

음성 정보를 이용한 자폐아 치료용 로봇의 동작 설계

Therapeutic Robot Action Design for ASD Children Using Speech Data

이진규*, 이보희*

Jin-Gyu Lee*, Bo-Hee Lee*

Abstract

A cat robot for the Autism Spectrum Disorders(ASD) treatment was designed and conducted field test. The designed robot had emotion expressing action through interaction by the touch, and performed a reasonable emotional expression based on Artificial Neural Network(ANN). However these operations were difficult to use in the various healing activities. In this paper, we describe a motion design that can be used in a variety of contexts and flexibly reaction with various kinds of situations. As a necessary element, the speech recognition system using the speech data collection method and ANN was suggested and the classification results were analyzed after experiment. This ANN will be improved through collecting various voice data to raise the accuracy in the future and checked the effectiveness through field test.

요약

이전 연구에서 자폐성 장애의 여러 특징적 증상을 갖는 아이들의 치료를 위해 사용될 수 있는 로봇을 설계 및 제작하여 현장실험이 진행 되었으며 기존 로봇은 터치에 의한 아이들과 상호작용을 통해 감정 표현 동작을 한다. 이러한 터치 상호작용에 감정 교육 및 치료를 위해 인공신경망을 이용한 동작 설계를 하였다. 하지만 이러한 물리적 접촉은 치료 활동의 초기에 사용되기 어려워 초기 치료 효과를 기대하기 어려웠다. 이에 본 논문에서는 동작 방식을 보완하여 음성 정보를 이용한 빠른 상호작용을 통해 치료 활동이 초기에 가능하고 유연한 대처와 다양한 상황에서 로봇이 사용될 수 있는 동작 설계를 기술한다. 이에 필요한 요소로서 음성 데이터 수집 방법 및 인공신경망을 이용한 음성 인식 구조가 설계되었으며 실험을 통하여 분류 결과를 분석하였다. 이렇게 설계된 인공신경망은 향후 다양한 음성 데이터를 수집하여 정확도를 향상시키고 현장실험을 통하여 동작의 효용성을 살펴볼 것이다.

Key words : Autism Spectrum Disorders, ASD children, therapeutic robot, speech recognition, Artificial Neural Network

1. 서론

자폐성 장애의 특징적 증상은 다섯 가지 분야에서 나타날 수 있는데, 사회적 상호작용의 결여 또는 부적절성, 언어와 의사소통에서의 부적절함, 그리고 상징적이며 상상적인 놀이의 결여, 지적능력의 지체 또는 이상한 형태, 반복적이며 상동적인 행동 등에서 비정상적이며 발달적인 장애를 보인다[1]. 이러한 자폐성 장애를 갖는 아이들을 위해 치료활동에 사용될 수 있는 감정 표현 동작을 하는 로봇이 설계 및 제작 되었으며 현장 실험을 통하여

* Dept. of Electrical Engineering, Semyung University

★ Corresponding author

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

※ Acknowledgment

This study was supported by the research project of Semyung University in 2018.

Manuscript received Dec. 4, 2018; revised Dec. 16, 2018; accepted Dec. 18, 2018

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.

사전 연구가 진행 되었다[2]. 기존 로봇은 터치에 의한 상호작용을 통해 감정 표현 동작이 가능하여 반복적인 감정 교육과 상호작용을 하며 이를 통해 사회적 교육에 사용되었다. 그러나 현장실험에서 자폐아들은 로봇에 대한 약간의 무서움, 두려움, 수줍음, 경계심을 보였으며 친밀감이 형성되기 전까지 물리적 접촉 방식인 터치 상호작용은 초기에 치료 효과를 볼 수 없었다. 이런 이유로 아이들의 경계심을 줄이거나 일찍 없애 로봇이 치료 초기에 활용될 수 있는 동작 방식이 필요하며 접촉하지 않고 로봇이 동작할 수 있는 원거리 접근 방식인 음성을 이용한 동작을 설계하였다.



Fig. 1. Field test.

