Journal of Biomedical Engineering Research 34: 8-13 (2013)
http://dx.doi.org/10.9718/JBER.2013.34.1.8

학 술 논 문

인공 신경망을 이용한 보청기용 실시간 환경분류 알고리즘

서상완[1] · 육순현[1] · 남경원[1] · 한종희[1] · 권세윤[2] · 홍성화[2] · 김동욱[3] · 이상민[4] · 장동표[1] · 김인영[1]

[1]한양대학교 의용생체공학과, [2]성균관대학교 의과대학 이비인후과학교실
[3]삼성종합기술원 바이오헬스 연구실, [4]인하대학교 전자공학과

Real Time Environmental Classification Algorithm Using Neural Network for Hearing Aids

Sangwan Seo[1], Sunhyun Yook[1], Kyoung Won Nam[1], Jonghee Han[1], See Youn Kwon[2], Sung Hwa Hong[2], Dongwook Kim[3], Sangmin Lee[4], Dong Pyo Jang[1] and In Young Kim[1]

[1]Department of Biomedical Engineering, Hanyang University
[2]Department of Otolaryngology-Head and Neck Surgery, Samsung Medical Center
[3]Bio and Health Lab, Samsung Advanced Institute of Technology
[4]Department of Electronic Engineering, Inha University
(Received September 7, 2012. Accepted December 12, 2012)

Abstract: Persons with sensorineural hearing impairment have troubles in hearing at noisy environments because of their deteriorated hearing levels and low-spectral resolution of the auditory system and therefore, they use hearing aids to compensate weakened hearing abilities. Various algorithms for hearing loss compensation and environmental noise reduction have been implemented in the hearing aid; however, the performance of these algorithms vary in accordance with external sound situations and therefore, it is important to tune the operation of the hearing aid appropriately in accordance with a wide variety of sound situations. In this study, a sound classification algorithm that can be applied to the hearing aid was suggested. The proposed algorithm can classify the different types of speech situations into four categories: 1) speech-only, 2) noise-only, 3) speech-in-noise, and 4) music-only. The proposed classification algorithm consists of two sub-parts: a feature extractor and a speech situation classifier. The former extracts seven characteristic features - short time energy and zero crossing rate in the time domain; spectral centroid, spectral flux and spectral roll-off in the frequency domain; mel frequency cepstral coefficients and power values of mel bands - from the recent input signals of two microphones, and the latter classifies the current speech situation. The experimental results showed that the proposed algorithm could classify the kinds of speech situations with an accuracy of over 94.4%. Based on these results, we believe that the proposed algorithm can be applied to the hearing aid to improve speech intelligibility in noisy environments.

Key words: hearing aids, classification, artificial neural network, hearing impaired

Corresponding Author : In Young Kim, Department of Biomedical Engineering, Hanyang University
TEL: +82-2-2220-0691 / FAX: +82-2-2220-4949
E-mail: iykim@hanyang.ac.kr
Sangwan Seo and Sunhyun Yook contributed equally to this paper and should be regarded as equivalent authors.
이 논문은 2012년도 지식경제부 바이오의료기기 전략기술개발사업(10031764)과 서울시 산학연 협력사업(SS100022) 의 지원을 받아 수행되었음.
이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행되었음(2012R1A1A2041508).

I. 서 론

감각신경성 난청(sensorineural hearing loss)을 가지고 있는 난청인들 은 정상인들에 비해 역동범위(dynamic range)가 좁고 높은 청력역치를 보여 어음청취에 어려움이 있다. 특히 배경잡음이 있는 경우 역동범위의 감소에 의한 누가현상(recruitment)과 시간분석능력 및 주파수분석능력의 저하로 인해 언어변별력이 감소되어 언어를 인지하는데 있어

서 더 많은 어려움을 느낀다[1,2].

감각신경 난청인들 은 음성신호와 배경잡음이 함께 있는 경우 언어 인지능력이 저하되고 청력역치가 난청도에 따라 2.5~7 dB 정도 상승하게 되서[3,4] 언어를 인지하는데 어려움을 느끼게 되는데 이러한 불편함을 해소하기 위해서 보청기를 착용하게 된다.

