig. 5 Android Applications

IV. 실험 및 검토

그림 6은 제안하는 시스템을 검증하기 위한 실험 환경이다.

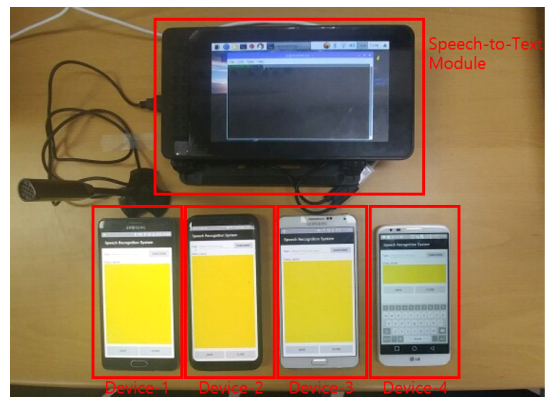


Fig. 6 Experiment Environment

실험에 사용되는 스마트폰 device 1, 2, 3, 4는 각각 SM-N910K, SM-G950N, SM-N900K, LG-F320S를 사용하였다. 음성인식 시스템에 부착된 마이크는 elcrow

사의 RPA02198R 제품을 사용하였다. 스마트폰과 speech-to-text 모듈은 Wi-Fi를 통해 네트워크와 연결하였다.

제안한 시스템에서 MQTT 적용에 대한 검토를 위해 4대의 사용자 어플리케이션이 다른 시점에 토픽을 구독하였을 때의 메시지 응답과 QoS level 설정에 따른 중단 간 지연을 통해 서비스 지연을 확인하였다. QoS level의 경우 화자의 speech-to-text module과 청자 device들의 QoS level을 동일하게 설정하였다. 메시지 전달을 위한 토픽은 'test_class'로 설정하였으며, 마이크로 음성을 인식시킨 뒤 청자 device들의 구독 시점을 달리하였다.

그림 7은 각 청자 device들의 스마트폰 어플리케이션 화면이다. 본 실험에서의 speech-to-text module과 청자 device들의 QoS level은 0으로 설정하였다. (a)와 (b) device는 동일한 시간에 구독 요청을 수행하였으며, (c)와 (d) 디바이스의 경우 (a), (b) 디바이스와 다른 시간에 구독 요청을 수행하였다. 실험 결과 각 device의 토픽 구독 시점을 기준으로 인식된 음성이 텍스트로 변환되어 청자 device로 전달됨을 확인할 수 있다. 또한 03:50:11 시점 이후의 메시지를 관찰하면 각 device로 하나의 메시지가 전달되는 과정에서 최대 1초의 차이를 가지며 메시지가 전달된 것을 확인할 수 있다.

