

논문 2011-48SP-5-14

청각장애인을 위한 상황인지기반의 음향강화기술

(Sound Reinforcement Based on Context Awareness for Hearing Impaired)

최 재 훈*, 장 준 혁**

(Jae-Hun Choi and Joon-Hyuk Chang)

요 약

본 논문에서는 청각장애인을 위한 음향 데이터를 이용한 음향강화 알고리즘을 Gaussian Mixture Model (GMM)을 이용한 상황인지 시스템 기반으로 제안한다. 음향 신호 데이터에서 Mel-Frequency Cepstral Coefficients (MFCC) 특징벡터를 추출하여 GMM을 구성하고 이를 기반으로 상황인지 결과에 따라 위험음향일 경우 음향강화기술을 제안한다. 실험결과 제안된 상황인지 기반의 음향강화 알고리즘이 다양한 음향학적 환경에서 우수한 성능을 보인 것을 알 수 있었다.

Abstract

In this paper, we apply a context awareness based on Gaussian mixture model (GMM) to a sound reinforcement for hearing impaired. In our approach, the harmful sound amplified through the sound reinforcement algorithm according to context awareness based on GMM which is constructed as Mel-frequency cepstral coefficients (MFCC) feature vector from sound data. According to the experimental results, the proposed approach is found to be effective in the various acoustic environments.

Keywords: 음향 강화, 상황 인지, Gaussian Mixture Model (GMM), Mel-Frequency Cepstral Coefficients (MFCC)

I. 서 론

청각장애인의 숫자가 통상 인구의 3% 정도의 큰 비중을 차지하고 있으며 앞으로의 사회가 고령화 사회로 옮겨감에 따라 노년인구가 늘어나고 난청인구가 증가하고 있다^[1]. 또한 어린 때부터 휴대용 오디오기기 및 휴

대폰의 사용이 증가하면서 청각에 손상을 입어 정상적인 나이보다 일찍 난청이 될 가능성이 커짐에 따라 사회복지적인 차원에서 청각장애인을 위한 통신기기 및 보조기기의 개발이 중요하다.

특히 위험한 상황이 청각장애인에게 닥쳤을 때 인지할 수 있는 감각이 시각에 국한되므로 위험한 상황에만 효과적으로 음향정보를 전달하는 보조기기의 개발이 중요하다. 이러한 기술을 개발하기 위해서는 특정 음역대의 소리만 증폭하는 기술이 아니라 음향 정보를 통한 상황인지 시스템이 필수적이다. 지금까지의 상황인지의 관한 연구는 일반적으로 Mel-Frequency Cepstral Coefficient (MFCC)를 기반의 Hidden Markov Model (HMM)을 사용한 연구가 이루어졌다^[2-5].

본 논문에서는 청각장애인을 위한 MFCC 특징벡터를 이용한 상황인식 기반의 음향강화 알고리즘을 제안한다. 입력된 음향신호에서 MFCC 특징 벡터를 추출하

* 학생회원, 인하대학교 전자공학부
(Department of Electronics Engineering, Inha University)

** 정회원, 한양대학교 융합전자공학부
(Department of Electronic Engineering, Hanyang University)

※ 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였음. [KI001824, 장애인 및 고령자를 위한 Digital Guardian 기술개발] 그리고, 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2009-0085162)

접수일자: 2011년1월18일, 수정완료일: 2011년5월11일

여 여러 가지 패턴을 구분하는데 강인한 Gaussian Mixture Model (GMM) 패턴인식 알고리즘을 사용하여 도출한 상황인지 결과에 따라 위험음향일 경우 음향신호를 강화한다. 실험결과 제안된 상황인지 기반의 음향강화 알고리즘이 다양한 음향학적 환경에서 우수한 성능을 보인 것을 알 수 있었다.