보청기에는 난청보상 알고리즘과 잡음제거 알고리즘 등 다양한 알고리즘들이 들어가게 되는데 이러한 알고리즘들은 실제 다양한 환경에서 최적의 성능을 보여주지를 못한다[4]. 이러한 점을 해결하기 위해서 일부 보청기에는 동작환경을 수동으로 조작할 수 있는 스위치가 있어서 볼륨을 조정하거나 일부 기능을 끄는 작업을 수행하게 된다.

하지만 보청기 착용자는 보청기 안에 있는 각각의 알고리즘에 대한 이해도 부족하고, 환경이 변할 때 마다 직접 수동으로 조작하는 것이 불편하기 때문에 자동으로 주변 환경을 인식하고 그에 맞게 보청기 알고리즘을 조정할 수 있는 자동 환경 구분 알고리즘이 연구되고 있다[6-8,12].

이러한 기존의 연구에서는 언어의 인지도만 높이기 위해 잡음 환경 분류에 초점을 맞추고 있거나, 특수한 상황의 잡음만 분류하여 제거 하였지만 본 연구에서 는 음성만 있는 상황, 음성과 잡음이 있는 상황, 잡음만 있는 상황, 그리고 음악을 듣는 상황으로 분류하여 일상 생활에서 자주 접하는 환경을 빠르고 정확하게 분류 할 수 있도록 알고리즘을 개선 하였다.

다시 말해서, 본 연구에서는 8 msec 마다 특징 값들을 추출하여 인공신경망 을 통해 음성만 있는 상황, 음성과 잡음이 있는 상황, 잡음만 있는 상황 그리고 음악을 듣는 상황 의 4 가지 환경에 대해 90% 이상의 성능을 보이며 분류 할 수 있는 실시간 환경 분류 알고리즘을 개발하였다.

II. 본 론

1. 환경 분류 모델

보청기를 통하여 들어 오는 음성 신호를 음성 환경에 따라 다르게 처리해 주기 위해 음성만 있는 상황, 음성과 잡음이 있는 상황, 잡음만 있는 상황 그리고 음악을 듣는 상황(4가지 음성 환경) 이 4 가지 환경으로 인공 신경망 모델링 기법을 사용하여 분류를 하였다. 인공 신경망 모델링 기법은 일반적으로 계층의 수에 따라 크게 단층 신경망과 다층 신경망의 2 가지로 구분되는데 본 논문에서는 다층 신경망 구조 중 구분의 정확도를 높이기 위해서 2 개의 은닉 층을 사용하는 피드포워드(Feed-forward) 다층 신경망을 사용하였다[5-7].

피드포워드 연산은 다음 절차를 따른다. 우선 입력 벡터들이 입력 층의 각 입력 뉴런에 제시되었다. 각 은닉 뉴런은 입력들의 가중된 합을 계산해서 스칼라 네트 활성화를 만들어낸다. 이 값은 입력 값과 입력-은닉 층 사이의 연결 가중치의 내적이다.

$$net_j = \sum_{i=1}^{n} x_i w_{ij} + b_{1j} = \sum_{i=0}^{n} x_i w_{ij} = w_j^t x \qquad \text{Eq. (1)}$$

Eq. (1)에서 i, j는 각각 입력 층, 은닉 층의 뉴런 인덱스, n 은 입력벡터의 차수 w 는 입력-은닉 층 사이의 연결 가중치, b_1 는 은닉 층 바이어스 뉴런의 연결 값을 의미한다. $b_{1j} = w_{0j}$, $x_{0=1}$ 을 가정하여 벡터의 내적으로 간단하게 표현하였다.

이 네트는 은닉 뉴런의 활성화 함수(activation function) f_1 에 대입되어 Eq. (2)와 같이 출력 y_1 을 계산한다.

$$y_j = f_1(net_j) \qquad \text{Eq. (2)}$$

활성화 함수(activation function) 로는 일반적으로 사용하는 sigmoid함수를 사용하였다[7-14].

