

# 인공지능 기반의 울음소리를 이용한 영아 상태 인식

황일규\*, 송호범\*\*

## AI-based Infant State Recognition Using Crying Sound

Il-Kyu Hwang\*, Ho-Bum Song\*\*

---

본 논문은 2017년도 동양미래대학교 학술연구과제 지원에 의하여 연구되었음.

---

### 요 약

영아기의 아기는 울음으로만 대부분의 의사소통을 한다. 이 시기는 뇌 발달 단계상 가장 중요한 시기이기 때문에 아기의 요구를 정확하게 이해하기 위한 많은 노력이 필요하다. 본 연구에서는 아기의 울음소리와 패턴에 대한 연구들의 분석과 이를 바탕으로 인공지능을 활용하여 상태를 인식하는 방법에 대해 소개한다. 음원은 기존 연구에서 작성된 패턴별 아기 울음과 관련된 음원을 저장 후 분류하여 구성한다. 이를 기반으로 스펙트로그램 형태의 이미지 데이터로 변형하여 시각화하여 분석하고 전처리하여 인공지능 기반의 CNN 모델의 학습 데이터로 활용한다. 학습된 모델을 이용하여 주파수 측정치의 평균으로 분류하던 기존 연구들과는 다른 비교적 높은 정확도의 아기 상태 인식 결과를 보여주는 시스템을 구현한다. 향후 전문가의 검증을 받은 음원 데이터가 충분히 확보된다면 즉시 상용화도 가능할 것으로 판단된다.

### Abstract

Babies in infancy communicate most by crying. Since it is a critical period in the brain development stage, a lot of effort is needed to understand the desire of the baby correctly. In this study, we introduce the analysis of research on baby's crying and patterns, and how to classify them using artificial intelligence. The sound source consists of the baby's crying sample from the existing research, and the pattern is stored and classified. It is transformed into spectrogram image data, visualized, analyzed and preprocessed, and used as learning data for CNN model. This study presents how to implement a system that classify baby's crying with comparatively higher accuracy than existing research that uses patterns from the average of frequency measurements. If much more input data is verified by experts in the near future, it can be commercialized easily and immediately.

### Keywords

infant, spectrogram, CNN, sound classification, emotion recognition

---

\* 동양미래대학교 로봇자동화공학부 교수(교신저자) · Received: Apr. 24, 2019, Revised: Jul. 15, 2019, Accepted: Jul. 18, 2019  
- ORCID: <https://orcid.org/0000-0001-9615-0377> · Corresponding Author: Il-Kyu Hwang  
\*\* 동양미래대학교 로봇자동화공학부 교수  
School of Robot & Automation Engineering, Dongyang Mirae University,  
Gyeongin-ro 445, Guro-gu, Seoul, 08221, Korea  
- ORCID: <https://orcid.org/0000-0002-3451-5729>  
Tel.: +82-2-2610-1834, Email: ikhwang@dongyang.ac.kr

## 1. 서 론

감정이란 외부의 물리적인 자극에 의한 감각, 지각으로부터 인간의 내부에 야기되는 심리적 체험[1]으로서, 이러한 감정 정보는 얼굴 표정, 음성, 심장박동수, 혈압, 체온 등으로 접근할 수 있다[2]. 그중에서도 음성기반 감정인식은 사용자의 음성 신호를 분석하여 사용자의 감정을 자동으로 인식하는 기술이다[3].

아기는 태어나서부터 구강 조음기관의 조절이 가능하여 웅얼이를 시작하는 생후 6개월까지 자신의 의사표현의 대부분을 울음이라는 방법으로 표현한다. 이 기간이 짧다고 소홀이 여길 수도 있지만 이 기간은 부모와의 많은 교감과 소통을 통해 집중적으로 관리해야 하는 시기이며, 만약 이 시기를 적절하게 보낼 수 없다면, 성인이 된 후 정신 질환과 중독 위험성이 높아지고[4], 스트레스 호르몬인 코르티솔이 과다 분비되면서 뇌에 안 좋은 영향을 받게 된다[5]. 그러나 현대사회의 부모나 보호자들은 아기가 사용하는 언어(울음)를 정확하게 인식하기엔 충분한 시간과 경험을 갖지 못하고 많은 시행착오와 어려움을 겪는 경향이 있어, 본 연구에서는 최근 기술 분야에서 가장 활발하고 유용하게 사용되는 인공지능(Artificial intelligence) 기술을 활용하여 아기의 울음을 분석하고 그 상태를 보호자에게 정확하게 전달함으로써, 효율적으로 육아에 도움을 주는 방법을 제시하고자 한다.

본 연구에서는 음성 신호를 통하여 영아들의 감정 상태를 인식하는 접근 방법을 다룬다. 아기의 욕구에 맞게 분류된 울음소리를 음원으로 사용하고, 이를 Python 언어를 기반으로 시각화하여 분석하고 전처리하여 인공지능 기반의 심층신경망 모델의 학습데이터로 활용하였으며 학습된 모델을 이용하여 시스템을 구현하였다.