그림 1. 로봇 현장 실험

한 연구에서는 자폐증을 가진 아이들의 발달을 지원하기 위해 음성을 이용한 접근 방법을 사용하였으며 최근 음성 인식 시스템을 설계하기 위해 인공신경망(Artificial Neural Network)이 사용되고 있다[3], [4], [5]. 인공신경망에는 다양한 신경망이 존재하며 특히 심층 순환신경망 기반의 음향 모델링은 대부분 자동 음성 인식에 표준이 되며, 최근에 순환신경망(Recurrent Neural Network : RNN)이 언어 모델링에 사용 되었고 지속적인 음성 인식을 위해 음향 모델링에 사용된다[6]. 이러한 RNN은 순차적 데이터 모델링을 위한 인공신경망 이므로 자연어 처리 또는 음성 신호 처리에 좋은 모델로서 복잡한 오디오 업무에 훌륭한 결과를 가지고 있다[7]. RNN에 반해, 순방향신경망(Feed-Forward Neural Network : FNN)은 맥락 정보의 부족 때문에 오디오, 비디오 그리고 문서와 같은 순차적인 입력 처리에 어려움을 가진다[8]. 이러한 이유로 RNN이 순차적인 정보를 바로 모델하는 시스템에 주로 사용되며 인공신경망 외에도 음성 특징 조합에 은닉 마르코프 모델(Hidden Markov Model), n-gram

모델을 이용하기도 하지만 RNN은 이들 보다 나은 성능을 보이므로 RNN 구조를 음향 모델에 사용한다. 그리고 인공신경망 성능은 학습 과정에 따라 차이가 나타나므로 RNN의 기울기 값의 소실 문제(Vanishing Gradient Problem)에 의한 RNN의 수정된 형태인 장단기 메모리(Long short-term memory : LSTM)를 사용한다. 순수한 형태의 LSTM 사용 외에도 요즘에는 RNN의 성능을 개선하기 위해 uRNN(Unitary evolution RNN), ceRNN(complex evolution RNN), CG-LSTM(coordinated-gate LSTM) 등 여러 변형된 형태의 RNN이 사용되지만 LSTM과 성능 차이가 작았다[9], [10]. 때로는 LSTM의 양방향 구조가 단방향 구조보다 더 효율적인 구조가 될 수 있다[11]. 하지만 같은 조건에서 LSTM이 RNN과 BiLSTM(Bi-directional LSTM) 보다 나은 결과를 가져오기도 한다. 또한, n-gram 언어 모델과 비교하여 LSTM은 별도의 크로스 엔트로피(cross-entropy) 수정 없이 평균 단어 오류율이 작은 차이로 좋은 성능을 보인다[12]. 이러한 이유로 본 논문에서는 음성 인식에 LSTM을 이용하였고 음성 데이터 수집에 있어 실제 실내 환경에서 우리는 항상 다중 소리의 혼합을 들으며 음성 데이터의 노이즈는 음성 인식 성능에 영향을 미치는 하나의 주요 원인이다[13], [14]. 이러한 노이즈 손상에 관하여 음성 인식의 훈련 과정과 서비스 환경에서 인식 부조화는 노이즈가 섞인 음성을 이용하여 줄일 수 있었다[15]. 이에 따라 대화소리, 기계 소음, 실내 소음이 섞인 환경에서 음성 데이터를 수집하였으며 순환 구조의 인공신경망인 LSTM을 이용하였다. 그러므로, 본 논문에서는 로봇이 치료활동 과정에서 더욱 유연한 대처와 합리적 감정표현 및 다양한 상황에서 사용될 수 있도록 지능적 동작이 설계 되었으며 기존의 터치에 의한 상호작용에 음성 명령을 추가한 로봇 동작을 설계한다.

II. 로봇 전체 구조

로봇 외형은 고양이 형태를 띠며 인형 외피를 씌워 동작 시키고 뼈대는 플라스틱 재질로서 3D 프린터를 이용하여 제작 되었다. 각 관절은 귀, 눈썹, 눈꺼풀, 입, 목, 앞다리, 뒷다리가 있으며 특히 앞다리는 3 자유도를 가지고 있다[16].

