

# 음성 분류 인공지능망을 활용한 자폐아 치료용 로봇의 지능화 동작 연구

## Motion Study of Treatment Robot for Autistic Children Using Speech Data Classification Based on Artificial Neural Network

이진규\*, 이보희\*

Jin-Gyu Lee\*, Bo-Hee Lee\*

### Abstract

Currently, the prevalence of autism spectrum disorders in children is reported to be higher and shows various types of disorders. In particular, they are having difficulty in communication due to communication impairment in the area of social communication and need to be improved through training. Thus, this study proposes a method of acquiring voice information through a microphone mounted on a robot designed through preliminary research and using this information to make intelligent motions. An ANN(Artificial Neural Network) was used to classify the speech data into robot motions, and we tried to improve the accuracy by combining the Recurrent Neural Network based on Convolutional Neural Network. The preprocessing of input speech data was analyzed using MFCC(Mel-Frequency Cepstral Coefficient), and the motion of the robot was estimated using various data normalization and neural network optimization techniques. In addition, the designed ANN showed a high accuracy by conducting an experiment comparing the accuracy with the existing architecture and the method of human intervention. In order to design robot motions with higher accuracy in the future and to apply them in the treatment and education environment of children with autism.

### 요약

현재 아이들의 자폐스펙트럼장애 유병률이 한층 더 높게 보고되고 있으며 다양한 형태의 장애 징후를 보이고 있다. 특히 이들은 사회적 의사소통 영역에서 의사소통장애로 인한 대화에 어려움을 겪고 있으며 이를 훈련을 통해 개선 시킬 필요가 대두된다. 이를 위해 본 연구에서는 사전 연구를 통해 설계된 로봇에 장착된 마이크를 통해 음성 정보를 취득하고 이러한 정보를 이용하여 지능적인 동작을 만드는 방식을 제안한다. 음성 정보를 로봇 동작으로 분류하기 위해 인공지능망을 이용하였으며 여러 신경망 기법중 합성곱 방식을 기본으로 한 순환신경망을 결합하여 정확도를 향상시키려고 하였다. 입력 음성 데이터의 전처리는 MFCC를 이용하여 분석하였으며 여러 데이터 정규화 및 인공지능망 최적화 기법을 활용하여 로봇의 동작을 추정하였다. 아울러 설계된 인공지능망은 기존에 사용한 구조 및 사람이 개입하여 분석하는 방법과의 정확도 비교 실험을 진행하여 분석 결과가 높은 정확도를 나타냈다. 향후 보다 높은 정확도를 가질 수 있는 로봇 동작을 설계하여 실제의 자폐아 치료 및 교육 환경에서 적용할 수 있기 위하여 다양한 형태의 데이터를 수집하고 효율적으로 전처리하는 방식에 대한 연구가 요구된다.

*Key words : Autism Spectrum Disorders, Autistic Children Training, MFCC, Convolutional Neural Network, Speech Data Classification*

\* Dept. of Electrical Engineering, Semyung University

★ Corresponding author

E-mail : bhlee420@nate.com, Tel : +82-43-649-1305

※ Acknowledgment

This paper was supported by the Semyung University Research Grant of 2019.

Manuscript received Dec. 11, 2019; revised Dec. 25, 2019; accepted Dec. 30, 2019.

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.