## II. 관련 연구

인간의 음성은 의사소통을 하기 위한 수단이자 감정을 전달하는 수단이며, 소리에 포함된 감정은 화자의 심리 상태를 표현한다. 그러나 일반적으로

음성 기반 감정 인식은 사람의 감정정보를 효과적으로 포함하는 특징의 부족으로 인하여 얼굴 표정 인식과 같은 다른 감정 인식에 비해 낮은 인식률을 보인다[2].

음성을 이용한 감정인식의 연구 분야에서 사용되었던 감정 모델은 크게 두가지로서[6], 첫째는 인간의 감정을 감정에 대한 성향을 나타내는 Valence 축과 감정의 강도를 나타내는 Arousal 축의 2차원 영역으로 다양한 감정상태를 표현하는 Valence-Arousal 모델이다[7]. 그러나 이 모델은 감정 상태를 연속적으로 표현하여 다양한 감정을 선정할 수 있지만, 구분 짓기 힘든 애매모호한 감정이 존재하는 단점이 있기도 하다. 따라서 두 번째 모델로서 즐거움, 놀라움, 두려움, 화남, 슬픔 등의 대표 감정을 선정하여 감정을 인식하는 모델을 사용한다. 이 모델의 경우에는 감정의 표현이 명확하여 감정에 따른 분류가 쉬우므로 감정 인식분야에서 많이 사용되고 있다[3].

따라서 감정 인식에 적합한 음성 신호의 특징 벡터를 찾거나, 분류 방법론을 다르게 적용하여 정확도를 개선하려는 노력이 있어왔다[2][8]-[12]. 특히, [11]은 특징 벡터로 MFCC를 이용하였으며, [2]는 특징 벡터로 MFCC, MEDC, LPC, LPCC, IAV, DAMV, Tempo, ZCR의 8가지를 사용하여, 감정 카테고리간의 분리도를 비교분석하였다. [12]는 PLP (Perceptual Linear Prediction)을 이용하여 음성신호의 감정인식에 사용하였다.

최근에는 인공지능 기법이 급속히 발전하고 있고, 이를 음성 감정인식에 활용하려는 시도가 많아지고 있다. MFCC를 이용하여 특징벡터를 추출하고, SVM(Support Vector Machine) 분류기법을 사용하여 감정을 분류한 연구들[3][13]과, 음성신호를 스펙트로그램(Spectrogram)과 같은 특징 벡터의 이미지로 변환하여 CNN등의 심층 신경망(Deep learning)을 적용하여 3가지 범주의 오디오 장르를 분류하거나 [14], 5가지 감정을 분류하고 인식하려는 연구도 있어왔다[15].

아기 울음의 경우에는 당장 필요한 욕구를 반영하게 되므로, 일반적인 음성을 통한 감정 인식에서 즐거움 등의 표현이 빠지고 별도의 울음 원인별 상

대분류가 필요하게 된다. 즉, 아기는 배고픔, 불편함, 졸림, 지루함 등 모든 욕구와 생각을 울음으로 표현한다. 이는 아기의 가장 중요한 소통수단이며, 호흡기나 심혈관계 이상을 나타내는 생물학적 기능을 반영하는 수단이기도 하다. 이러한 아기울음의 중요성은 오래전부터 인지되어 그 중요성을 인정받아 다양한 연구가 진행되어왔다[16]-[20].

[18]은 모든 아기는 인종과 성별의 구분없이 원인별 울음패턴이 비슷하다는 것을 알아내고, 이를 특허출원 및 제품화하였다. 약 100명의 아기들의 울음을 수집하고 분석하여 Annoyed, Bored, Hungry, Sleepy, Stressed 총 다섯 가지의 주파수 패턴으로 아기의 울음패턴을 정의하였으며, 주파수 패턴 분석 칩을 개발하여 상용화하고 여러 병원에서 임상실험 결과 비교적 좋은 임상결과를 획득하였다고 주장한다.

또한, 호주인 Dunstan은 울음소리별 요구패턴에 대하여 의구심을 갖고 연구를 시작하였으며, 전 세계적인 연구 조사를 통해 인종, 피부색, 문화와 관계없이 모든 아기가 울기 전에 요구에 따라 5가지 패턴의 소리를 낸다는 것을 알게 되었고 이를 정리하여 ‘DBL(Dunstan Baby Language)’로 발표하였다 [19]. 이는 Neh(배고픔), Owh(졸림), Eh(트림), Eairh(가스참), Heh(불편함) 등의 총 다섯 가지의 발음패턴으로 아기의 상태를 분류한다.