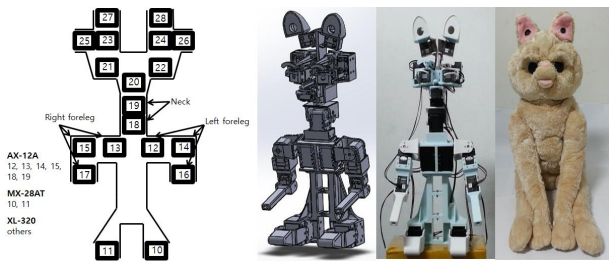


Fig. 2. Robot structure and skin.
그림 2. 로봇 뼈대 및 외관

로봇 내부 제어는 주로 주제어기인 Cortex-M3 계열의 LPC1768에서 이루어지며 각종 센서 정보와 보조제어기에 의해 로봇 동작이 결정된다. 보조제어기는 인공신경망을 형성하고 터치에 의한 상호작용과 음성 명령 동작을 위해 사용하며 소리제어기는 타이머 인터럽트에 의한 실시간 소리 데이터 수집을 위해 사용되었다.

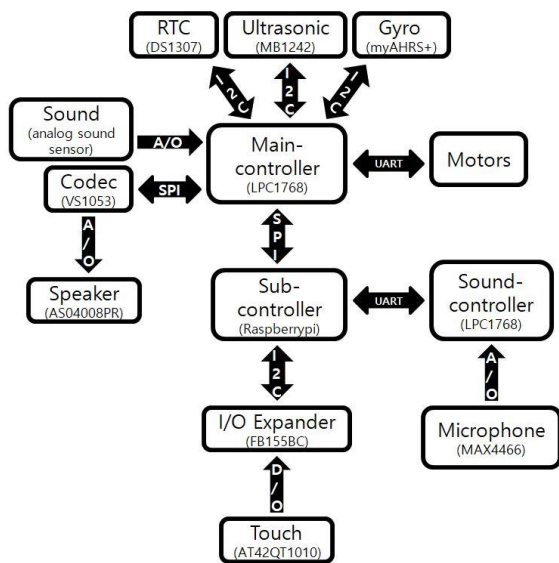


Fig. 3. Controller block diagram.
그림 3. 제어기 블록 선도

소리 데이터 수집의 경우 주제어기의 센서 처리에 의한 포화상태와 보조제어기의 인공신경망 확장, 그리고 목소리 인식의 전체 시스템은 거대하므로 오직 하나의 단일 칩 프로세서는 충분하지 않기 때문에 소리제어기가 추가로 사용된다[17]. 이러한 제어 시스템은 로봇 외피 안에 장착되고 마이크는 로봇 내부 공간 여유를 고려하여 전면부 목 부위에 위치한다. 그리고 로봇의 전원은 제어기에 7.4[v] 4400[mAh]와 모터에 11.1[v] 5200[mAh]의 배터리를 사용하며 뒷다리 부위에 위치한다.

III. 지능적 동작 설계

자폐아 치료활동에 사용되는 고양이 로봇의 동작은 현장실험을 통하여 수정 및 보완이 진행되었고 여러 센서를 이용하여 다양한 상황에서 동작하며 주로 터치 센서에 의한 아이들과의 상호작용으로 감정 표현을 하도록 동작이 설계 되었다.

1. 동작 분석

설계된 로봇의 동작은 총 18개이며 대표적인 동작 목록과 감정표현은 표 1과 그림 4와 같다.

Table 1. Action table of robot.

표 1. 고양이 로봇의 동작 테이블

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



Fig. 4. Representative emotional action.

그림 4. 대표적인 감정표현 동작

이러한 감정표현 및 상호작용 목적의 동작이 다양한 치료활동에서 사용되고 유연한 상황 대처가 가능하도록 인공신경망을 이용한 지능적 로봇 동작을 설계 한다. 인공신경망에는 터치 센서와 음성

데이터가 입력되어 동작이 출력되며 터치 시의 경우 감정 표현 동작이 치료활동에서 일관 되거나 급변할 수 있어 다중 터치 입력에 의한 감정 분류가 가능한 인공지능망을 설계 하였다[18]. 인공지능망을 사용하지 않는 동작은 식 (1)와 같이 고유 동작의 반복과 해당 터치 센서에 의해 동작되며 로봇이 동작 가능한 상태에서 출력이 나타나고 식 (2)는 인공지능망을 이용한 식으로 모든 터치 데이터를 이용하여 계산된 값에 의해 동작이 출력된다.