2. 특징 추출

위 에서 제시한 4 가지 음성 환경으로 구분하기 위해서 먼저 우리가 일상 생활에서 듣는 소리들의 특징을 추출하여 분류하기 위한 값들을 얻어낸다. 특징 값으로는 시간축상에 특징을 나타내는 short time energy(STE) [15,16], zero crossing rate(ZCR) [15], 주파수 특성을 나타내는 spectral roll-off(SRO) [15,16], spectral centroid(SC) [15,16], spectral flux(SF) [15,16], 그리고 음성이 저주파 대역에 주로 분포되어 있는 것을 고려하여 비선형으로 주파수 밴드를 나누어 특징 값을 구하는 mel frequency cepstrum coefficients(MFCCs) [5,15,17], mel band power(MP) [18,19]를 사용하였다. 2 가지 시간축상의 특징 값 추출 방식, 3 가지 주파수축상의 특징 값 추출 방식, 그리고 2 가지 mel 밴

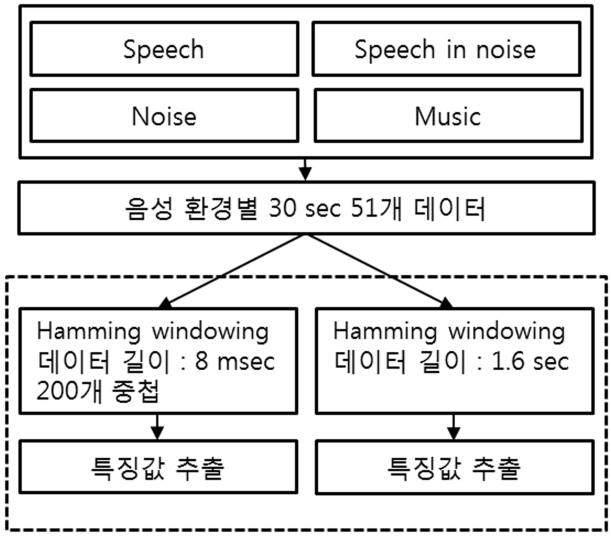

그림 1. 환경 분류 모델 생성을 위한 데이터 처리 과정
Fig. 1. Data processing for generating sound classification model

드를 나누는 방식으로 입력신호로부터 실시간으로 해밍 창 함수(hamming windowing)를 사용하여 데이터를 분할 후 특징 값을 추출하여 보청기 사용자의 환경을 구분하는 모델을 인공 신경망 모델링 기법을 사용하여 생성하였다.

3. 실험 방법

음성 환경 구분 모델을 만들기 위해서는 같은 잡음 군이라도 다양한 장소에서 녹음된 데이터가 필요했다. 이 같은 상황을 모두 포함 하기 위해서 Phonak사로부터 제공 받은 음성 및 잡음 데이터를 사용하였다[20]. 음성 환경 구분 모델을 만들기 위한 데이터는 그림 1과 같이 음성만 있는 상황, 음성과 잡음이 있는 상황, 잡음만 있는 상황 그리고 음악을 듣는 상황의 4 가지 상황에 대하여 음성 환경별로 길이가 30 sec인 51 개의 데이터를 8 msec의 길이로 분할하