한편, [17]에 의해 배고픔과 통증에 따른 영유아 울음에는 평균 주파수가 다르다는 것이 알려졌으며, 졸릴 때와 트림할 때의 아기 울음소리에 패턴이 있음을 밝힌 연구[20]를 통해 Dunstan이 발표한 DBL에서 졸림 ‘Owh’와 트림 ‘Eh’ 소리가 검증되었다.

이와 같이 아기의 상태를 알아내기 위해 울음소리의 주파수 패턴을 분석하는 연구들이 있어왔으나, 아기의 울음소리를 특징 벡터의 이미지로 변환하여 심층 신경망에 적용하여 아기 상태 인식의 정확도를 높이려는 시도는 거의 없었다. 따라서 본 연구에서는 아기의 울음 패턴을 증명한 기존 연구결과들을 바탕으로, 울음소리를 시각화하고 CNN(Convolutional Neural Network) 합성곱 신경망 모델을 기반으로 학습을 진행하여 아기의 정확한 요구 사항이나 상태를 분류하고 인식하는 방법을 제시하고자 한다.

### III. 학습 데이터의 생성

#### 3.1 데이터 선정

인공지능을 사용하는데 있어서 학습용 데이터는 매우 중요하다. 더구나 영상 데이터가 아닌 음성 데이터는 시각화하여도 지도학습(Supervised learning)을 위한 판단근거를 세우는 것이 어렵기 때문에 그 취급에 있어서 각별한 주의가 필요하다.

본 연구를 위한 음원 데이터는 전문가가 아니면 울음소리만으로 아기의 상태를 판별하기가 매우 어려우므로, 일반인이 직접 만든 음원이나 영상, 온라인 영상 스트리밍 사이트(예, YouTube 등), 또는 SNS 상의 영상이나 음원을 취득하고, 이를 아기의 상태에 맞게 정확하게 분류하는 것은 불가능하다. 따라서 본 연구에서는 Dunstan이 TV쇼에서 강연했던 영상 스트리밍 자료에 나온 패턴별 아기음성 예제와 홈페이지[19]에서 판매하는 자료 속 패턴별 예제음원을 저장 후 분류하여 구성하였다.

각 음원데이터는 트림(Trim), 가스참(Burp gas), 배고픔(Hungry), 졸림(Sleepy), 불편함(Uncomfortable)의 다섯 가지 범주로 분류하였다. 또한 평온한 상태를 인식하기 위한 조용(Quiet)이 추가되었다. 표 1은 본 연구에 사용된 학습 데이터들의 구분과 총량이다.

표 1. 학습 데이터의 분류 및 총량

Table 1. Classification and total amount of learning data

Classification	Time(s)	Size(MB)	Number of image files
burp Gas (G)	76	6.43	51
Hungry (H)	58	4.93	55
Trim (T)	43	3.66	78
Sleepy (S)	90	7.65	86
Uncomfortable (U)	36	3.10	73
Quiet (Q)	2102	353	1048
Total	2405	378.77	1391

표 1에서 시간은 각 범주별로 녹음된 음원의 총 시간의 합, 용량은 이들 음원의 전체 파일 크기이며, 이미지 파일의 개수는 이 음원들이 3.2절에서의 전처리와 추출과정을 통하여 음성 신호를 스펙트로

그램으로 변환한 배열의 형태로 저장된 이미지 파일의 개수를 의미한다. 올바른 인공지능 학습을 위해 가능한 데이터의 양을 범주별로 균일하게 맞춰주는 것이 좋으나 보유한 범주별 데이터의 총량이 매우 부족하기에 데이터의 불균형을 감수하고 학습을 진행하였다.

### 3.2 데이터의 전처리, 특징추출

획득한 아기 울음 음원들을 인공지능의 학습데이터로 바로 사용하기에는 한 음원 파일마다 데이터가 너무 많기 때문에 부적합하다. 따라서 학습하기 전에 음원별로 특징추출과 전처리 과정이 필요하다. 아래의 예로 드는 시각화 샘플음원은 hungry 패턴의 음원 중의 하나이며 길이는 1.4초이다. 그림 1은 샘플음원의 wave plot이다.

스펙트로그램은 시간의 변화에 따라 음성신호의 주파수 성분의 차이를 농도나 색상의 차이로 표현한 것이다. 스펙트로그램의 가로축은 시간을, 세로축은 주파수를 나타내며, 소리의 강도가 변화함에 따라 색이 달라진다. 따라서 특정 감정이나 상태에 따라서 스펙트로그램의 모양이 다르게 표현된다.

본 연구에서는 음원의 특징추출 방법으로 푸리에

변환을 통해 소리 데이터에서 시간대별 주파수를 분리해 스펙트로그램 형태의 이미지로 변경하기 위해서 Python 언어의 scipy라이브러리를 사용하여 window size를 20으로, step size를 10으로 하여 FFT 스펙트로그램으로 변환하였다.