$$\alpha_n = \alpha_n(t-1) + \beta_m$$

$$y_m(t) = (\alpha_n + n(t)) \times s \tag{1}$$

$$hypothesis = \sum_{n=1}^{10} (\omega_n x_n + b_n)$$

$$predicted_output = softmax(hypothesis)$$

$$y_n(t) = (\alpha_n + predicted_output) \times s \tag{2}$$

y : action, α : iteration number, β : action count, n : touch sensor, m : action number, s : state, x : touch frequency, w : weight, b : bias

이에 따라 식 (2)는 터치 데이터를 입력 받아 그림 5와 같이 감정 표현 동작을 분류하였으며 분류 결과 터치 부위에 관련하여 감정 표현의 관계를 확인 하였다.

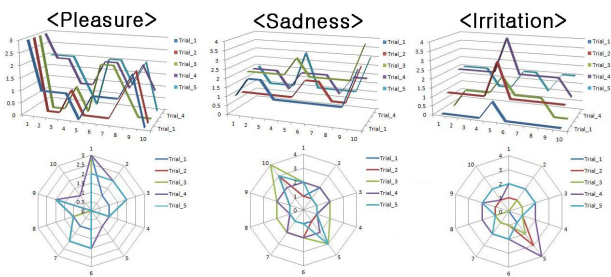


Fig. 5. Data graph for three representative Emotions. 그림 5. 3가지 감정의 데이터 그래프

하지만 치료활동을 위한 물리적 접촉인 터치 상호작용은 친밀감이 형성되기 전까지 아이들이 접근하기 어려웠으며 이러한 아이들에게 보다 쉬운 접근 방법으로 음성을 이용한 비접촉 로봇 동작 방식을 설계하였다. 음성 인식에는 하드웨어 구성과 데이터 수집 방법, 음성 분류 방법에 대한 기술이 필요하므로 설계된 구조는 음성 명령에 의한 로봇의 동작을 위해 단일 마이크를 설치하여 음성 데이

터를 수집하였으며 음성 인식 시스템의 경우 다중 마이크 사용으로 보통 2개의 스테레오를 사용하지만 고양이 로봇의 경우 내부 공간 여유에 따라 단일 마이크가 목 부위에 설치되어 음성 데이터를 수집한다. 또한, 표본화 비율은 16~44.1[kHz]를 갖는 것이 일반적인 경우이지만 보조제어기의 인터럽트 지연과 직렬 통신 속도, 메모리 사이즈를 고려하여 2[kHz]의 표본화 비율을 가지며 2초 동안의 음성을 60[dB]로 녹음 하였다. 이러한 음성 표본화 구조는 그림 6와 같이 소리제어기에서 수행되며 보조제어기의 상태 신호에 영향을 받는다. 보조제어기는 로봇이 동작하지 않을 때 음성 데이터를 전달 받으며 2개의 디지털 입출력 핀에 의해 상태를 확인한다. 이러한 방식은 로봇이 동작 중에 음성 데이터가 버퍼에 쌓이는 것을 없애 오동작을 방지한다.