II. MFCC 특징벡터를 이용한 상황인지 시스템

1. Mel-Frequency Cepstral Coefficients (MFCC)

인간이 음성을 인지할 때 각 주파수 성분을 선형적으로 인지하지 않고 비선형적인 Mel 스케일로 음성을 인지한다. Mel 스케일은 사람이 인지하는 톤의 변화를 측정하는 단위로, 사람의 청각 특성을 반영하고 있기 때문에 Mel 스케일과 캡스트럼을 적용한 MFCC 특징 파라미터를 음성 신호 기반의 인식 시스템에서 많이 사용되고 있다. 그림 1은 MFCC 특징벡터를 추출하는 과정을 보여주고 있으며 이 과정을 통해 도출한 최종 결과식은 아래와 같다^[6].

$$x_k = \sum_{t=1}^N X'(t) \cos(k \frac{\pi}{t} (t - \frac{1}{2})) \text{ for } t = 1, 2, \dots, N \quad (1)$$

여기서 c_k 와 M 은 각각 k 차 MFCC 그리고 필터뱅크의 수이다. 본 논문에서는 13차의 MFCC를 특징벡터로 사용하였고 $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ 와 같이 정의한다.

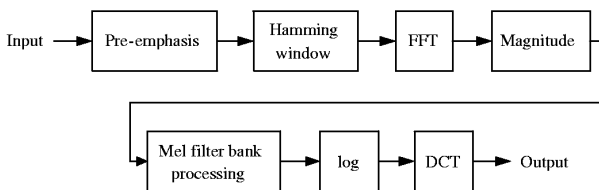


그림 1. MFCC 특징 벡터 추출 과정

Fig. 1. The block diagram of MFCC calculation process.

2. Gaussian Mixture Model (GMM)

GMM은 주어진 데이터 집합의 분포밀도를 복수개의 가우시안 확률밀도함수로 모델링하고 실제 데이터를 기반으로 최대 우도를 가지는 클래스를 선택하는 패턴인식 방법이다^[7-8]. 이미 기술한 음향신호에서 추출한 MFCC 특징벡터를 상태 열 N 개의 특징 벡터

$\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ 라 하면 M 개의 혼합성분(Mixture Component)을 가지는 가우시안 확률밀도함수의 우도(Likelihood)는 다음과 같이 주어진다.

$$p(x_t | \lambda) = \sum_{i=1}^M p_i b_i(x_t) \quad (2)$$

$$b_i(x_t) = \frac{\exp\left\{-\frac{1}{2}\right\}}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}}} (x_t - \mu_i)^T (\Sigma_i)^{-1} (x_t - \mu_i) \quad (3)$$

$$0 \leq p_i \leq 1, \sum_{i=1}^M p_i = 1 \quad (4)$$

그리고 GMM모델의 x_t 번째 성분 파라미터 λ 는 가우시안 혼합 성분 밀도의 가중치 (p_i : Mixture Weight), 평균 벡터 ($\vec{\mu}_i$: Mean Vector) 그리고 공분산 행렬(Σ_i : Covariance Matrix)로 다음과 같이 구성된다.

$$\lambda = \{p_i, \vec{\mu}_i, \Sigma_i\}, i = 1, \dots, M \quad (5)$$

$$p_i = \frac{1}{T} \sum_{t=1}^N p(i | x_t, \lambda) \quad (6)$$

$$\vec{\mu}_i = \frac{\sum_{t=1}^N p(i | x_t, \lambda) x_t}{\sum_{t=1}^N p(i | x_t, \lambda)} \quad (7)$$

$$\Sigma_i = \frac{\sum_{t=1}^N p_i b_i(x_t) x_t x_t^T - \vec{\mu}_i \vec{\mu}_i^T}{\sum_{t=1}^N p_i b_i(x_t)} \quad (8)$$

여기서 i 번째 성분의 사후확률(A Posteriori Probability)은 다음과 같이 주어진다.

$$p(i | x_t, \lambda) = \frac{p_i b_i(x_t)}{\sum_{k=1}^M p_k b_k(x_t)} \quad (9)$$

본 논문에서는 Expectation Maximization(EM)을 사용해 최적 모델 λ 를 추정하며 $p(x | \lambda') \geq p(x | \lambda)$ 가 되는 새로운 모델 λ' 이 정해진 문턱 값(Threshold)에 도달 할 때까지 반복하여 모델을 설정한다^[9-10]. 실제로

구성된 각 클래스별 모델의 실제 데이터의 특징벡터를 입력받아 환경플래그(Environment Flag)를 다음과 같이 선택한다^[11].