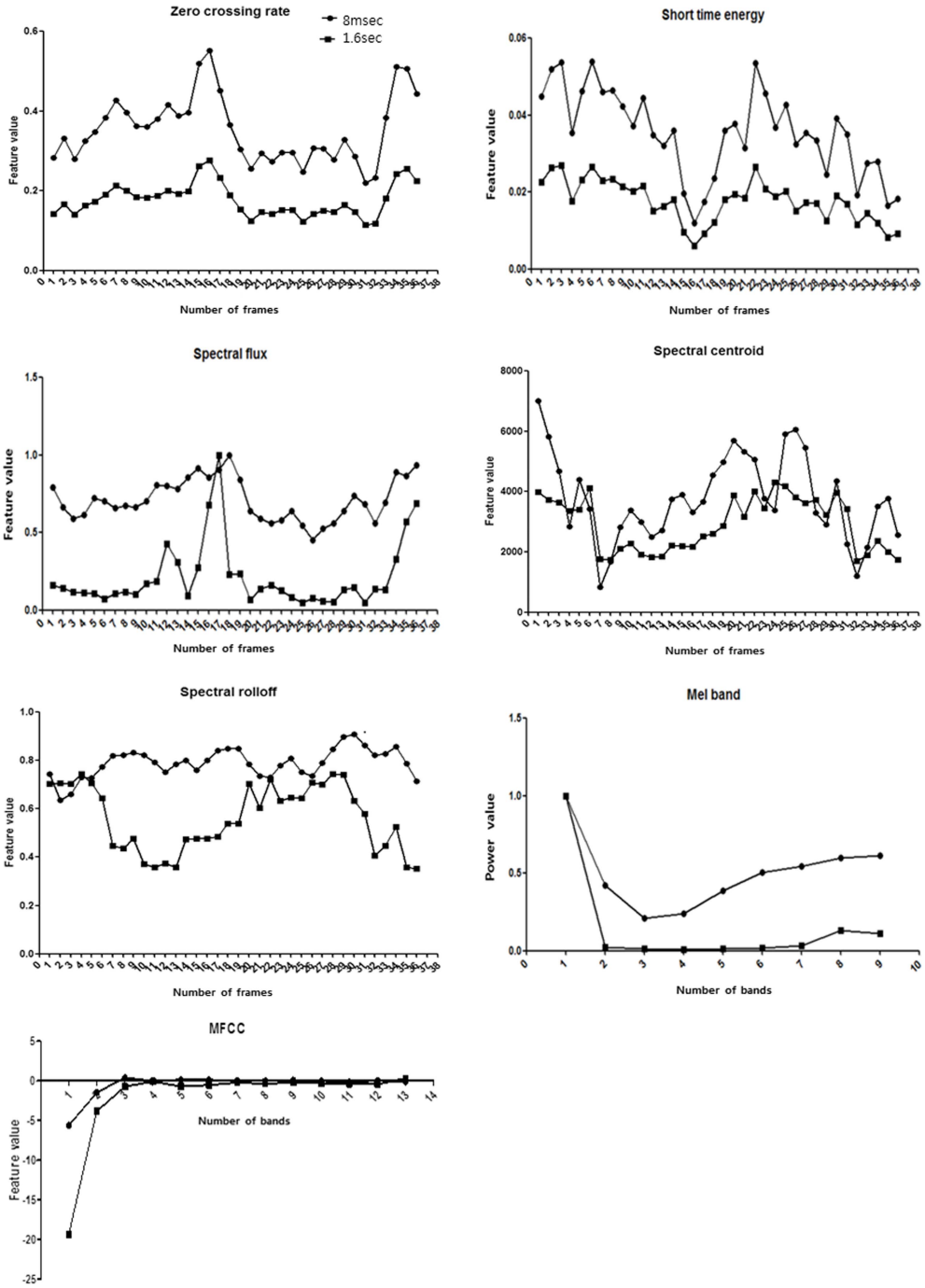

그림 2. 8 msec 와 1.6 sec 특징 값 결과 비교 ; (a) ZCR, (b) STE, (c) SF, (d) SC, (e) SRO, (f) Mel band power, (g) MFCCs
Fig. 2. the comparison of 8 msec and 1.6 sec feature values ; (a) ZCR, (b) STE, (c) SF, (d) SC, (e) SRO, (f) Mel band power, (g) MFCCs

여 각각의 환경별로 191,250 개의 데이터를 모델을 생성하는데 사용하였다[21]. 잡음이 포함 되어진 상황인 음성과 잡음이 있는 상황과 잡음만 있는 상황인 경우, 2명 이상의 사람의 말소리 잡음, 차량 실내 잡음, 도로에서의 잡음 그리고 백색잡음 이 4 가지 잡음 상황을 모두 고려해 주었다.

특징 값 추출 시 각각의 데이터의 길이는 길수록 높은 정확도의 구분 성능을 구현할 수 있지만, 실시간 알고리즘의 메모리 및 데이터 처리시간의 한계에 의하여 프레임을 길게 잡는 것에는 한계가 있다. 제안하는 환경 분류 알고리즘은 8 msec 길이의 데이터를 실시간으로 특징값을 계산한 후 이 값을 200 개씩 중첩시켜서 데이터 길이가 1.6 sec 인 환경 구분 모델과 유사한 형태의 모델을 생성하였다. 그림 1에서와 같이 비교를 위해서 1.6 sec로 데이터를 분할하여 특징 값을 추출하였고 제안하는 8 msec로 데이터를 분할 후 중첩 시켜 특징 값을 추출하여 그림 2와 같은 특징 값을 추출해 내었다[21].