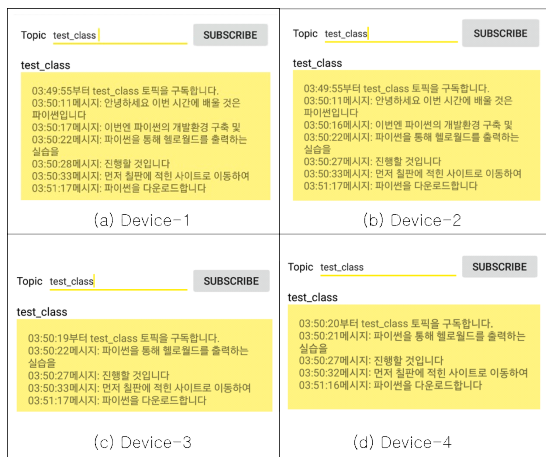


Fig. 7 Screen of Each Smartphone

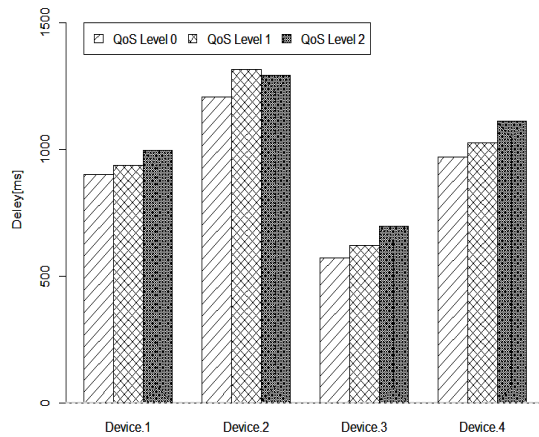


Fig. 8 End-to-End Delay According to QoS Level of Each Smartphone

제안한 시스템의 서비스 지연을 확인하기 위해 QoS level에 따른 중단 간 지연을 확인하였다. 그림 8은 speech-to-text module과 청자 device들의 QoS level에 따른 중단 간 지연을 나타낸다. QoS level이 낮을수록 메시지 전달 과정이 단순하여 중단 간 지연이 낮음을 확인할 수 있다. 각 스마트폰에 따른 QoS level 0, 1, 2에 대한 평균 중단 간 지연을 측정하였을 경우 level 2는 1보다 62ms 높게 측정되었다. 또한, QoS level 1은 0보다 49ms 높게 측정되었다. 가장 신뢰성이 높은 level 2의 경우 level 0보다 111ms의 지연이 발생되나, 대화 인식에 큰 영향을 끼치는 정도의 지연은 발생되지 않았다. 따라서 제안하는 시스템을 적용할 때 MQTT의 QoS level 설정에 따른 서비스 지연으로 인한 불편함은 낮을 것으로 판단된다.

V. 결론

본 논문에서는 강의실 및 회의실에서 청각장애인들의 원활한 의사소통을 지원하기 위해 음성인식 기반 메시지 전송 시스템을 제안하였다. 제안한 시스템은 화자와 가까운 마이크를 통해 인식된 음성을 텍스트로 변환한다. 변환된 텍스트는 다수의 청각장애인에게 동시다발적으로 메시지 전송이 가능하도록 MQTT 프로토콜을 통해 스마트폰으로 전달된다. Speech-to-text를 통해 변환된 텍스트는 broker로 전달되어 해당하는 토픽 명

에 게시되며, 해당 토픽을 구독한 다수의 청각장애인에게 텍스트 메시지 형태로 전달된다. 제안하는 시스템의 검증을 위해 마이크로 음성을 인식시켜 서로 다른 시점에 토픽을 구독하였을 때의 메시지 응답을 확인하였다. 실험 결과 구독 시점에 따라 화자의 인식된 음성이 각 스마트폰을 통해 메시지로 변환되어 전달됨을 확인하였다. 또한, QoS level 설정에 따른 중단 간 지연을 측정하여 제안한 시스템의 서비스 지연을 확인하였다. 가장 신뢰성이 높은 QoS level 2와 0간의 지연은 111ms로 나타났다. 대화 인식에 큰 영향을 끼치는 정도의 지연을 발생되지 않았다. 향후 연구로는 제안하는 시스템의 음성인식을 향상을 위해 외부 잡음 제거에 대한 연구를 진행할 계획이다. 또한 실질적인 서비스 운용이 가능하도록 사용자 어플리케이션의 인터페이스를 설계할 예정이다.