## 1. 서론

자폐스펙트럼장애는 유병률이 한층 더 높게 보고 되어 호주의 경우 160명당 1명, 영국은 100명당 1명, 미국은 91명당 한 명으로 평균적으로 100명당 1명의 발생률을 나타내고 있으며 자폐성 장애 징후는 사회성-사회적 의사소통의 장애, 반복적 행동과 관심사, 놀이기술의 지연, 타인의 고통에 대한 반응, 기질 및 감각반응의 이상으로 요약할 수 있다 [1]. 또한, 미국정신의학회가 발간한 DSM-5에서는 자폐스펙트럼장애로 진단되기 위해서는 사회적 의사소통 영역의 3가지 항목과 제한적이고 반복적인 영역에 적어도 2가지 항목이 공통적으로 해당되어야 한다고 규정하고 있고 여러 항목 중 DSM-5의 자폐스펙트럼장애 심각도 수준에 따르면 사회적 의사소통의 초기 심각도 단계는 의사소통에 참여하지만 다른 사람들과 대화를 주고받는 데에는 실패할 수 있으며 친구들을 만들기 위한 시도는 과상하고 대개 실패한다[2]. 이와 같이 자폐스펙트럼장애를 가진 아동들은 대부분 모방행동과 의사소통적 행동이 적고 비전형적인 눈 맞춤을 하며 장난감을 이용한 모방행동과 기능적 활용이 적기 때문에 의사소통 영역에서 언어성 의사소통장애를 훈련시킬 필요가 있다. 이에 따라 사전연구를 거쳐 현장 실험에서 장난감 또는 동물 모방 치료를 목적으로 고양이 로봇을 이용한 놀이 및 통합감각치료, 언어치료 등에 로봇이 사용되었으며 추가적으로 언어성 의사소통장애 훈련에 중점을 두어 로봇의 향상된 지능적인 동작 설계를 위해 기존의 현장실험에서의 동작과 더불어 인공지능 기반의 음성 정보를 활용한 지능화 동작을 제안한다. 음성 정보를 이용한 로봇 동작을 설계하기 위해서는 음성 데이터의 전처리, 주파수 해석, 지역화 등의 여러 전문적인 지식과 기술을 활용하여 입력에 해당하는 출력을 분류해야 한다. 이와같이 대부분의 데이터 특징 분류는 전문가에 의해 구분되거나 수동으로 데이터 타입과 도메인을 코드로 작성하는 것을 통해 이루어지며 이러한 작업은 전문적인 지식이 필요하여 어렵고 작업 시간에 있어서 오래걸리며 추가적인 개발비용도 발생한다[3]. 이에 비해 인공지능망은 입력과 출력을 분류하는데 사용되어 가공되지 않은 순수 데이터를 학습하여 묘사하는데 강력한 잠재력을 가지고 있다[4]. 따라서 인공지능망을

이용하여 음성 정보를 학습시키고 로봇 동작을 분류하는데 효과적으로 사용될 수 있으며 음성 정보의 특징을 추출하기 위해 MFCC 이미지를 사용하고 이러한 이미지를 분류하는 신경망 구조를 설계한다[5]. MFCC 특징 계수는 delta-MFCC, delta-delta-MFCC를 사용하여 39 계수 또는 60 계수를 사용하기도 하지만 두 길이의 계수에서 음성 분류 정확도 차이가 2% 이하로 낮은 개선률을 보였으며 MFCC 13 계수만 사용하여 90% 이상의 정확도를 보이는 연구 사례를 보았을 때 MFCC 입력 전처리를 통한 정확도 개선은 있으나 최적화된 알고리즘의 정확도 향상에 상대적으로 작은 영향을 보이므로 설계된 음성 분류 인공지능망에 13 계수를 사용한다[6,7]. 이러한 이미지를 이용하여 사전에 이미지 분류에는 AlexNet, GoogleNet 등의 여러 인공지능망이 소개 되어있으며 이들은 합성곱 신경망(CNN)을 이용하여 높은 정확도로 이미지를 분류한다. 요즘은 이러한 전통적인 합성곱 신경망 구조를 개선하여 정확도를 향상 시키는 목적으로 순환신경망(RNN)을 추가한 구조가 이용되고 있다[8]. 이와 같은 구조는 RNN의 순차적인 입력과 역동적인 시간적 행동을 처리하는 특징을 이용하여 이미지 정보에서 시간적인 의존성을 활용하기 위해 RNN을 추가로 사용하는데 이러한 방법은 공간과 시간적 맥락을 결합하여 모델링 하므로 다른 전통적인 방법과 단일 RNN보다 CRNN 구조가 가장 높은 성능을 보인다[9,10]. 이러한 구조로 인공지능망을 설계할 때 일반적으로 데이터셋이 제한되는 환경에서 큰 사이즈의 네트워크는 많은 수의 파라미터를 가지며 이러한 네트워크가 더 확대되면 과적합되기 쉬워지는 단점이 존재하여 CNN의 성공을 이끌어내기 위해 필수적인 pooling 동작과 여러 스케일의 특징을 동시에 추출할 수 있는 Inception 기법을 이용하여 최적화한다[11]. 그리고 최적화 방법의 일환으로 가중치 초기화 및 입력 전처리, 데이터 정규화 등의 기법을 적용하여 인공지능망을 설계한다[12,13]. 또한, 추가적으로 일반화 성능을 향상 시키기 위해 신경망 구조내에 여러 최적화 기법을 사용한다[14,15,16]. 그러므로 본 연구에서는 자폐성 장애를 갖는 아동들의 치료 및 교육훈련에 활용될 수 있는 인공지능망 기반의 음성 정보를 이용한 로봇의 지능적 동작을 설계하였으며 로봇 동작의 정확도를 높이기 위해 인공지능망의 여러 최적화 방법과 학습 방법을 이용하였다.