모든 음원을 시각화 처리 코드를 사용하여 아래 그림 2와 같이 나열하여 패턴별 음원의 주파수 구성성분을 확인하였고, 불필요한 성분을 제거하기 위해 나이퀴스트 - 새넨 표본화 정리(Nyquist-Shannon sampling theorem)에 따라 음원을 특정주파수(본 연구에서는 24,000Hz)로 resample하여 표본화를 진행하였다.

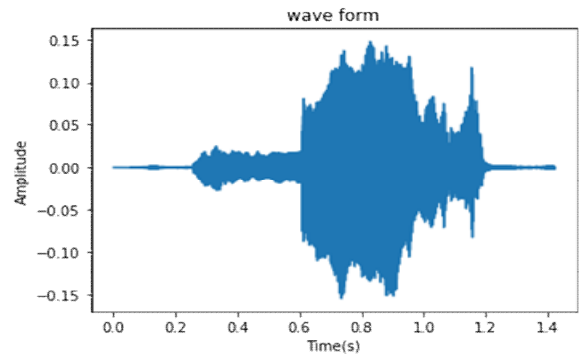


그림 1. wave form의 예 (hungry패턴 No. 25)  
Fig. 1. Examples of wave form (hungry pattern No. 25)

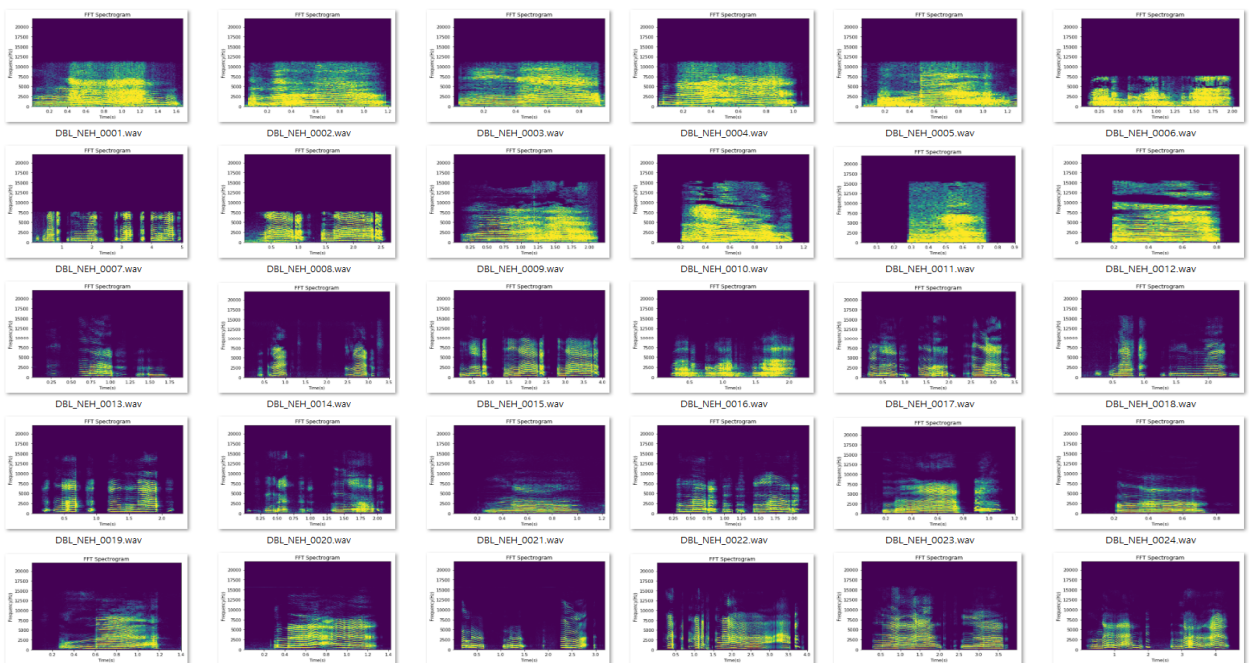


그림 2. Hungry 패턴 FFT 스펙트로그램  
Fig. 2. Hungry pattern FFT spectrogram

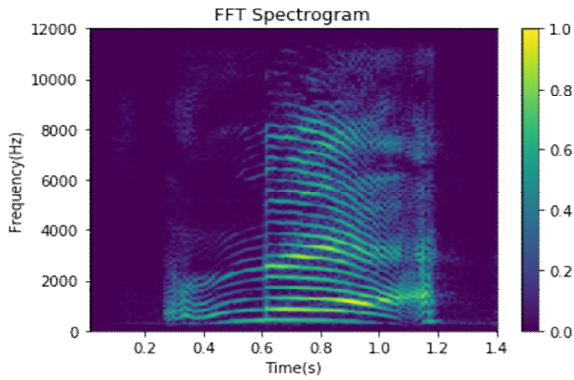


그림 3. 표본화와 정규화된 Hungry패턴 No.25의 FFT 스펙트로그램 (이미지 크기: 241X199)  
 Fig. 3. Sampling and normalized hungry pattern No.25 FFT spectrogram (image size: 241X199)

