

# 응급구조 음향데이터 분석을 위한 Gabor 필터뱅크 기반의 특징추출 알고리즘에 대한 연구

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

## A study on Gabor Filter Bank-based Feature Extraction Algorithm for Analysis of Acoustic data of Emergency Rescue

Inyoung Hwang, Joon-Hyuk Chang  
Dept. of Electronic and Computer Engineering, Hanyang University

### 요 약

본 논문에서는 응급상황이 신고되는 상황에서 수보자에게 전달되는 신고자의 주변음향신호로부터 신고자의 주변상황을 추정하기 위하여 음향의 주파수적 특성 및 변화특성의 모델링 성능이 뛰어난 Gabor 필터뱅크 기반의 특징벡터 추출 기술 및 분류 성능이 뛰어난 심화신경망을 도입한다. 제안하는 Gabor 필터뱅크 기반의 특징벡터 추출 기법은 비음성 구간 검출기를 통하여 음성/비음성을 구분한 후에 비음성 구간에서 23차의 Mel-filter bank 계수를 추출한 후에 이로부터 Gabor 필터를 이용하여 주변상황 추정을 위한 특징벡터를 추출하고, 이로부터 학습된 심화신경망을 통하여 신고자의 장소적 정보를 추정한다. 제안된 기법은 여러 가지 시나리오 환경에서 평가되었으며, 우수한 분류성능을 보였다.

### 1. 서론

신고자가 응급상황을 신고할 때, 수보자에게는 신고자의 음성 및 신고자 주변의 음향이 전달되며, 이로부터 짧은 시간 내에 상황을 정확히 파악하는 것이 매우 중요하다. 하지만 긴박한 상황에 놓이게 되면 신고자가 당황한 상태이기 때문에 짧은 시간에 응급상황을 정확하게 전달하는데 한계가 있어 수보자가 신고상황을 정확히 파악하고 이에 대응하는데 어려움이 있다. 특히, 신고자의 비음성 구간에서 들을 수 있는 신고자 주변의 음향은 신고자의 주변 환경을 추정하는데 도움이 되지만, 일반적으로는 수보자가 이를 듣고 판단하는 것은 매우 어렵다. 신고자의 주변음향으로부터 주변 환경을 추정해주는 것이 자동화되면 수보자의 상황판단 및 대응에 매우 유리하며, 특히 장난전화인지 여부를 판단하는데도 매우 유용하다.

Gabor [1]가 제안한 Gabor 필터뱅크 기반의 음향의 특징벡터 추출 기술은 음향의 주파수적 특성뿐만 아니라, 변화적 특성을 잘 모델링 할 수 있는 것으로 알려져 있으며, 영상처리, 음성/음향 신호처리 등에 도입되어 매우 우수한 객체인식 및 음향 이벤트 검출성능을 보였다. 또한, Hinton과 Salakhutdinov [2]가 제안한 심화신경망 (deep neural network)은 다수의 은닉 층을 가지는 신경망을 pre-training 과정을 통해서 초기화하고 이를 통해서 얻어진 가중치 및 바이어스 파라미터를 신경망의 초기 값으로써 fine-tuning함으로써 지역최적화와 과학습문제를 방지함으로써 고성능 분류성능을 도출할 수 있다.

본 논문에서는 신고자의 주변음향으로부터 신고자의 장

소적 정보를 보다 잘 추정하기 위하여 음향에 대한 모델링 성능이 뛰어난 Gabor 필터기반의 특징벡터 추출 기법과 분류 성능이 뛰어난 심화신경망을 이용한 신고자의 주변 상황 추정 기술을 제안한다. 제안하는 기법은 다양한 시나리오 환경에서 평가되었으며, 우수한 분류 성능을 보였다.