## II. 로봇 시스템

사전 연구를 통해 설계된 로봇은 그림 1과 같으며 로봇 프레임은 3D 프린터를 이용하여 제작되었고 플라스틱 재질을 선택하여 경량화 하였다. 이렇게 구성된 프레임에 외피를 씌워 고양이 인형과 같이 로봇 외형을 형상화 하였으며 로봇의 각 관절은 귀, 눈썹, 눈꺼풀, 입, 목, 앞다리, 뒷다리가 있고 특히 앞다리는 3 자유도를 가지고 있다.

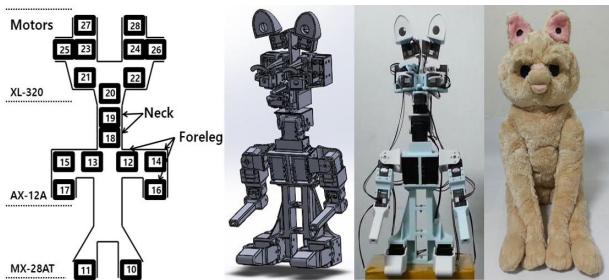


Fig. 1. Robot structure.  
그림 1. 로봇 구조

설계된 로봇 몸체는 내부 제어 시스템에 의해 동작 되는데 로봇 제어 시스템에는 그림 2와 같이 주 제어기와 보조 제어기가 있어 주 제어기(LPC1768)에 의해 각종 센서 데이터 처리 및 모터 제어가 이루어지고 보조 제어기(RaspberryPi)에 의해 지능화 알고리즘 처리 및 인공지능경망 설계가 가능한 구조를 갖는다. 사용된 센서로는 자이로, 초음파, 소리, 실시간 클럭, 터치가 있으며 음향 재생을 위한 코

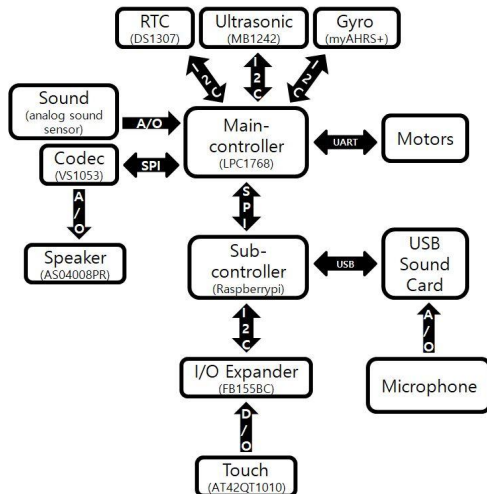


Fig. 2. Robot control system.  
그림 2. 로봇 제어 시스템

덱과 스피커 그리고 음성 데이터 수집을 위해 로봇 내부에 장착된 마이크와 사운드 카드가 있다.

내부 제어 시스템에 의한 로봇 동작 방식은 그림 3과 같이 터치와 음성에 의해 동작되며 자폐성 장애 아동과의 상호작용을 목적으로 동작을 설계하여 입력 이벤트에 상응하는 동작을 출력한다. 고양이 로봇의 동작은 총 18개이며 대표적인 감정 표현 동작은 표 1과 같고 실제 동작 모습은 그림 4와 같다.

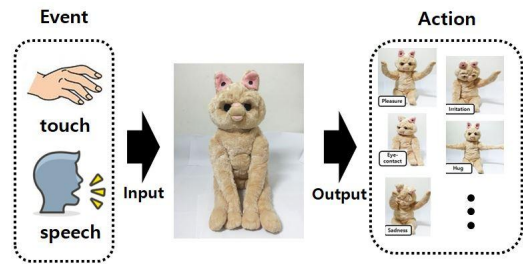


Fig. 3. Operation methods of robot.  
그림 3. 로봇 동작 방식

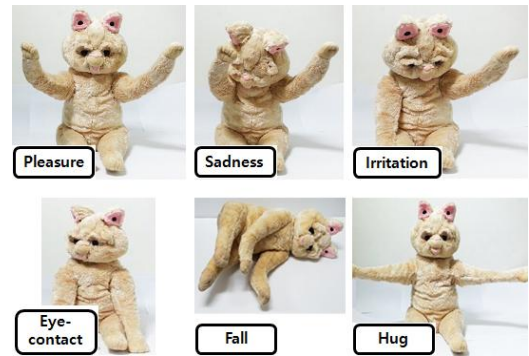


Fig. 4. Representative robot motions.  
그림 4. 대표적인 로봇의 동작

Table 1. Robot representative actions.