$$EF = \arg \max_{1 \leq s \leq k} \sum_{t=1}^N \log p(\vec{x}_t | \lambda_s), k = \text{상황의 수} \quad (10)$$

III. 상황인지기반의 음성강화기술

잡음이 섞인 오염된 음성 신호에서 잡음 신호와 실제 음성 신호를 구분하기 위해 Soft Decision에 기반한 음성 존재와 음성 부재의 확률을 적용하여, 오염된 원단 음성 신호에서 잡음 신호 전력과 실제 음성 신호 전력을 정확하게 추정하는 것이다. 이를 바탕으로 잡음이 섞이지 않은 깨끗한 음성 신호만을 추정하여, 근단의 배경 잡음 전력과 연산을 통해 얻어진 이득만큼 추정된 깨끗한 음성 신호만을 증폭하여 근단 화자가 듣게 되는 오염된 음성 신호의 명료성을 강화 시키는 것이다. 따라서 원단에서 입력되는 오염된 음성 신호에서 깨끗한 음성 신호 전력의 정확한 추정이 가장 중요하다.

원단에서의 깨끗한 음성 신호 $s(t)$ 과 잡음 신호를 $d(t)$ 라 한다면 오염된 음성 신호를 $y(t)$ 으로 나타낼 수 있으며, 각각의 성분을 DFT(Discrete Fourier Transform)을 통해서 주파수 축으로 나타내면 다음과 같이 나타낼 수 있다.

$$Y(t, k) = S(t, k) + D(t, k) \quad (11)$$

여기서 $Y(t, k)$, $S(t, k)$, $D(t, k)$ 는 $y(t)$, $s(t)$, $d(t)$ 의 t 번째 프레임에 대한 k 번째 주파수 성분을 표시한다. 추정된 근단의 배경 잡음 전력과 오염된 원단 음성 신호에서 깨끗한 음성 신호 전력의 추정치로 구해진 이득 $G(t, k)$ 을 곱해서 강화된 음성 신호 $\hat{y}(t)$ 을 다음과 같이 나타낼 수 있다.

$$\hat{Y}(t, k) = G(t, k) \cdot Y(t, k) \quad (12)$$

추정된 근단 화자의 배경 잡음 전력과 오염된 원단 음성 신호에서 잡음 신호를 제외한, 깨끗한 음성 신호 전력의 추정을 통하여 이득 $G(t, k)$ 에 대한 식을 다음과 같이 다시 나타낼 수 있다.

$$G(t, k) = \min \left\{ \max \left\{ \sqrt{\xi \cdot \frac{\hat{\Phi}_{NV}(t, k)}{\hat{\Phi}_{SS}(t, k)}}, 1 \right\}, G_{\max} \right\} \quad (13)$$

식 (13)에서 나타낸 $\hat{\Phi}_{SS}(t, k)$ 는 오염된 원단 음성 신호에서 잡음이 섞이지 않은 깨끗한 음성 신호 전력 $\Phi_{SS}(t, k)$ 의 추정 값을 나타낸다. 기존의 Sauert와 Vary의 주파수별 SNR 복구 기법에서는 원단의 깨끗한 음성 신호 전력 $\Phi_{SS}(t, k)$ 을 구함에 있어, 반복된 실험에 의해 구해진 고정된 시간 상수 α_s 을 적용하였다. 또한 원단에 입력된 음성 신호가 오염된 음성 신호일 경우, 잡음 신호와 음성 신호가 섞여 있게 되므로, 잡음 신호와 음성 신호 모두를 증폭시키는 문제점을 가지고 있다. 그러나 제안된 음성 강화 기법에서는 오염된 음성 신호로부터 잡음 신호와 실제 음성 신호를 정확하게 추정하고, 잡음 신호를 제외한 추정된 깨끗한 음성 신호 전력만을 증폭시키게 된다. 따라서 식 (13)의 $\hat{\Phi}_{SS}(t, k)$ 는 오염된 원단 음성 신호 중에서 실제 잡음이 섞이지 않은 깨끗한 음성 신호 전력만을 추정한 값을 나타낸다. 오염된 원단 음성 신호의 깨끗한 음성 신호 전력의 추정 값 $\hat{\Phi}_{SS}(t, k)$ 는 Soft Decision에 기반하여 다음과 같이 구할 수 있다.