### 2. Gabor 필터뱅크 기반의 특징추출기술

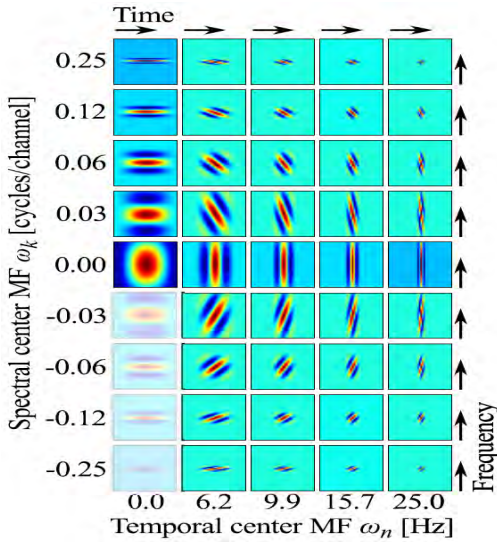
Gabor 필터뱅크 기반의 특징추출 기술은 웨이블릿 변환의 일종으로, Gabor 필터뱅크 함수는 가우시안 함수를 통해서 변조된 사인파로 정의된다. 각 프레임의 특징벡터는 입력된 음향신호의 재표현된 Mel주파수 스케일의 스펙트로그램과 Gabor 필터뱅크를 이용하여 구할 수 있다. 이를 위하여 입력된 음향의 이산신호  $y(t)$ 에 해밍윈도우  $w$ 를  $n_s$ 만큼 이동하면서 적용한다. 해밍윈도우가 적용된 입력 음향신호를 이산푸리에변환 (discrete Fourier transform) 을 통해서 주파수도메인으로 다음과 같이 표현할 수 있다.

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

본 논문에서는 23차의 Mel필터뱅크를 사용한다. Mel필터뱅크 계수는 주파수도메인에서 해밍윈도우가 적용된 음향신호에 삼각형 모양의 Mel필터뱅크  $F_{k,m}$ 을 적용한 후 로그를 취하여 아래와 같이 구할 수 있다.

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

Mel필터뱅크 특징벡터 영역에서 Gabor 필터뱅크는 다음과 같이 정의된다.



(그림 1) 2차원 Gabor 필터뱅크 예

$$h(x,y;f,\theta) = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \times \exp \left[ -\frac{1}{2} \left( \frac{(x')^2}{\sigma_x^2} + \frac{(y')^2}{\sigma_y^2} \right) \right] \exp(2\pi j f x') \quad (3)$$

여기서,  $x' = x \sin\theta + y \cos\theta$ 이고  $y' = x \cos\theta - y \sin\theta$ 이며,  $f$ 는 채널의 필터 중심 주파수,  $\theta$ 는 채널의 필터방향,  $\sigma$ 는 가우시안 성분의 분포를 나타낸다. 그림 1은 식 (3)을 통해서 구해진 Gabor 필터뱅크의 진폭을 보여주고 있다.

Gabor 필터기반의 특징벡터 추출 기술은 Mel필터뱅크 계수로부터 Gabor 필터뱅크  $h$ 를 적용하여 구할 수 있다.

$$G_{l,m}(m_0, l_0, f_m, \theta_m) = \sum_{\mu, \lambda} \hat{Y}_{l,k} h(\mu + m, \lambda + l; m_0, l_0, f_m, \theta_m). \quad (4)$$

Gabor 필터뱅크를 통해서 추출된 신고자 주변음향으로부터 추출된 특징벡터는 신고자의 장소적 정보추정을 위한 심화신경망으로 입력되며, 아래와 같이 표현된다.

$$y(l) = g(g(g(G(l)W_1 + b_1)W_2 + b_2)W_3 + b_3)W_4 + b_3 \quad (5)$$

심화신경망의 출력 값으로부터 각 장소에 대한 확률을 추정하기 위하여 soft-max 기법을 적용한다.

$$\hat{y}_i(l) = \frac{\exp(y_i(l))}{\sum_{j=1}^N \exp(y_j(l))} \quad (6)$$

여기서,  $y_i$ 와  $\hat{y}_i$ 는 심화신경망의  $i$ 번째 출력노드의 값과 이로부터 추정된 확률을 각각 의미한다.