표 1. 로봇의 대표 동작

Action	Description	Related Hardware	Purpose
Pleasure	Wave forelegs up and down.	Touch, codec, motors	Positive action
Sadness	Wave forelegs and neck side by side.	Touch, codec, motors	Improper action
Irritation	Wave right foreleg up and down, and move neck right.	Sound, touch, codec, motors	Strong improper action.
Eye-contact	Wave ears left and right, operate neck toward child.	Ultrasonic, touch, codec, motors	Emotion exchange
Fall	Release the torques and play sound.	Gyro, codec, motors	Strong improper action
Hug	Operate each foreleg to outside	Touch, Gyro, codec, motors	Strong emotion exchange

또한 로봇 동작의 지능화를 위해 인공신경망을 이용하여 터치 데이터에 따른 감정 영역을 형성하도록 설계되어 로봇은 터치 위치에 따라 여러 터치 데이터를 분류하여 감정을 갖는 듯한 동작을 한다. 그림 5는 터치 데이터 패턴에 따른 로봇 동작 분류를 나타내며 여러 위치의 터치 데이터를 신경망 구조에서 분류하여 특정 동작을 한다. 이러한 사전 연구를 기반으로 본 연구에서는 음성 정보를 이용한 로봇의 지능적 동작 설계를 위해 추가적인 음성 데이터 처리 및 인공신경망 설계를 통하여 로봇의 지능화 및 자폐아 치료 분야에 활용도를 높인다.

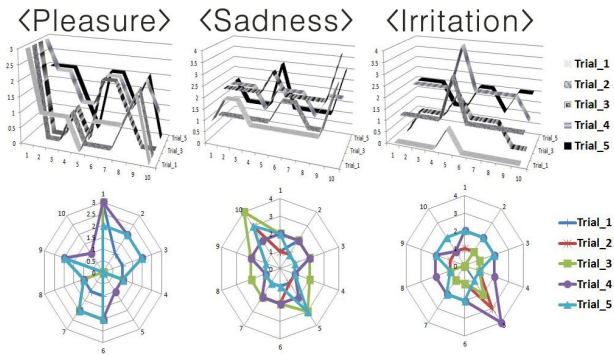


Fig. 5. Touch data classification.  
그림 5. 터치 패턴 데이터 분류

### III. 음성 데이터 분석 시스템

음성 정보를 이용하여 로봇을 동작시키기 위해서는 음성 데이터가 인공신경망에 입력되도록 적합하게 전처리를 거쳐야 한다. 이를 위해 음성 데이터는 음성 분류에 많이 사용되는 MFCC (Mel-Frequency Cepstral Coefficient)를 이용하여 이미지로 생성하는 전처리 과정을 거친다. MFCC는 그림 6과 같이 FFT(Fast Fourier Transform) 및 Mel-Filter Bank를 거쳐 마치 사람의 달팽이관과 같이 소리 대역의 특징 추출이 가능한 방법으로 음성 인식에서 주로 사용된다. 이후 생성된 이미지는 인공신경망에 입력되어 감독학습을 통해 출력을 분류하게 된다. 그

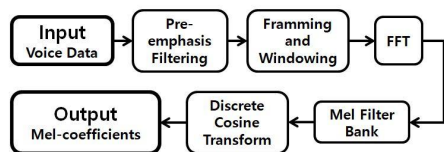


Fig. 6. MFCC classification process.  
그림 6. MFCC 분류 과정

림 7은 인공신경망의 입력과 출력 흐름을 나타낸 것으로 앞서 설명한 MFCC 이미지가 입력 레이어에 해당하며 설계된 인공신경망을 거쳐 출력 레이어에서 입력에 해당하는 로봇 행동을 분류하여 동작 명령이 전달되고 로봇이 동작한다.

