응급구조에서의 음향데이터 분석을 위한 음성 부재구간 검출 기술

황승현, 장준혁
한양대학교 전자컴퓨터통신공학과
e-mail:jchang@hanyang.ac.kr

Voice inactivity detection for Analysis of Acoustic data of Emergency Rescue

Seng Hyun Huang, Joon-Hyuk Chang
Dept of Electronics Computer Engineering, Hanyang University

요 약

본 논문에서는 응급구조의 신고 상황에서의 수보자의 보다 정확하고 신속한 대응를 위하여 수화자의 음향환경을 분석하여 주변상황에 대한 정보를 알고자 심화 신경망 기반의 음성 부재구간 검출 기법을 제안한다. 제안한 알고리즘은 음성 신호에서의 23차의 Mel-filter bank를 추출하고 이를 심화 신경망 기법을 이용하여 음성 부재구간을 검출한다. 객관적인 성능 평가를 위해 제안된 기법은 실제 응급구조 상황에서 평가되었으며, 기존의 음성검출기를 이용한 음성 부재구간 검출 성능에 비하여 향상된 성능을 보였다.

1. 서론

응급구조 상황에서의 특수성 때문에 수화자의 발화만으로 수보자가 정확한 상황을 파악하고 대응하기에 어려움이 있다. 수보자의 보다 정확하고 신속한 대응을 위하여 수화자의 발화 정보뿐만 아니라 음향환경 분석을 통한 환경정보가 필요하다. 이러한 음향환경 정보는 수화자의 음성신호가 존재하지 않는 음성 부재구간을 이용하면 보다 정확한 정보 획득이 가능하다. 음향환경 정보는 음향환경의 모델링 특성이 뛰어난 MFCC나 Gabor-filter 기반의 특징벡터 추출 기술의 성능이 뛰어난 것으로 알려져 있다 [1]. 이러한 특징벡터 추출 기술은 Mel-filter bank를 이용하여 특징벡터를 추출하기 때문에 Mel-filter bank 기반의 음성 부재구간 검출기술을 이용하여 보다 정확한 음향환경 분석 및 계산량을 줄일 수 있다.

전통적인 음성 검출기술은 음성 부재구간에서 음성인식기의 실행을 중단하여 소비전력 및 오류를 줄이는 역할을 한다. 그렇기 때문에 음성 부재구간보다 음성 존재구간에 우선순위를 두게 되어 상대적으로 음성 존재구간의 인식확률에 비하여 음성 부재구간의 인식확률이 낮아지게 된다.

본 논문에서는 응급구조 상황에서의 음성 부재구간 검출을 위하여 23차의 Mel-filter bank를 기반으로 심화 신경망 [2] 기반의 음성 부재구간 검출기술을 제안한다. 제안된 기법은 기존의 음성 검출기술을 이용한 방법보다 우수한 성능을 보였다.

2. 제안된 음성 부재구간 검출기술

음성 부재확률을 추정하기 위한 심화신경망에 입력되는 특징벡터의 추출을 위하여 각 프레임의 특징벡터는 입력된 음향신호를 Mel주파수 스케일의 스펙트로그램으로 다시 표현할 수 있다. 이를 위하여 입력된 음성신호 $y(t)$에 해밍윈도우 w를 n_s만큼 이동하면서 적용한다. 해밍윈도우가 적용된 음성신호를 discrete-time Fourier transform (DTFT)를 통해 주파수 영역에서 다음과 같이 표현할 수 있다.

$$Y_{l,k}=\sum_{n=0}^{N-1} y(n+l\times n_s)w(n)e^{-j2\pi kn/N},\ 0\leqq k\leqq N-1 \quad (1)$$

Mel필터뱅크 계수는 주파수도메인에서 삼각형 모양의 Mel필터뱅크 $F_{k,m}$를 해밍윈도우가 적용된 음성신호에 적용한 후 로그를 취하여 다음과 같이 구할 수 있다.

$$\hat{Y}_{l,k}=\log\left(\sum_{k=0}^{N-1}|Y_{l,k}|\times F_{k,m}\right) \quad (2)$$

식 (2)에서 l과 k는 각각 프레임 인덱스와 주파수 성분을, m은 filter 인덱스를 나타낸다.