제안하는 알고리즘은 컴퓨터 시뮬레이션을 통해 테스트 되고 성능을 정확도로 나타내었다. 시뮬레이션을 위해 Matlab (2011a; Mathworks Inc., Massachusetts, USA) 을 사용하였고 30 sec 인 204 개(4 가지 환경 별 51 개씩)의 데이터를 1.6 sec와 8 msec 두 가지 형태로 나누어서 특징 값을 추출한 후 4 가지 환경에 대한 분류 정확도를 검증 하였다. 204 개의 데이터 중 신경회로망 패턴 분류(Neural network pattern classification)에서 I. Kaastra 와 M. Boyd[22]에 따라 데이터를 분할하여 학습(training)으로 사용한 데이터는 68% (140 개)이고 16% (32 개)의 데이터가 시험(testing)에 사용되고 나머지 16% (32 개)의 데이터가 검증(validation)에 사용 되어 환경 분류 모델의 정확도를 평가하였다[15].

III. 결 과

1. Neuron 개수에 따른 성능 비교

최적의 뉴런 개수를 찾기 위해 특징 값을 7 개로 고정하고 인공 신경망의 뉴런의 개수들을 다르게 하여 성능 비교를 하였다. 첫 번째 은닉 층과 두 번째 은닉 층의 개수들을 다르게 하여 성능 비교를 하였다. 뉴런의 개수가 많아 질수록 좋은 성능을 보여주었지만, 뉴런의 개수가 20 개 가 넘어가면 오히려 정확도가 떨어지는 것을 볼 수 있다(그림 3). 뉴런의 개수가 많아지면 계산시간은 증가하는 것을 볼 수 있다(표 1). 표 1의 결과를 통하여 제안하는 각 은닉층의 뉴런의 개수는 20개로 하였다.

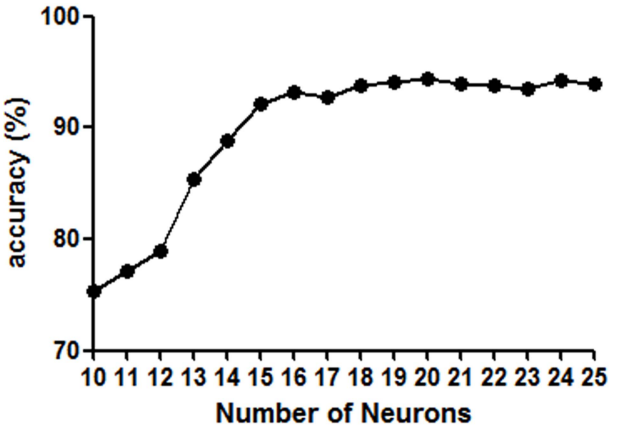

그림 3. 두 개 은닉 층에 뉴런 개수가 같을 시 개수에 따른 분류 정확도
Fig. 3. the result of classification accuracy when the number of neurons in two hidden layers were same

2. Feature 개수에 따른 성능비교

표 1을 통하여 가장 높은 정확도를 보인 두 개의 은닉 층에 대하여 각각 20개의 뉴런을 사용한 인경 신경망 모델을

표 2. 특징값 조합에 따른 성능 비교표
Table 2. The results according to the feature sets

인공신경망 모델의 특징값 조합	정확도	
	8 msec	1.6 sec
ZCR, STE, SF, SC, SRO	82.7%	83.4%
ZCR, STE, SF, SC, SRO, MP	91.3%	91.0%
ZCR, STE, SF, SC, SRO, MFCCs	88.5%	91.3%
ZCR, STE, SF, SC, SRO, MP, MFCCs	94.4%	95.2%
MP, MFCCs	84.2%	87.2%
MFCCs	69.5%	71.3%
MP	77.0%	75.9%