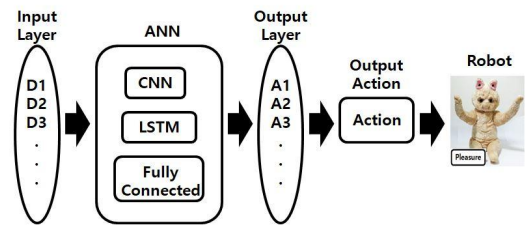


Fig. 7. Motion decision process using speech data.  
그림 7. 음성 정보 처리 구조

인공신경망의 입력으로 사용되는 음성 데이터의 형태는 그림 8과 같이 나타나며 맨위는 순수 데이터의 진폭에 관한 그래프, 중간은 주파수와 데시벨 그래프, 맨밑은 MFCC 출력 그래프로 각 음성 데이터에 따라 특징이 나타난다. 이외에도 음성 데이터 분류에 사용되는 음성 명령은 “좋아(like)”, “바보(fool)”, “안녕(hello)”, “반가워(Nice To Meet You)”, “멍청이(twit)”, “같이놀자(Let Us Play Together)”, “일어서(get up)”, “엎드려(lie down)”, “앉아(sit)”, “고양(cat)”으로 총 10개를 사용한다. 음성 데이터의 속성으로는 데이터 표본화 44.1[kHz], 녹음 시간은 2[s], 마이크와 사람의 거리 약 60[cm], 음성 성량은 60~70[dB]에서 녹음되어 평균 67[dB]이며 음성의 주파수 대역은 200~3500[Hz] 사이에서 FFT를 통해 MFCC 특징을 추출하고 계수는 13개를 사용하여 이미지를 구성한다.

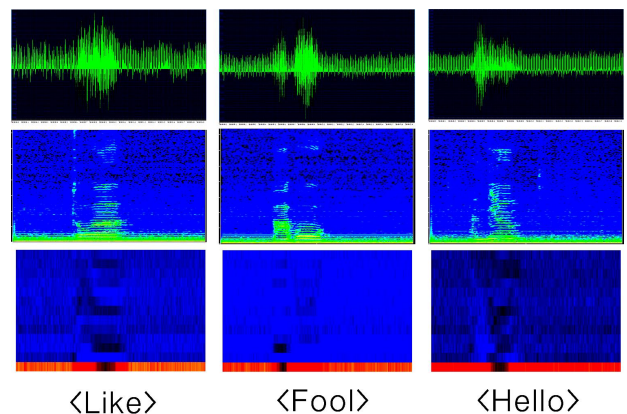


Fig. 8. Speech data features.  
그림 8. 음성 데이터 그래프 세부 특징



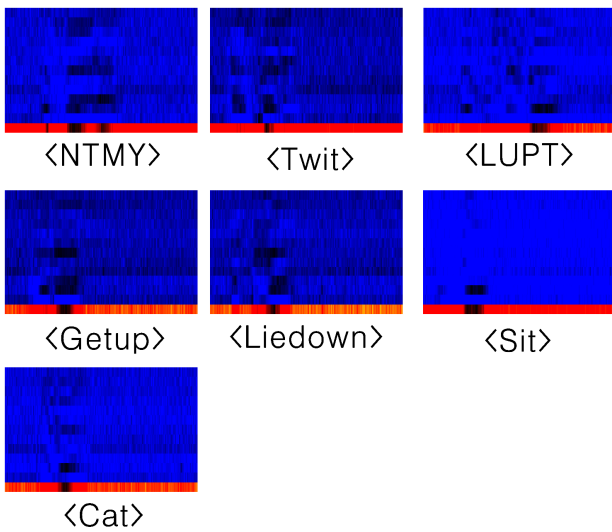


Fig. 9. MFCC images of speech data.  
그림 9. 음성 데이터 MFCC 이미지

#### IV. 인공신경망 설계

앞에서 설명한 MFCC 특징 추출을 통해 생성된 이미지는 인공신경망에 입력되는데 사전 연구에서 음성 데이터 분류에 사용된 인공신경망 구조는 그림 10에서 나타낸 것과 같이 음성의 데이터 특징을 고려하여 시간 순차적 데이터 처리에 적합한 LSTM (Long Short Term Memory)으로 히든 레이어를 구성하고 FC(Fully Connected layer)를 통해 동작을 분류하도록 설계되었다. 하지만 이러한 구조는 낮은 정확도로 로봇 동작이 제한되었으며 이에 본 연구에서는 향상된 정확도로 로봇의 효용 가치를 높여 실제 자폐아 치료에 활용될 수 있도록 인공신경망 구조를 설계한다.

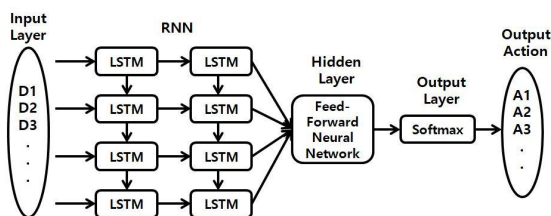


Fig. 10. RNN based Architecture.  
그림 10. RNN 인공신경망 구조

이에 따라 설계된 인공신경망 구조는 그림 11과 같이 나타나며 이미지 분류에 대표적 인공신경망인 AlexNet의 CNN 5층, F.C 3층으로 구성되는 것과 같이 기본적인 CNN 구조로 설계하고 이에 정

확도 향상을 위해 LSTM을 추가하며 여러 최적화 기법을 사용한다. CNN에서 LSTM의 사용은 서론에서 설명한 것과 같이 이미지 정보에서 공간과 시간적 맥락을 결합하여 모델링하기 위해 사용한다.