원단 신호의 음성 부재와 음성 존재에 대한 가설을 각각 H_0 , H_1 이라고 하면, 주파수 채널에 따라 다음과 같이 가정할 수 있다.

$$\begin{aligned} H_0 : \text{speech absence} : Y(t) &= D(t) \\ H_1 : \text{speech presence} : Y(t) &= S(t) + D(t) \end{aligned} \quad (14)$$

음성 신호와 잡음 신호의 스펙트럼이 zero-mean 복소 가우시안 분포를 보인다고 가정하면, (14)에서 제시한 가설 H_0 , H_1 에 따라 다음과 같은 확률 밀도 함수로 표현할 수 있다.

$$\begin{aligned} p(Y(t, k) | H_0) &= \frac{1}{\pi \lambda_d(t, k)} \exp \left\{ -\frac{\Phi_{YY}(t, k)}{\lambda_d(t, k)} \right\} \\ p(Y(t, k) | H_1) &= \frac{1}{\pi [\lambda_d(t, k) + \lambda_s(t, k)]} \\ &\cdot \exp \left\{ -\frac{\Phi_{YY}(t, k)}{\lambda_d(t, k) + \lambda_s(t, k)} \right\} \end{aligned} \quad (15)$$

위의 식에서 $\lambda_s(t, k)$ 과 $\lambda_d(t, k)$ 는 각각 t 번째 프레임에 대한 k 번째 주파수 성분에서의 음성과 잡음의 분산을 나타낸다. 따라서 음성의 존재와 부재에 관한 가설로

부터 주파수 채널별 음성부재확률인 LSAP는 다음과 같다.

$$\begin{aligned} p(H_0|Y(t,k)) &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k))} \\ &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k)|H_0)p(H_0) + p(Y(t,k)|H_1)p(H_1)} \\ &= \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\Lambda(Y(t,k))} \end{aligned} \quad (16)$$

여기서 $p(H_0)$ 는 음성 부재에 대한 사전 확률 (a priori probability)이고, $\Lambda(Y(t,k))$ 는 k 번째 주파수 대역의 우도비 (likelihood ratio) 로써 다음과 같이 표현된다.

$$\begin{aligned} \Lambda(Y(t,k)) &= \frac{p(Y(t,k)|H_1)}{p(Y(t,k)|H_0)} \\ &= \frac{1}{1 + \xi(t,k)} \exp\left[\frac{\gamma(t,k)\xi(t,k)}{1 + \xi(t,k)}\right] \end{aligned} \quad (17)$$

여기서 $\gamma(t,k)$, $\xi(t,k)$ 는 각각 a posteriori SNR과 predicted SNR로써 다음과 같이 나타내어진다.

$$\begin{aligned} \gamma(t,k) &\equiv \frac{\Phi_{YY}(t,k)}{\Phi_{DD}(t,k)} \\ \xi(t,k) &\equiv \frac{\Phi_{SS}(t,k)}{\Phi_{DD}(t,k)} \end{aligned} \quad (18)$$

제안된 음성 강화 기법의 알고리즘의 경우 오염된 원단 음성 신호 중에서 잡음이 섞이지 않은 깨끗한 음성 신호만이 증폭의 대상이기 때문에, 잡음 전력 $\Phi_{DD}(t,k)$ 과 음성 전력 $\Phi_{SS}(t,k)$ 의 추정이 알고리즘 성능에 중요한 역할을 하게 된다. 일반적으로, 음성검출기(VAD, voice activity detector)를 사용하여 음성 부재구간에서 잡음 전력을 갱신한다^[12]. 그러나 실제 잡음 환경이 비정상(non-stationary)인 경우, 잡음 전력은 음성의 존재 구간에서도 갱신되어야 한다. 따라서 신뢰성 있는 음성 전력 $\Phi_{SS}(t,k)$ 와 잡음 전력 $\Phi_{DD}(t,k)$ 을 추정하기 위해, 음성 전력과 잡음 전력 각각에 long-term smoothed 전력 스펙트럼을 사용하여 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \hat{\Phi}_{SS}(t+1,k) &= \zeta_s \hat{\Phi}_{SS}(t,k) + (1 - \zeta_s)\Phi_{SS}(t,k) \\ \hat{\Phi}_{DD}(t+1,k) &= \zeta_d \hat{\Phi}_{DD}(t,k) + (1 - \zeta_d)\Phi_{DD}(t,k) \end{aligned} \quad (19)$$