본 논문에서는 음성 부재구간을 검출하기 위하여 은닉층이 3개인 심화신경망을 도입하였다. Mel-filter bank를 통해서 추출된 특징벡터는 심화신경망으로 입력되며 도입된 심화신경망을 이용한 특징벡터의 맵핑함수는 다음과 같이 표현될 수 있다.

$$p=y(y(y(YW_1+B_1)W_2+b_2)W_3+b_3)W_{out}+b_{out} \quad (3)$$

여기서, W 와 b 는 가중치 파라미터와 바이어스 파라미

<표 1> 제안하는 음성 부재구간 검출 성능비교

	P_E	P_{MA}	P_{FA}
AMR VAD option 2	13.43	1.66	39.37
Proposed	9.43	6.44	16.03

터를 나타낸다. 심화신경망의 출력 값으로부터 음성 부재 확률을 추정하기 위하여 soft-max 기법을 이용한다..

$$\hat{p_i} = \frac{\exp(p_i)}{\sum_{j=1}^{N}\exp(p_j)} \quad (4)$$

여기서 p_i 와 $\hat{p_i}$ 은 각각 심화신경망의 I번째 출력노드 값과 이로부터 추정된 확률을 나타낸다. 최종적인 음성 부재구간 검출은 음성의 존재확률과 부재확률 중 가장 확률이 높은 상황으로 결정된다.

3. 실험 및 결과 비교

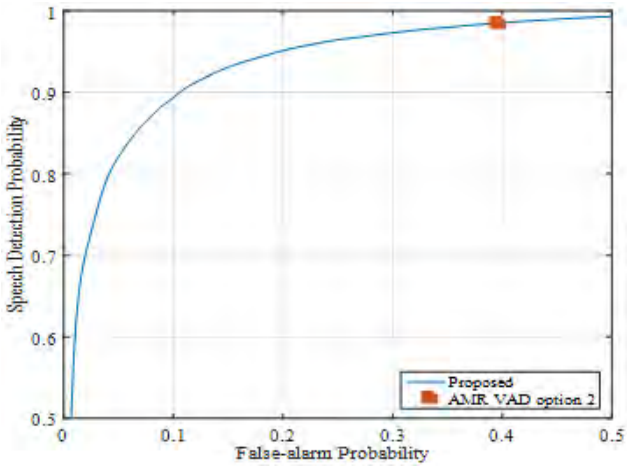
본 논문에서 제안하는 심화신경망 기반의 음성 부재구간 검출기술을 구현 및 평가하기 위하여 총 2400초의 8 kHz로 샘플링된 응급구조의 신고전화 데이터를 이용하였다. 또한 음성 존재구간과 부재구간의 구분을 위하여 음성의 존재와 부재구간을 10 ms 마다 수동으로 표시하였으며 심화신경망의 학습을 위하여 전체 데이터의 75 % 에 해당되는 데이터를 학습하였으며 음성 검출 성능을 평가하기 위하여 전체 데이터의 25 % 에 해당되는 음성데이터를 사용하였다. 음성 부재구간의 추정을 위한 심화신경망의 은닉유닛 수는 500, 500, 500 개로 각각 설정하였으며 심화 신경망의 pre-training을 위하여 0.002의 학습률로 80번 반복 학습하였다. 또한, 심화신경망의 fine-tuning을 위하여 0.005의 학습률로 50번 반복 학습하였다.

제안하는 음성 부재구간 검출기술의 평가를 위하여 기존 방법과의 P_{MA} (probability of missing alarm), P_{FA} (probability of false amalrm)를 비교하였다.

표 1은 기존 음성검출기술을 사용한 음성 부재구간 검출과 제안하는 음성 부재구간 검출기의 성능을 나타내었다. 특히 기존 AMR코덱의 음성검출기 옵션2를 이용한 방법에 비하여 음성부재구간의 검출확률인 P_{nh}에서 높은 정확도를 보였다. 제안하는 기법이 기존의 방법보다 향상된 음성 부재구간 검출성능을 보였다.

4. 결론

본 논문에서는 음급구조 신고 상황에서의 음향환경 분석을 위하여 Mel-filter bank를 특징벡터로 하는 심화신경망을 이용한 음성 부재구간 검출 기법을 제안하였다. 제안된 음성 부재구간 검출기술은 실제 응급구조 신고 상황에서 기존의 기법에 비하여 우수한 성능을 보였다.



(그림 1) 제안된 음성 부재구간 검출기와 기존방법과의 ROC그래프 비교

감사의 글

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2014R1A2A1A10049735). 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [R0126-15-1119, 음성·음향 분석 기반 상황 판단 솔루션 기술 개발]

참고문헌

[1] J. Schroder, B. Cauchi, M. R. Schadler, N. Moritz, K. Adiloglu, J. Anemuller, S. Doclo, B. Kollmeier, and S. Godtze “Acoustic event detection using signal enhancement and spectro-temporal feature extraction,” in *Proc. IEEE Workshop Appicat. SIgnal Process. Audio Acoust. (WASPAA)*, 2013

[2] G. Hinton and R. Salakhutdinov, “Reducing the dimensionality of data with neural network,” Science, vol. 313, no. 5786, pp. 504-507, Jul. 2006.