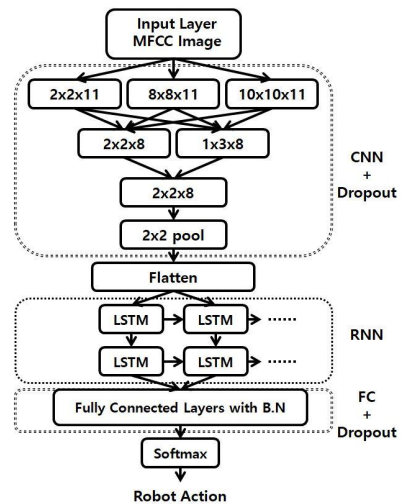


Fig. 11. Proposed ANN architecture.  
그림 11. 제안된 인공신경망 구조

설계된 인공신경망을 학습 시킬 때는 입력 레이어를 학습에 적합하게 전처리를 거쳐야 인공신경망의 최적화에 효과적으로 작용하며 인공 신경망을 최적화할 때 하이퍼 파라미터 조정에 따른 정확도 개선은 균일하지 않지만 대부분 입력 이미지 사이즈 증가에 따라 인공신경망의 정확도는 증가하고 최고 정확도 지점 이후의 이미지 사이즈 증가는 정확도 감소를 나타낸다. 이에 따라 이미지 크기 선정에서 128×128 크기는 최고 정확도를 보였고 64×64와 256×256 같은 작거나 큰 이미지에서는 정확도 감소를 나타내어 128×128 크기의 이미지를 사용한다. 또한, 입력 이미지 데이터 정규화에서 RGB 색상 0~255 사이 값을 활성화 함수인 ReLU (Rectified Linear Unit)의 선형화 특성에 따라 양수만 이용하는 것을 고려하여 0~1.0의 소수 사용으로 정규화 분해능을 정수보다 증가 시킨다. 이렇게 입력 데이터를 가공하여 전처리하는 것 외에도 신경망 구조 내에서 사용하는 일반화 및 정규화 방법이 있으며 머신 러닝에서 일반적으로 알려진 L2, ridge(Tikhonov's), Dropout과 같은 정규화 수단 중 큰 네트워크에서 높은 성능과 낮은 변동성을 보이는 Dropout을 사용하여 과적합의 발생을 감소시켜 일반화한다. 이에 추가적으로 데이터 정규화를 위

해 배치 정규화(Batch Normalization)를 이용하며 모든 레이어에 적용할 경우 빠른 학습 속도를 보이거나 큰 과적합 문제를 보이기 때문에 설계된 인공신경망에 마지막 FC 레이어만 배치 정규화를 사용하여 최적화하며 순환신경망에서는 훈련 단계에서 50000 이상의 학습을 반복하지 않는 이상은 BN-RNN(배치 정규화가 적용된 RNN)을 사용했을 경우와 LSTM의 손실 함수(loss function)의 차이가 상당히 작기 때문에 LSTM을 사용하여 최적의 성능을 갖도록 설계하였다.

### V. 실험 결과

제안한 분석 방법을 이용하여 인공신경망을 설계하여 훈련과 학습을 진행하였다. 학습에 필요한 파라미터는 에폭 5, 반복 200, 배치 10, 배치 크기 100, Dropout 0.7, 학습률 0.0002로 학습시켜 총 학습 데이터 1000개와 실험 데이터 100개를 사용하였다. 그리고 Adam 최적화 알고리즘, Glorot 초기화, ReLU 활성화 함수 그리고 Softmax 분류를 사용하였다. 이와 같은 방법으로 설계 및 학습된 인공신경망의 정확도 실험을 수행하고 다양한 조건하의 파라미터 변경에 따라 최적화된 구조의 신경망을 변경한 결과 제안된 구조에서 정확도가 84.25%로 나타났고 기존 LSTM만 사용한 경우 53.3% 보다 향상된 정확도를 보였다. 그림 12는 입력 데이터에 따른 출력 개수의 평균값이며 10개의 음성 중 “안녕(hello)”의 데이터 분류가 낮은 정확도를 보이는 반면 다른 음성은 입력의 과반수가 정확한 분류를 나타낸다. 이를 보다 세부적으로 분석한 그래프는 그림 13과 같이 나타나며 “안녕” 음성의 대다수가 같은 음절인 “앉아(sit)”으로 분류되고 “일어서(get up)” 음성의 상당수가 같은 음절인 “반가워(NTMY)”로 분류되는 것으로 나타났다. 이러한 결과는 주로 같은 음절에서 일어나는 것을 가정하였을 경우 입