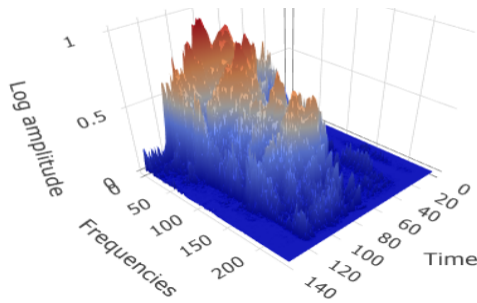


그림 4. 그림 3의 3D화  
 Fig. 4. 3D transformation in Fig. 3

제작한 인공지능 모델 내부에서 활성화 함수로 Relu 함수를 사용하는데 Relu 함수는 0보다 큰 값은 그대로 통과시키고 0 보다 작은 값은 0으로 통과시키는 함수이기에 입력 데이터의 값을 0에서 1사이의 값으로 변경해주는 정규화(Normalization) 작업이 필요하다.

그림 3은 정규화 작업을 마친 표본화한 샘플음원 (Hungry패턴)의 FFT 스펙트로그램의 예이다. 그림 4는 그림 3을 삼차원 형태로 표현한 것이다.

이와 같은 절차에 따라 전처리와 추출과정을 마치고 Python 언어 기반의 numpy 라이브러리를 사용하여 배열 형태로 저장하였다.

생성된 데이터를 이용하여 FFT 결과 값을 친밀도가 가까운 값끼리 묶는 알고리즘인 t-SNE (t-Distributed Stochastic Neighbor Embedding)를 이용하여 이들 배열을 시각화하여 표현한 것이 그림 5에 보여지고 있다.

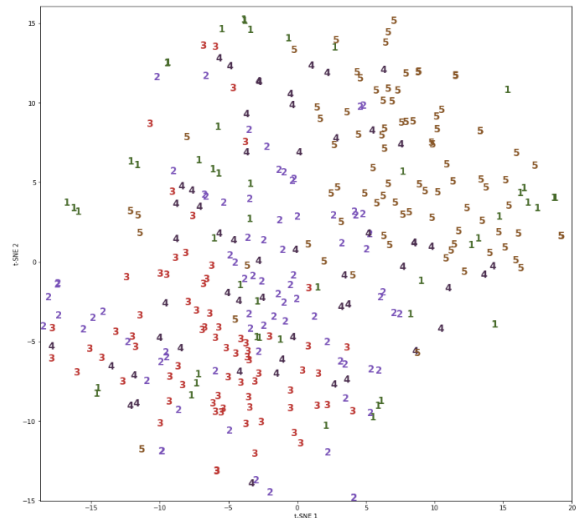


그림 5. t-SNE 시각화 분포도  
 Fig. 5. t-SNE visualization distribution chart (1:burp gas, 2:hungry, 3:trim, 4:sleepy, 5:uncomfortable)

그림 5에서 나타난 바와 같이 많은 부분의 데이터가 패턴별 군집화가 극히 어렵게 분포되어 있어, 아기 울음을 단순히 패턴별 주파수 평균만을 사용하여 분류하기가 쉽지 않음을 알 수 있다.

#### IV. 인공지능 모델 구현

##### 4.1 인공지능 신경망의 구현

본 연구의 인공지능 프레임 워크로는 텐서플로우 (Tensorflow)와 케라스(Keras)라는 심층 신경망 라이브러리를 사용하였다. 그리고 음원의 분류를 위해 분류 데이터와 그 분류 값을 같이 네트워크에 입력해야 하는 지도학습방식을 사용하였으며, 추출한 FFT 스펙트로그램은 그 구조가 이미지와 유사하기 때문에 필터로 특정 범위에서 특징을 뽑아내어 네트워크를 훈련시켜 이미지 분류에 뛰어난 성능을 보이는 CNN 합성곱 신경망 모델을 사용하였다.

학습에 사용한 CNN 모델은 convolution layer 3개 층, max pooling layer 2개 층, 그리고 출력이 각각 64개와 6개인 fully connected layer 2개 층으로 이루어져 있다. 모델은 점점 노드 수가 줄어드는 피라미드형 구조이며, 각 활성화 함수는 함수값의 포화문제가 없으며 함수 구조상 수렴 속도가 빠른 연산속도를 보여주는 ReLU 함수로 구성되어 있다.