표 1. 7 개 특징 값 사용시 Neuron 개수에 따른 성능표
Table 1. The results according to the number of neurons when using 7 feature values

1st layer / 2nd layer	15 neuron		20 neuron		25 neuron	
	정확도	소요시간	정확도	소요시간	정확도	소요시간
15 neuron	92.1%	32 msec	92.5%	35 msec	92.4%	46 msec
20 neuron	92.9%	37 msec	94.4%	46 msec	93.3%	58 msec
25 neuron	93.2%	52 msec	92.6%	61 msec	93.9%	72 msec

사용하여 최적의 특징 값 조합을 찾고자 하였다.

시간 축 영역과 주파수 축 영역에서 볼 수 있는 특징 값들 5 가지를 묶어서 결과를 보고, 기존 연구에서 높은 성능을 보여주는 MFCCs 와 Mel Band power값을 선택적으로 특징 값으로 지정하여 모델의 정확도를 평가하였다(표 2) [5,18]. 최종적으로 8 msec의 데이터를 받아 평가 결과 두 개의 은닉 층에 대하여 각각 20 개의 뉴런을 사용하고 7 개의 특징 값을 모두 다 쓴 경우 인공 신경망 기법을 통하여 생성된 모델의 정확도는 가장 높은 94.4%를 보였다.

IV. 결론 및 고찰

보청기 사용자들 역시 일반적인 사람들과 같이 빠르게 변화하는 상황에 정확하게 적응하여야 하고 상황에 맞는 잡음 제거 알고리즘을 사용하여 불편함 없는 생활을 유지해야 한다[8]. 실시간으로 변하는 환경에 빠르게 적응하기 위해 본 연구에서는 8 msec의 프레임으로 데이터를 받아서 음성 환경 분류를 하였고 보청기에 적용이 어려운 1.6 sec 프레임으로 데이터를 받은 것과 비교 하였을 때 정확도가 7개의 특징 값을 사용하고 20개 뉴런 사용시 0.8% 의 차이를 보이며 유사한 결과를 얻어내었다. 결론적으로 실시간에서도 환경 구분 모델을 구현하면서도 높은 정확도가 유지되는 결과를 얻어 낼 수 있었다.

또한 다양한 특징 값 조합과 뉴런 개수의 조합을 통한 환경 구분 모델의 정확도를 분석하였다. 모든 특징 값을 다 사용하는 것이 가장 높은 정확도를 보였지만, MFCCs를 뺀 경우에도 91.3%로 높은 정확도를 보여 실시간 알고리즘 구현 시 계산 량이 상대적으로 많은 MFCCs는 빼고 환경 분류 모델을 구현할 수도 있고 향후 특징 값의 종류 및 조합을 늘린다면 성능이 개선될 것으로 보여진다.

뉴런의 개수 또한 늘어날수록 계산 량이 늘어 실시간 구현에 부담을 주게 되는데 두 개의 은닉 층에서 각각 20 개 뉴런을 가지는 경우 가장 높은 성능을 보여 실시간 구현 시 뉴런의 개수는 20 개 이하로 최적화 시키는 것이 계산 량도 적고 높은 구분 성능을 보일 것으로 판단되어진다.

기존에 연구되던 환경분류 알고리즘은 환경을 분류하는데 2.5 sec, 5 sec 의 데이터를 받아서 처리하는데 보청기에서 입력되는 소리가 최소 9 msec 지연이 일어나지 않아야만 알고리즘이 실시간 구동이 가능하기 때문에 본 연구에서는 8 msec 단위로 입력 신호를 받아 실시간 구현이 가능하도록 하였다[23].

또한 본 연구에서는 기존의 2.5 sec나 5 sec의 데이터 길이가 뚜렷한 기준이 없었고 1.6 sec로도 충분히 90% 이상의 정확도가 유지된다고 판단되어 데이터 길이를 1.6 sec로 하였고 적절한 환경 분류에 걸리는 시간에 대한 정의는 추후에 필요할 것으로 판단되었다.