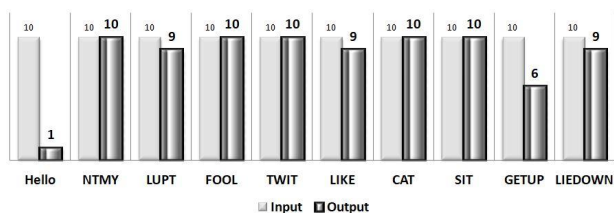


Fig. 12. Classification results in output layers.  
그림 12. 출력 레이어 분류 결과

력 이미지의 유사성에 따라 분류될 수 있지만 “바보(fool)” 또는 “좋아(like)” 그리고 “멍청이(twit)” 또는 “엎드려(lie down)”의 정확도를 고려하였을 경우에는 같은 음절의 다른 음성으로부터 해당 음성의 특징 정보를 분류하여 정확한 결과를 나타내었다.

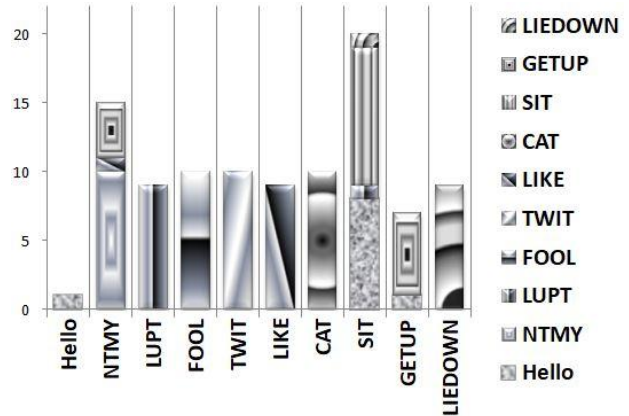


Fig. 13. Speech classification results.  
그림 13. 음성 분류 결과

이에 추가적으로 정확도 비교를 위해 피실험자 1명을 대상으로 학습 데이터를 보여주고 실험 데이터를 분류하도록 실험을 진행하였으며 총 3회 실험 진행에서 평균 79.04%의 정확도로 분류하였다. 일반적으로 사람의 물체 인식률 98.84%와 비교하여 실험 결과를 보았을 때 설계된 인공신경망 구조는 피실험자인 사람의 인식 정확도보다 높은 결과를 나타낸다.

Table 2. Test accuracy table.

표 2. 시험 정확도 비교

Test	Accuracy [%]
RNN	53.3
Human	79.04
CRNN	84.25

하지만 피실험자는 이미지 변환 전의 녹음된 음성을 청취하였을 때 모두 분류 하였으므로 인공신경망에서 같은 음절과 같은 유사한 특징을 가지는 음성의 정확도 향상을 위해서는 이미지 데이터에 대한 전처리 과정의 개선이 필요하다. 그러므로 인공신경망의 정확도를 향상시켜 음성에 따른 로봇

의 동작으로 자폐아 치료 활동에 보다 효과적으로 로봇이 적용되기 위해서 향후 입력 데이터의 전처리에 대한 연구가 필요하며 음성 데이터 또한 다양한 환경과 사람에 의한 수집이 요구된다.

## VI. 결론

자폐아 치료 및 교육을 위한 로봇의 지능적 동작을 설계하였으며 자폐아 치료 활동 중 언어 영역의 의사소통장애에 로봇이 활용될 수 있도록 음성 정보를 활용한 로봇 동작을 설계하였다. 이를 위해 인공신경망을 이용하여 음성 정보를 로봇 동작으로 분류하도록 설계하며 인공신경망의 일반화, 정규화 및 최적화에 필요한 여러 기법을 사용하였다. 사전 연구와 제안된 인공신경망 구조를 비교하고 정확도 실험을 통해 향상된 결과를 나타내며 피실험자와의 정확도 실험을 통해 사람과 인공신경망의 정확도 결과를 비교하여 제안된 인공신경망의 최적화된 구조와 정확도를 분석하였다. 향후 정확도를 개선하여 자폐아 치료에 보다 효과적으로 로봇이 활용되기 위해 데이터 전처리에 대한 연구 및 다양한 환경과 사람에 의한 음성 데이터 수집이 필요하다.