## 18 인공지능 기반의 울음소리를 이용한 영아 상태 인식

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 237, 195, 24)	624
max_pooling2d_1 (MaxPooling2D)	(None, 59, 97, 24)	0
activation_1 (Activation)	(None, 59, 97, 24)	0
conv2d_2 (Conv2D)	(None, 55, 93, 48)	28848
max_pooling2d_2 (MaxPooling2D)	(None, 13, 46, 48)	0
activation_2 (Activation)	(None, 13, 46, 48)	0
conv2d_3 (Conv2D)	(None, 9, 42, 48)	57648
activation_3 (Activation)	(None, 9, 42, 48)	0
flatten_1 (Flatten)	(None, 18144)	0
dropout_1 (Dropout)	(None, 18144)	0
dense_1 (Dense)	(None, 64)	1161280
activation_4 (Activation)	(None, 64)	0
dropout_2 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 6)	390
activation_5 (Activation)	(None, 6)	0

Total params: 1,248,790  
 Trainable params: 1,248,790  
 Non-trainable params: 0

그림 6. CNN 모델 구성 요약

Fig. 6. CNN model summary

또한 학습 시 임의의 뉴런을 삭제하여 신호전달을 차단하여 과적합(Over-fitting)을 방지하기 위해 2개의 Dropout 층을 추가하였다.

마지막에는 softmax 함수를 사용하여 각 감정의 category가 나올 확률을 퍼센트(%)로 계산하도록 하였다. 그림 6은 본 연구에서 사용한 CNN 모델의 구성 요약이다.

학습에 필요한 batch 크기는 40, epoch는 5000번으로 하였으며, 매개변수 최적화를 위한 방법으로는 Adam(Adaptive moment estimation) 알고리즘을 사용하였다.

### 4.2 학습결과

일반적으로 제한된 데이터들을 가지고 학습과 테스트를 할수 있는 방법 중의 하나로서 보유한 데이터들에 대하여 일부는 학습용으로 또 다른 일부는 테스트용으로 사용하도록 한다. 학습용 데이터들은 학습시킬 파라미터들을 결정하는데 사용되며, 테스트용 데이터는 학습된 모델을 평가하는데 사용된다.

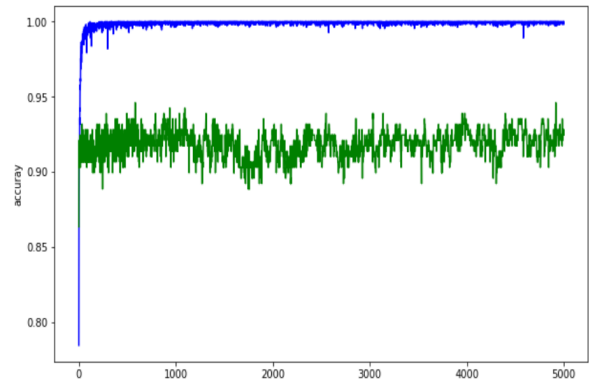


그림 7. Training/Test 정확도 그래프 (학습 정확도(파랑색)는 97%, 테스트 정확도(초록색)는 91%)

Fig. 7. Training/test accuracy graph (training accuracy(blue): 97%, test accuracy(green): 91%)

본 과제에서는 음원들을 각 범주별로 비례한 크기를 갖도록 학습용으로 80%, 테스트용으로 20%의 비율로 랜덤하게 구분하여 그림 6의 CNN 모델에 입력하여 학습을 시행하였다.

학습된 모델이 얼마나 정확하게 예측했는지를 평가하는 기준으로서 일반적으로 정확한 예측수를 총 예측수로 나눈 정확도(Accuracy)가 사용된다.

그림 7에서의 학습정확도는 학습용 데이터를 이용하여 학습 후에 각 범주별 음원의 예측 정확도를 나타내고, 테스트 정확도는 학습된 모델에 테스트 데이터를 적용했을 때의 예측 정확도를 의미한다. 그림 7에 나타난 바와 같이 전체적으로 학습 결과와 이 학습 모델을 사용한 테스트 결과값의 정확도가 90%를 넘고 있음을 알 수 있다.

또한 오류 경향을 좀더 세밀하게 분석하기 위하여 모델에 비 학습된 테스트 데이터를 넣어 표 2와 같은 다중 클래스 분류 모델을 평가하는 혼동행렬을 사용하여 학습 모델의 평가를 진행하였다. 좌측 레이블이 실제 Label 값, 하단 레이블이 인공지능이 예측한 Label 값이다. 예를 들어, T 범주의 경우 실제 값인 T 범주로 분류하는 확률이 71%, U 범주로 분류하는 확률이 14%, H와 S 범주로 분류하는 확률이 각각 7%라는 의미이다. 표 2를 기준으로 학습 결과를 분석하면, G 범주의 경우 결과값이 약간 낮은 경향을 보이고 있으나, 이는 너무 적은 데이터 양에 의한 것으로 판단되며, 향후 데이터의 양이 충분하게 확보된다면 예측 결과가 훨씬 더 향상될 것

으로 기대된다. 그 외의 대부분의 범주에서는 비교적 좋은 예측 결과를 보이고 있음을 알 수 있다.