구현된 알고리즘은 Phonak 사에서 녹음한 데이터 로써, 실제 환경에서 녹음한 데이터 와 성능을 비교할 수 없었다는 단점과 실제 보청기 마이크로폰을 통하여 들어온 음향이 아니기에 실제 소리와는 다소 차이가 있을 수도 있다. 이러한 단점들이 보완하기 위해 좀더 다양한 데이터를 통하여 모델을 생성할 필요가 있을 것이다.

Reference

[1] A. Duquesnoy, “Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons,” *J. Acoust. Soc. Am.*, vol. 74, pp. 739, 1983.

[2] J.M. Festen, and R. Plomp, “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,” *J. Acoust. Soc. Am.,* vol. 88, pp. 1725, 1990.

[3] S. Hygge, J. Ronnberg, B. Larsby, and S. Arlinger, “Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds,” *J. Speech and Hearing Research,* vol. 35, pp. 208, 1992.

[4] R. Plomp, “Noise, amplification, and compression: considerations of three main issues in hearing aid design,” *Ear and Hearing,.* vol. 15, pp. 2, 1994.

[5] E. Alexandre, L. Cuadra, L. Alvarez, M.R. Zurera, and F.L. Fereras, “Two-layer automatic sound classification system for conversation enhancement in hearing aids,” *Integrated Computer-Aided Engineering,* vol. 15, pp. 85-94, 2008.

[6] A. Bugatti, A. Flammini, and P. Migliorati, “Audio classification in speech and music: A comparison between a statistical and a neural approach,” *EURASIP J. Applied Signal Proc.,* vol. 2002, pp. 372-378, 2002.

[7] C. Freeman, Audio environment classification for hearing aids, Ontario, Canada.: Guelph Univ. Press, 2008, pp. 41-44.

[8] M.C. Büchler, Algorithms for sound classification in hearing instruments, Zurich, German,: Zurich Univ. Press, 2002, pp. 90-91.

[9] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification and Scene Analysis 2nd ed, NJ, US: Wiley-Interscience Press, 2000.

[10] B.D. Barkana, and I. Saricicek. “Environmental Noise Source Classification Using Neural Networks,” in *Information Technology: New Generations Seventh International Conference,* 2010, pp. 259-263.

[11] P. Dhanalakshmi, S. Palanivel and V. Ramalingam, “Classification of audio signals using aann and gmm,” *Applied Soft Computing,.* vol. 11, pp. 716-723, 2011.

[14] D. Changhong, “Matlab neural network and application,” *National Defense Industry Press,.* vol. 1, 2005.

[15] H. Subramanian, “AUDIO SIGNAL CLASSIFICATION,” M. Tech. Credit Seminar Report, 2004.

[17] F. Beritelli, and R. Grasso, “A pattern recognition system for environmental sound classification based on MFCCs and neural networks,” *Signal Proc. Communication Systems 2nd Int. Conf.,* 2008, pp.1-4.

[18] S.H. Yook, Y.S. Ji, H.P. Kim, D.B. Shin, and I.Y. Kim, “Envi-

ronmental Noise Classification System for Adaptive Noise Reduction Algorithm in Hearing Aids," *The Korea Society of Medical & Biological Engineering*, 2009.

[19] S.S. Stevens, and J. Volkmann, "A scale for the measurement of the psychological magnitude pitch," *J. Acoust. Soc. Am.,* vol. 8, pp. 185-190, 1937.

[20] M. Büchler, S. Allegro, S. Launer, and N Dillier, "Sound classification in hearing aids inspired by auditory scene analysis," *EURASIP J. Applied Signal Proc.,* vol. 2005, pp. 2991-3002, 2005.

[21] L.R. Rabiner, and B. Gold, "Theory and application of digital signal processing," *Englewood Cliffs,* NJ, US: Prentice-Hall Inc. Press, 1975, pp. 777.

[22] I. Kaastra, and M. Boyd, "Designing a neural network for forecasting financial and economic time series" *Neurocomputing*, vol. 10, pp. 215-236, 1996.

[22] J.M. Kates, and K.H. Arehart, "Multichannel dynamic-range compression using digital frequency warping" *Eurasip J. on Applied Signal Proc.*, vol. 2005, pp. 3003-3014, 2005.