## References

[1] Kyung-Sook Lee, Hyeon-Sook Yoon, Hui-seung Jeong, Hui-jeong Yoo, "A Study on Early Screening of Young Children with Autism Spectrum Disorder in Korea and Its Support System," *Journal of the Korean Association for Persons with Autism(JKAPA)*, vol.15, no.2, pp.93-120, 2015.

[2] Jang-Won Moon, Byeong-Jong Jeong, "Issues and Tasks for Autism Spectrum Disorders in the DSM-5," *Journal of Emotional & Behavioral Disorders(JEBD)*, vol.34, no.3, pp.251-270, 2018.

[3] Nadia Jmour, Sehla Zayen, Afef Abdelkrim, "Convolutional neural networks for image classification," *2018 International Conference on Advanced Systems and Electric Technologies (IC\_ASET)*, pp.397-402, 2018. DOI: 10.1109/ASET.2018.8379889

[4] Sharath Adavanne, Pasi Pertilä, Tuomas Virtanen, "Sound event detection using spatial features and convolutional recurrent neural network," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 771-775, 2017. DOI: 10.1109/ICASSP.2017.7952260

[5] Zahra Shah, Minsu Kim, Gil-Jin Jang, "Image Pattern Classification Using MFCC and HMM," *2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pp.1-4, 2018. DOI: 10.1109/ICCE-ASIA.2018.8552112

[6] Chandrasekhar Paseddula, Suryakanth V. Gangashetty, "DNN based Acoustic Scene Classification using Score Fusion of MFCC and Inverse MFCC," *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, pp.18-21, 2018. DOI: 10.1109/ICIINFS.2018.8721379

[7] Elvira Sukma Wahyuni, "Arabic speech recognition using MFCC feature extraction and ANN classification," *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp.22-25, 2018. DOI: 10.1109/ICITISEE.2017.8285499

[8] Chi Zhang, Thang Nguyen, Shagan Sah, Raymond Ptucha, Alexander Loui, Carl Salvaggio, "Batch-normalized recurrent highway networks," *2017 IEEE International Conference on Image Processing (ICIP)*, pp.640-644, 2018. DOI: 10.1109/ICIP.2017.8296359

[9] Jose Bermudez Castro, Raul Queiroz Feitosa, Patrick Nigri Happ, "An Hybrid Recurrent Convolutional Neural Network for Crop Type Recognition Based on Multitemporal Sar Image Sequences," *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp.3824-3827, 2018. DOI: 10.1109/IGARSS.2018.8517280

[10] Lichao Mou, Xiao Xiang Zhu, "A Recurrent Convolutional Neural Network for Land Cover Change Detection in Multispectral Images," *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp.4363-4366, 2018. DOI: 10.1109/IGARSS.2018.8517375

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1-9, 2015.

DOI: 10.1109/CVPR.2015.7298594

[12] Toshi Sinha, Brijesh Verma, Ali Haidar, "Optimization of convolutional neural network parameters for image classification," *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp.1-7, 2018.

DOI: 10.1109/SSCI.2017.8285338

[13] Cecilia F. Silva, Claurton A. Siebra, "An investigation on the use of convolutional neural network for image classification in embedded systems," *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pp.1-6, 2018. DOI: 10.1109/LA-CCI.2017.8285727

[14] Ekachai Phaisangittisagul, "An Analysis of the Regularization Between L2 and Dropout in Single Hidden Layer Neural Network," *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, pp.174-179, 2017. DOI: 10.1109/ISMS.2016.14

[15] Mirco Ravanelli, Philemon Brakel, Maurizio Omologo, Yoshua Bengio, "Batch-normalized joint training for DNN-based distant speech recognition," *2016 IEEE Spoken Language Technology Workshop (SLT)*, pp.28-34, 2017.

DOI: 10.1109/SLT.2016.7846241

[16] César Laurent, Gabriel Pereyra, Philémon Brakel, Ying Zhang, Yoshua Bengio, "Batch normalized recurrent neural networks," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.2657-2661, 2016.

DOI: 10.1109/ICASSP.2016.7472159

## BIOGRAPHY

### Jin-Gyu Lee (Student Member)



2016 : BS degree in Electrical Engineering, Semyung University.  
2018 : MS degree in Electric and Electronic Engineering, Semyung University.

### Bo-Hee Lee (Member)



1985 : BS degree in Electronics Engineering, Inha University.  
1992 : MS degree in Automation Engineering, Inha University.  
1996 : PhD degree in Automation Engineering, Inha University.

1985~1988 : Researcher at telecommunication system of Samsung Electronics Co. Ltd., Seoul, Korea.

1997~ : Professor, Dept. of Electrical Engineering, Semyung University.