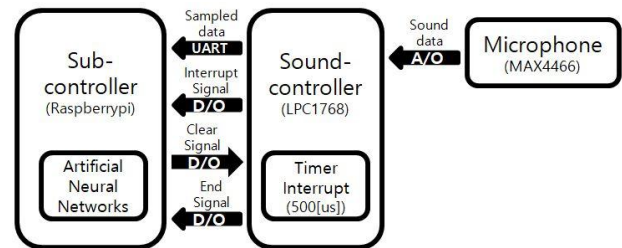


Fig. 6. Speech data acquisition and transmission. 그림 6. 음성 데이터 수집 및 전송

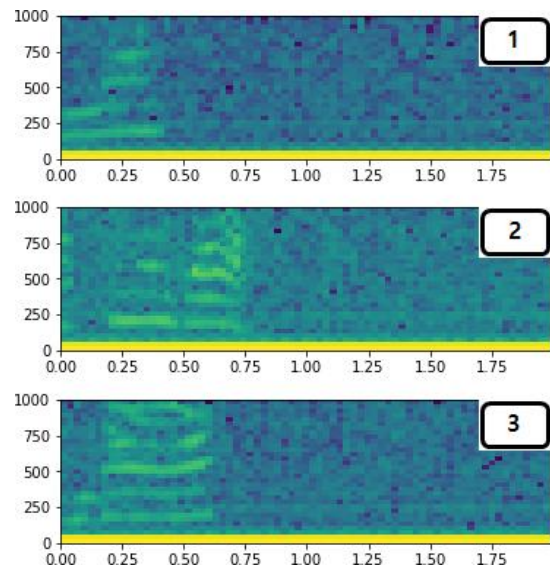


Fig. 7. Spectrogram of three speeches. 그림 7. 3가지 음성의 스펙트로그램

해당 표본화된 음성 데이터는 재생하여 녹음된 음성을 확인하였고 스펙트로그램을 통해 음성 데이

터의 사용 가능성을 확인 하였다. 그림 7는 녹음된 데이터의 스펙트로그램으로 안녕(1번), 같이놀자(2번), 반가워(3번) 그래프이다. 그림에서 나타난 색의 변화는 음성 크기를 뜻하며 가로 축과 세로 축은 시간과 주파수를 나타낸다. 이를 통해 시간에 따라 주파수 값이 변화하는 특징을 확인 할 수 있다.

2. 인공신경망 설계

수집된 음성 데이터를 이용하여 음성 인식이 가능한 인공신경망 구조를 그림 8와 같이 설계하였으며 음성 데이터는 입력층에 입력되어 인공신경망을 통해 출력층에서 동작과 노이즈를 구분하고 로봇에 동작 신호를 전달한다. 이렇게 설계된 구조 외에도 음성에 9가지 감정 분류를 하는 로봇이 있다[19]. 이는 자폐아를 위해 많은 감정 동작이 사용되지만 감정적 이해가 부족한 자폐아들에게 복잡하다. 그리고, 현장실험에서 자폐아들은 언어장애, 발달장애, 지체장애 등을 동반하여 말하는 능력이 부족한 경우가 대부분이었기 때문에 언어 구사 능력이 부족했다. 이에 따라 고양이 로봇은 음성 명령에 3가지 분류를 한다. 이는 설계된 음성 인식의 목적이 아이들의 경계심을 일찍 없애거나 줄이는 것이므로 이 단계에서는 복잡한 교육을 하기 보다 빠른 친밀감 형성을 통해 원활한 치료활동과 상호작용이 목적이다. 그러므로 음성 명령에는 “안녕”, “같이놀자”, “반가워”의 3가지 음성이 사용되며 노이즈 데이터를 입력으로 사용하여 음성과 노이즈를 분류하는 것으로 잘못된 입력에 의한 동작을 분류한다. 그림 8의 로봇에서 A1~3은 음성 명령에 의한 동작이며 인공신경망의 출력층에서 A1,A2,...은 동작과 노이즈 분류이다.

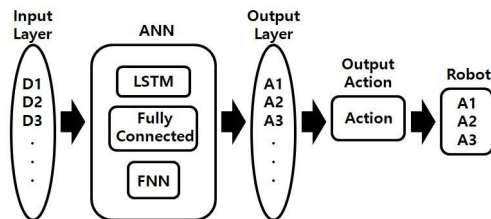


Fig. 8. Control process.
그림 8. 제어기 동작 순서

일반적으로 음성 인식에 사용되는 음성 데이터는 수십~수백 명의 대상자를 상대로 수집되는 거대한 양이지만 본 연구에서는 녹음 대상자를 모집하

기 쉽지 않았으며 자체 녹음을 통해 3명의 대상자가 참여하여 상대적으로 적은 양의 음성 데이터가 녹음 되었다. 표본화된 음성은 총 85개로 인공신경망의 학습 데이터로 사용되며 깨끗한 음성 60개와 소음 섞인 음성 15개가 사용 되었고 노이즈 10개가 사용된다. 또한, 2초 동안의 2[kHz] 녹음에 따른 4000개의 데이터 길이를 가지며 그림 9와 같은 특징을 갖는다. (<Hello> : “안녕”, <LUPT> : “같이놀자”, <NTMY> : “반가워”, <noise1> : “소리지름”, <noise2> : “기계소리”).