최종적인 신고자의 장소적 상황추정 결과는 확률이 가장 높은 상황으로 결정된다.

### 3. 실험 및 결과

본 논문에서 제안하는 Gabor 필터뱅크에 기반하는 음향의 특징벡터 추출 기법 및 이를 이용한 장소적 환경 분류 기술의 성능을 평가하기 위하여 화재에 대한 신고상황을 가정한 스크립트를 작성하고, 이를 길거리, 사무실, 지하철역에서 발생하는 약 300초의 음성을 각각 8000 Hz의 샘플링주파수로 녹음했다. 핸드폰을 통해서 신고하는 상황을

(표 1) 제안하는 특징벡터 추출 기법 및 심화신경망을 이용한 신고자의 장소적 환경 분류결과 (%)

	길거리	사무실	지하철역
길거리	99.66	0.04	0.00
사무실	6.28	85.48	8.24
지하철역	11.35	30.71	57.94

고려하기 위하여 저대역의 AMR (adaptive multi-rate) 복부호화기를 이용하여 부호화 및 복호화 하였다. 또한, 음성/비음성을 구분하기 위하여 AMR복화기의 음성검출기를 이용하여 음성구간과 비음성구간을 구분하였다.

장소적 정보추정을 위한 심화신경망의 학습을 위하여 상황별 240초의 음성/음향 데이터를 30ms 길이의 각 프레임마다 라벨링하였으며, 심화신경망은 은닉층의 수는 3개, 각 은닉 층의 은닉유닛 수는 1024개로 설정하였다. 또한, 활성화함수는 시그모이드 함수를 적용했다. 심화신경망의 pre-training과정에서는 0.005의 학습율로 contrastive divergence 1 알고리즘 [3]을 50번 반복하여 학습하였으며, fine-tuning과정에서는 conjugate gradient에 기반하는 Line 탐색 알고리즘을 이용하여 100번 반복 학습했다.

본 논문에서 제안하는 신고자의 장소적 정보 추정을 위한 Gabor 필터뱅크 기반의 특징벡터 추출기법 및 이를 이용한 심화신경망 기반의 분류기술의 성능을 평가하기 위하여 각 상황별 60초의 음성/음향 데이터를 30ms길이의 각 프레임마다 라벨링했다. 이로부터 Gabor 필터뱅크를 이용하여 특징벡터를 추출하고 분류를 위한 심화신경망의 결과와 라벨 정보를 비교하여 성능을 평가했다. 표 1은 신고자의 장소적 정보추정의 결과의 혼동행렬을 보여주고 있다. 지하철역 환경에서는 상대적으로 낮은 정확도를 보였으나, 평균적으로 81.13%의 우수한 분류성능을 보였다.

### 4. 결론

본 논문에서는 응급구조가 필요한 상황이 신고되는 상황에서 수보자에게 전달되는 신고자의 주변 음향을 잘 모델링하여 분석하기 위하여 음향모델링 성능이 뛰어난 Gabor 필터뱅크 기반의 특징벡터 추출 기술과 분류 성능이 뛰어난 심화신경망을 이용한다. 제안된 Gabor필터기반의 특징벡터 추출 기법 및 심화신경망 기반의 분류기술은 길거리, 사무실, 지하철역 환경에서 평가되었으며, 지하철역에 대해서는 상대적으로 낮은 분류성능을 보였으나, 평균적으로 우수한 신고자의 장소적 정보추정 기술을 보였다.

### 5. 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] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, vol. 93, pp. 429-459, 1946.
- [2] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural network." *Science*, vol 313, no. 5786, pp. 504-507, 2006.
- [3] G. Hinton and S. Osindero, T. A. The, "A faster learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527-1554, 2006.