표 2. 혼동행렬  
Table 2. Normalized confusion matrix

True label	Q	1.00	0.00	0.00	0.00	0.00	0.00
	G	0.00	0.45	0.09	0.09	0.09	0.27
	T	0.00	0.00	0.71	0.14	0.07	0.07
	U	0.00	0.00	0.27	0.60	0.00	0.13
	H	0.00	0.00	0.20	0.00	0.70	0.10
	S	0.00	0.21	0.00	0.00	0.05	0.74
		Q	G	T	U	H	S
	Predicted label						

데이터의 종류 및 획득 방법, 연구 방법의 차이 등으로 단순 비교는 어려우나, 일반적으로 음원의 주파수로부터 특징 벡터를 구하여 감정 인식을 하는 기존의 연구들[3][12][13][15]의 인식률이 53~78% 정도의 결과를 보이는 것에 비하여는 본 연구에서는 적은 자료량에도 불구하고 90% 이상의 정확도를 얻었다. 또한 이는 Dustan[19]이 제시한 예제 자료가 일관성이 있음을 의미하기도 한다.

### V. 결 론

영아기의 아기는 오직 울음으로만 대부분의 자기 의사를 표현한다. 긴급한 상황에 대한 의사표현도 울음으로 하게 되는데 대부분의 보호자들은 울음소리만 듣고 아기가 전달하고자 하는 구체적인 의미를 알기 어려운 경우가 많다. 이와 관련하여 기존의 연구들은 음원의 주파수 평균을 이용하는 방식을 사용하였으나 이는 잡음에 매우 취약한 것으로 알려져 있다. 또한 현재까지 상용으로 출시된 아기울음 번역 제품과 어플리케이션[18][20]들은 높은 정확도를 얻기 위해 측정시 아기의 체중별 측정거리를 다르게 지정해야 할 정도로 사용자에게 많은 요구를 하고 있다.

본 연구의 목적은 이러한 기존 연구들의 단점을 극복하고, 가능한 높은 정확도를 가진 인공지능 기반의 음향을 이용한 아기 상태인식 시스템을 구현

하여 그 유효성을 검증하는 것이다.

본 연구를 위한 음원 데이터는 전문가가 아니면 울음소리만으로 판별하기가 매우 어려우므로, 기존 연구에서 사용된 패턴별 아기음성 음원들을 저장한 후 트림, 가스잡, 배고픔, 졸림, 불편함의 다섯 가지 범주로 분류하여 사용하였다. 본 연구에서는 이들 음원들을 시간 구간별 주파수 분리를 통해 스펙트로그램 형태의 이미지 데이터로 변형하여 시각화하였다. 이러한 전처리과정을 통한 데이터는 심층 신경망 기법인 CNN을 사용하여 90% 이상의 정확도를 가지는 상태 인식 결과를 보여주었다. 따라서 본 연구에서는 아기 울음을 인식할 수 있는 기존의 주파수 측정방식과는 다른 정확도 높은 우수한 모델을 설계하는 방법을 제시하였음을 확인하였다.

본 연구는 데이터를 수집하고 정제하는 과정 중에 실사용 데이터의 유효성 검증 및 연구 방향 설정을 위한 목적도 가지고 있으며, 이러한 관점에서 는 비록 사용되는 데이터 양이 적음에도 불구하고 연구방향이 적절하다고 판단된다.

향후 본 연구 방법과 결과를 반영하여 전문가의 검증을 받은 충분한 음원데이터를 획득하고 활용하여 보다 발전된 연구를 진행할 계획이다. 따라서 이들 음원데이터를 이용하여 본 연구에서 제시된 CNN 모델로 학습을 진행한다면, 훨씬 높은 정확도를 가지는 시스템을 쉽게 구현할 수 있고, 이를 스마트폰의 앱(Application)이나 전용기기의 형태로 실제 상용화도 충분히 가능할 것으로 기대된다.

### References

- [1] Y. Kim and Y. C. Bae, "Design of Emotion Recognition Model Using Fuzzy Logic", Journal of Korean Institute of Intelligent Systems, Vol. 10, No. 1, pp. 268-282, May 2000.
- [2] B. R. Shin and S. P. Lee, "A Comparison of Effective Feature Vectors for Speech Emotion Recognition", The transactions of The Korean Institute of Electrical Engineers, Vol. 67, No. 10, pp. 1364-1369, Oct. 2018.
- [3] J. H. Bang and S. Lee, "Call Speech Emotion