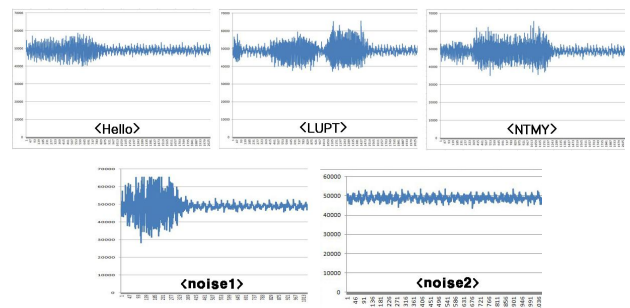


Fig. 9. Sampled data graph.
그림 9. 음성과 노이즈 데이터 그래프

그림 9와 같이 표본화된 음성 데이터는 그림 10와 같은 인공신경망 구조에 입력되며 2층 구조의 LSTM에서 순차적 입력 데이터의 특징을 추출하고 3층 구조의 FNN에서 특징을 분류하는 구조이므로 입력층, 히든층 5개, 출력층에 따른 총 7층 구조를 갖는다. 4000개의 음성 데이터 값을 LSTM에서 490개로 특징값을 추출하고 추출된 490개 값을 FNN에서 Softmax함수를 통해 동작과 노이즈로 분류한다.

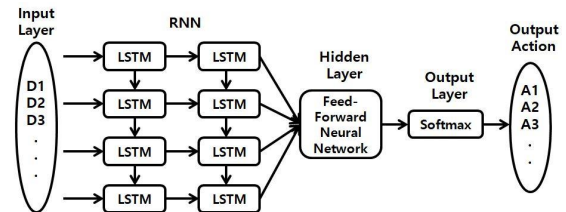


Fig. 10. Designed ANN.
그림 10. 설계된 인공신경망 구조

음성 데이터의 특징을 추출하는 과정은 그림 11와 같이 LSTM 학습을 통해 출력값이 나타나며 LSTM은 총 5단계 학습을 진행하여 입력에 매칭되는 특징값을 출력한다. LSTM과 FNN의 학습은 분

리되어 이루어지며 가중치 초기화와 최적화 알고리즘은 보편적 수단인 Xavier와 Adam을 사용하여 학습하고 각 하이퍼 파라미터(Hyper-parameter)는 LSTM에 2500 반복(Epoch), 10 노드(Node), 490 배치 사이즈(Batch Size), 0.01 학습율(Learning rate)이 사용되며 FNN에 1000 반복, 0.0001 학습율, 0.7 드롭아웃(Dropout)이 사용된다.

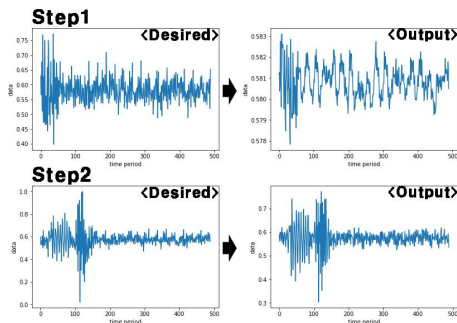


Fig. 11. LSTM learning process.
그림 11. LSTM 학습 과정

이러한 방식으로 설계되어 학습된 인공신경망은 최종 출력에서 그림 12와 같이 음성 입력에 매칭되는 로봇 동작이 나타난다. 앞서 설명한 A1은 Hello 으로 눈 맞춤 동작을 하며 A2는 LUPT으로 같이 놀자 동작을 하고 A3는 NTMY으로 반가워 동작을 수행한다. 그리고 나머지 음성에 대해서는 잡음으로 분류하여 동작하지 않는다.