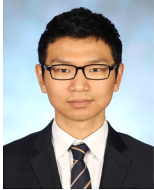
REFERENCES

- [1] T. Aujeszyk, and M. Eid, "A gesture recognition architecture for arabic sign language communication system," *Multimedia Tools and Applications*, vol. 75, no. 14, pp. 8493-8511, Jul. 2016.
- [2] E. W. Healy, and S. E. Yoho, "Difficulty understanding speech in noise by the hearing impaired: underlying causes and technological solutions," in *Proceedings of the 38th Annual International Conference of The IEEE in Medicine and Biology Society*, pp. 89-92, Orlando: Florida, 2016.
- [3] R. Akmeliawati, D. Bailey, S. Bilal, S. Demidenko, N. Gamage, S. Khan, Y. C. Kuang, M. Ooi, and G. S. Gupta, "Assistive technology for relieving communication barrier between hearing/speech impaired and hearing people," *The Journal of Engineering*, vol. 2014, no. 6, pp. 312-323, Jun. 2014.
- [4] D. Watanabe, Y. Takeuchi, T. Matsumoto, H. Kudo, and N. Ohnishi, "Communication support system of smart glasses for the hearing impaired," in *Proceedings of the 16th International Conference on Computers Helping People with Special Needs*, Linz: Austria, pp. 225-232, Jul. 2018.
- [5] S. E. Han, S. A. Kim, and G. H. Hwang, "E-book to sign-language translation program based on morpheme analysis," *Journal of Korea Institute of Information and Communication Engineering*, vol. 21, no. 2, pp. 461-467, 2017.
- [6] A. Chern, Y. H. Lai, Y. P. Chang, Y. Tsao, R. Y. Chang, and H. W. Chang, "A smartphone-based multi-functional hearing assistive system to facilitate speech recognition in the classroom," *IEEE Access*, vol. 5, pp. 10339-10351, Jun. 2017.
- [7] P. Patil, and J. Prajapat, "Implementation of a real time communication system for deaf people using internet of things," in *Proceedings of the International Conference on Trends in Electronics and Informatics*, Tirunelveli: India, pp. 313-316, May 2017.
- [8] J. Ming, and D. Crookes, "Speech enhancement based on full-sentence correlation and clean speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 3, pp. 531-543, Mar. 2017.
- [9] C. K. A. Reddy, N. Shankar, G. S. Bhat, R. Charan, and I. Panahi, "An individualized super-gaussian single microphone speech enhancement for hearing aid users with smartphone as an assistive device," *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1601-1605, Nov. 2017.
- [10] OASIS Standard, MQTT version 3.1.1[Internet], Available: <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html>.
- [11] Y. T. Lee, W. H. Hsiao, C. M. Huang, and S. C. T. Chou, "An integrated cloud-based smart home management system with community hierarchy," *IEEE Transactions on Consumer Electronics*, vol. 62, no. 1, pp. 1-9, Feb. 2016.
- [12] S. H. Kim, D. H. Kim, H. S. Oh, H. S. Jeon, and H. J. Park, "The data collection solution based on MQTT for stable IoT platforms," *Journal of Korea Institute of Information and Communication Engineering*, vol. 20, no. 4, pp. 728-738, Apr. 2016.
- [13] Google Cloud, Cloud Speech-to-Text[Internet], Available: <https://cloud.google.com/speech-to-text/>.
- [14] Mosquitto, Mosquitto[Internet], Available: <http://mosquitto.org/>.
- [15] eclipse paho, paho[Internet], Available: <http://eclipse.org/paho/>.



김성진(Sung-Jin Kim)

2016년 3월 ~ 현재 한국기술교육대학교 전기전자통신공학과 석사과정
2016년 2월 나사렛대학교 정보통신학과 이학사
※관심분야 : Internet of Things, Speech recognition, Machine Learning, Network



조경우(Kyoung-Woo Cho)

2015년 3월 ~ 현재 한국기술교육대학교 전기전자통신공학과 박사과정
2013년 3월 ~ 2015년 2월 한국기술교육대학교 전기전자통신공학과 공학석사
2013년 2월 공주대학교 전기전자제어공학부 전자공학나노정보공학전공 전자공학트랙 공학사
※관심분야 : Internet of Things, Machine Learning, Wireless Sensor N/W



오창현(Chang-Heon Oh)

1988년 2월 한국항공대학교 항공통신공학과 졸업 공학사
1990년 2월 한국항공대학교 항공통신정보공학과 졸업 공학석사
1996년 2월 한국항공대학교 항공전자공학과 졸업 공학박사
1990년 2월 ~ 1993년 8월 한진전자(주) 기술연구소 선임연구원
1993년 10월 ~ 1999년 2월 삼성전자(주) CDMA 개발팀 선임연구원
1999년 2월 ~ 현재 한국기술교육대학교 전기전자통신공학과 교수
2018년 1월 ~ 현재 한국정보통신학회 회장
※관심분야 : 무선/이동통신, IoT, 기계학습 기반 통신시스템