- Recognition for Emotion based Services", Journal of KISS : Software and Applications, Vol. 41, No. 3, pp. 208-213, Mar. 2014.
- [4] S. S. Janetsian-Fritz, N. M. Timme, M. M. Timm, A. M. McCane, A. J. Baucum II, B. F. O'Donnell, and C. C. Laphish, "Maternal deprivation induces alterations in cognitive and cortical function in adulthood", *Translational Psychiatry*, Vol. 8, No. 1, Article number: 71, Mar. 2018.
- [5] K. Sylva, A. Stein, P. Leach, J. Barnes, L. Malmberg, and FCCC-team, "Effects of early child-care on cognition, language, and task-related behaviours at 18 months: an English study", *Br J Dev Psychol.*, 29 (Pt 1), pp. 18-45, Mar. 2011.
- [6] J. Park, D. Park, J. H. Park, and J. Park, "Development of Human Sensibility Recognition System using Hidden Markov Model", *Proceedings of HCI 2004*, pp. 605-610, Feb. 2004.
- [7] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology", *Development and Psychopathology*, Vol. 17, No. 3, pp. 715-734, Sep. 2005.
- [8] A. B. Kandali, A. Routray, and T. K. Basu, "Emotion recognition from Assamese speeches using MFCC features and GMM classifier", *TENCON 2008-2008 IEEE Region 10 Conference*, pp. 19-21 Nov. 2008.
- [9] Z. Xiao, E. Dellandrea, L. Chen, and W. Dou, "Recognition of emotions in speech by a hierarchical approach", *ACII 2009. 3rd International Conference*, 10-12, pp. 401-408, Sep. 2009.
- [10] Y. H. Cho and K. S. Park, "A Study on The Improvement of Emotion Recognition by Gender Discrimination", *Journal of IEEK*, Vol. 45, No. 4, pp. 107-114, Aug. 2008.
- [11] J. W. Choi, Y. S. Park, C. M. Jung, and Y. J. Kim, "Development of A Sound Detection System for Security using the MFCC and HMM", *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, pp. 352-353, Nov. 2016.
- [12] B. Jung, S. Cheun, Y. Kim, and S. Kim, "An Emotion Recognition Technique using Speech Signals", *Journal of Korean Institute of Intelligent Systems*, Vol. 18, No. 4, pp. 494-500, Aug. 2008.
- [13] S. Y. Kim, E. Y Kim, and E. H. Goo, "Emotion Recognition Method from Speech Using the Youtube Video and Emotion API", *Proceedings of The Institute of Electronics and Information Engineers*, pp. 1096-1099, Jun. 2018.
- [14] H. W. Yun, S. H. Shin, W. J. Jang, and H. Park, "On-Line Audio Genre Classification using Spectrogram and Deep Neural Network", *Journal of Broadcast Engineering*, Vol. 21, No. 6, pp. 977-985, Nov. 2016.
- [15] S. Park, D. Kim, S. Kwon, and N. Park, "Speech Emotion Recognition based on CNN using Spectrogram", *Information and Control Symposium*, pp. 240-241, Oct. 2018.
- [16] P. S. Zeskind and B.M. Lester, "Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications", *Child Dev.*, Vol. 49, No. 3, pp. 580-589, Sep. 1978.
- [17] T. Murry and P. Amundson, "Acoustical characteristic of infant cries: fundamental frequency", *Child Lang.*, Vol. 4, No. 3, pp. 321-328, Oct. 1977.
- [18] WhyCry Technology - <http://www.why-cry.com>. [accessed: Apr. 05, 2019]
- [19] Dunstan Baby Pty Ltd - <http://www.dunstanbaby.com> [accessed: Apr. 05, 2019]
- [20] U. J. Son, D. W. Seo, and M. J. Bae, "On an Acoustical Discrimination of Baby Crying during Sleepy and Burp State", *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, pp. 1040-1041, Jan. 2017.



저자소개

황 일 규 (Il-Kyu Hwang)



1988년 2월 : 서울대학교  
조선공학과(공학사)  
1990년 2월 : 한국과학기술원  
산업공학과(공학석사)  
1990년 ~ 1994년 : 삼성종합기술원  
주임연구원  
2002년 12월 : 미국 University of

Wisconsin-Milwaukee 기계공학과 (공학박사)  
1994년 3월 ~ 현재 : 동양미래대학교 로봇자동화공학부  
교수

관심분야 : 인공지능, 임베디드 시스템, CAD/CAM

송 호 범 (Ho-Bum Song)



1984년 2월 : 서울대학교  
전기공학과(공학사)  
1986년 2월 : 서울대학교  
전기공학과(공학석사)  
1986년 ~ 1994년 : 삼성전자  
시스템개발실 선임 연구원  
2005년 : 성균관대학교 정보통신

공학부 박사 수료  
1995년 3월 ~ 현재 : 동양미래대학교 로봇자동화공학부 교수  
관심분야 : 딥-러닝, 모바일 프로그램, 임베디드 시스템