Fig. 12. Actions using speech data.
그림 12. 음성 명령에 의한 로봇 동작

IV. 결과 분석

이후 학습된 인공신경망에 음성 데이터를 입력하여 실험 하였으며 입력된 데이터는 “안녕”음성으로 그림 13와 같이 목표값에 상응하는 LSTM의 특징 추출값이 나타났고 FNN의 출력에서 해당하는 동작을 분류 하였다. 이러한 방식으로 실험을 진행한 후 30개의 각종 음성 데이터에서 정확도가 53.3%로 나타났다. 실험에는 LSTM 외에도 RNN과, BiLSTM

을 사용하여 진행했으며 같은 조건에서 RNN 27%, BiLSTM 26%의 정확도를 나타내어 결과적으로 높은 정확도가 나타난 LSTM을 사용한다. 현재, 음성 인식에 사용된 인공신경망 구조는 53.3%의 정확도로 절반 수준이지만 이는 적은 양의 훈련 데이터에 따른 결과로 보여지며 일반적인 음성 인식 시스템에는 수십~수백 명의 대상자를 상대로 데이터가 녹음되지만 본 연구에서는 3명의 대상자가 참여하여 적은 양의 데이터를 사용했으므로 향후 음성 인식의 정확도 개선을 위해 많은 대상자를 상대로 음성 데이터 녹음이 요구되며 축적된 데이터를 이용한 연구 진행이 필요하다.

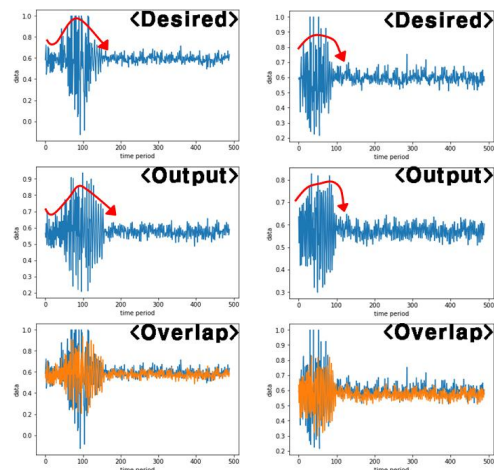


Fig. 13. comparison to Desired value and LSTM output.
그림 13. 목표값과 LSTM 출력값 비교

V. 결론

자폐아 치료활동에 사용될 수 있는 고양이 로봇의 지능적 동작을 설계하였으며 기존 물리적 접촉 방식인 터치에 의한 상호작용은 현장실험 초기에 아이들의 무서움, 두려움, 수줍음, 경계심에 의해 친밀감이 형성되기 전까지 접근이 어려웠다. 이런 이유로 본 논문에서는 음성을 이용한 비접촉 접근 방식을 고안하였고 터치와 음성이 함께 사용되는 동작 구조를 제안하였다. 새롭게 추가된 음성 인식 구조에는 LSTM과 FNN을 이용하여 설계되었으며 실험을 통하여 최적의 순환신경망 구조를 설정하였다. 설계된 음성 인식 구조는 실험에서 53.3%의 정확도를 보였으나 향후, 현장 실험을 통하여 음성 데이터를 축적하고 이를 통해 설계된 지능적 동작을 보완하여 치료활동에 보다 정확한 감정 분류와

유연한 상황대처 및 합리적 표현이 될 수 있도록 하고 현장실험을 통해 설계된 동작의 효용성을 살펴볼 것이다.

References

- [1] Yang-Soon Kim, "A Case Study of an Animal Assisted Play Therapy with Two Autistic Children," *The Korea Journal of Counseling*, vol.6, no.2, pp.485-497, 2005.
- [2] Jin-Gyu Lee, Bo-Hee Lee, Jin-Soun Jung, Ja-Young Kwon, "Robot Design and Action Study for the Treatment of Autistic Spectrum Disorders Children," *Institute of Korean Electrical And Electronics Engineers*, vol.20, no.2, pp.196-199, 2016. DOI:10.1145/1463689.1463716
- [3] Sujirat Attawibulkul, Boonserm Kaewkamnerdpong, Yoshikazu Miyana, "Noisy speech training in MFCC-based speech recognition with noise suppression toward robot assisted autism therapy," *2017 10th Biomedical Engineering International Conference (BMEiCON)*, pp.1-5, 2017. DOI:10.1109/BMEiCON.2017.8229135
- [4] Sungho Jeon, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyouon Kwon, Hae-Yong Yang, "Empirical study of drone sound detection in real-life environment with deep neural networks," *2017 25th European Signal Processing Conference (EUSIPCO)*, pp.1858-1862, 2017.
- [5] Fei Tao, Carlos Busso, "Aligning Audiovisual Features for Audiovisual Speech Recognition," *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pp.1-6, 2018. DOI:10.1109/ICME.2018.8486455
- [6] Sreeram Ganji, Rohit Sinha, "Exploring recurrent neural network based acoustic and linguistic modeling for children's speech recognition," *TENCON 2017~2017 IEEE Region 10 Conference*, pp.2880-2884, 2017. DOI:10.1109/ICME.2018.8486455
- [7] Po-Wei Hsiao, Chia-Ping Chen, "Effective Attention Mechanism in Dynamic Models for Speech Emotion Recognition," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.2526-2530, 2018. DOI:10.1109/ICASSP.2018.8461431
- [8] Giambattista Parascandolo, Heikki Huttunen, Tuomas Virtanen, "Recurrent neural networks for polyphonic sound event detection in real life recordings," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.6440-6444, 2016. DOI:10.1109/ICASSP.2016.7472917
- [9] Izhak Shafran, Tom Bagby, R. J. Skerry-Ryan, "Complex Evolution Recurrent Neural Networks (ceRNNs)," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.5854-5858, 2018. DOI:10.1109/ICASSP.2018.8462556
- [10] Gregory Gelly, Jean-Luc Gauvain, "Optimization of RNN-Based Speech Activity Detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol.26, no.3, pp.646-656, 2018. DOI:10.1109/TASLP.2017.2769220
- [11] Naima Zerari, Samir Abdelhamid, Hassen Bouzgou, Christian Raymond, "Bi-directional recurrent end-to-end neural network classifier for spoken Arab digit recognition," *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, 2018, pp.1-6. DOI:10.1109/ICNLSP.2018.8374374
- [12] Kazuki Irie, Zhihong Lei, Ralf Schlüter, Hermann Ney, "Prediction of LSTM-RNN Full Context States as a Subtask for N-Gram Feedforward Language Models," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp.6104-6108. DOI:10.1109/ICASSP.2018.8461743.
- [13] Ha Manh Do, Weihua Sheng, Meiqin Liu, Senlin Zhang, "Context-aware sound event recognition for home service robots," *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, pp.739-744, 2016. DOI:10.1109/COASE.2016.7743476
- [14] Qingyang Hong, Caihong Zhang, Xiaoyang Chen, Yan Chen, "Embedded speech recognition system for intelligent robot," *2007 14th International*

Conference on Mechatronics and Machine Vision in Practice, pp.35-382. 2007.

DOI:10.1109/COASE.2016.7743476

[15] Youngjoo Suh, Younggwon Kim, Hyungjun Lim, Jahyun Goo, Youngmoon Jung, Yeonjoo Choi, Hoirin Kim, Dae-Lim Choi, Yongju Lee, "Development of distant multi-channel speech and noise databases for speech recognition by in-door conversational robots," *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*, pp.5-8, 2017.

DOI:10.1109/ICSDA.2017.8384419

[16] Jin-Gyu Lee, Bo-Hee Lee, Ju-Yeong Jang, Ja-Young Kwon, Keum-Hi Mun, Jin-Soun Jung, "Study on Cat Robot Utilization for Treatment of Autistic Children," *International Journal of Humanoid Robotics*, vol.14, no.2, pp.1-15, 2017.

DOI:10.1142/S0219843617500013

[17] Yonghui Xing, Wenzhuo Chen, "Design of Speech Recognition Robot Based on MCU," *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*, pp.253-256, 2012. DOI:10.1109/IHMSC.2012.70

[18] Jin-Gyu Lee, Bo-Hee Lee, "Therapeutic Behavior of Robot for Treating Autistic Child Using Artificial Neural Network," *Fuzzy Systems and Data Mining IV Proceedings of FSDM 2018*, vol.309, pp.358-364, Sep. 2018. DOI:10.3233/978-1-61499-927-0-358

[19] Jonathan C. Kim, Paul Azzi, Myounghoon Jeon, Ayanna M. Howard, Chung Hyuk Park, "Audio-based emotion estimation for interactive robotic therapy for children with autism spectrum disorder," *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp.39-44, 2017. DOI:10.1109/URAI.2017.7992881

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.