여기서 $\hat{\Phi}_{DD}(t,k)$ 와 $\hat{\Phi}_{SS}(t,k)$ 는 각각 $\Phi_{DD}(t,k)$ 과 $\Phi_{SS}(t,k)$ 의 추정 값이고, ζ_d 과 ζ_s 는 정상(stationary) 상태를 가정한 스무딩 파라미터로써 $0 < \zeta_d, \zeta_s < 1$ 값을 가진다. 음성 신호 $S(t)$ 와 잡음 신호 $D(t)$ 의 통계적 가정과 (19)식을 이용하여 현재 프레임에 대한 음성 신호와 잡음 신호의 전력 추정 값을 계산하면 다음 식으로 나타낼 수 있다.

$$\begin{aligned} \Phi_{SS}(t,k) &= E[|S(t,k)|^2 | Y(t,k), H_0]p(H_0 | Y(t,k)) \\ &\quad + E[|S(t,k)|^2 | Y(t,k), H_1]p(H_1 | Y(t,k)) \end{aligned} \quad (20)$$

$$\begin{aligned} \Phi_{DD}(t,k) &= E[|D(t,k)|^2 | Y(t,k), H_0]p(H_0 | Y(t,k)) \\ &\quad + E[|D(t,k)|^2 | Y(t,k), H_1]p(H_1 | Y(t,k)) \end{aligned} \quad (21)$$

여기서,

$$\begin{aligned} E[|D(t,k)|^2 | Y(t,k), H_0] &= \Phi_{YY}(t,k) \\ E[|D(t,k)|^2 | Y(t,k), H_1] &= \left(\frac{\hat{\xi}(t,k)}{1 + \hat{\xi}(t,k)}\right)\hat{\Phi}_{DD}(t,k) + \left(\frac{1}{1 + \hat{\xi}(t,k)}\right)^2\Phi_{YY}(t,k) \end{aligned} \quad (22)$$

$$\begin{aligned} E[|S(t,k)|^2 | Y(t,k), H_0] &= 0 \\ E[|S(t,k)|^2 | Y(t,k), H_1] &= \left(\frac{1}{1 + \hat{\xi}(t,k)}\right)\hat{\Phi}_{SS}(t,k) + \left(\frac{\hat{\xi}(t,k)}{1 + \hat{\xi}(t,k)}\right)^2\Phi_{YY}(t,k) \end{aligned} \quad (23)$$

이때 사용된 $\hat{\xi}(t,k)$ 는 predicted SNR로 정의되며, 추정된 잡음 전력과 음성 전력의 비로 다음과 같이 나타내어진다.

$$\hat{\xi}(t,k) \equiv \frac{\Phi_{SS}(t,k)}{\Phi_{DD}(t,k)} \quad (24)$$

predicted SNR은 잡음 전력 갱신을 위해 사용된다.

IV. 실험

제안된 음향신호를 이용한 상황인지기반의 음성강화 기술 성능을 평가하기 위해서 다양한 종류의 환경에 대한 음향 데이터를 수집하였다. 구체적으로 실제 환경에

표 1. GMM 패턴인식기반의 상황인지 결과

Table 1. The accuracy of context awareness based on GMM.

Accuracy	오토바이	경적음	천둥	대형자동차	공사장	UBM
오토바이	35.8%	0.0%	35.0%	26.3%	2.9%	0.0%
경적음	29.1%	32.7%	6.4%	10.0%	2.7%	19.1%
천둥	12.3%	0.0%	82.3%	2.5%	0.5%	2.5%
대형자동차	22.3%	0.8%	27.0%	33.2%	10.3%	6.4%
공사장	24.2%	1.2%	2.0%	15.4%	51.7%	5.5%
UBM	0.2%	0.0%	2.7%	0.7%	0.0%	96.4%

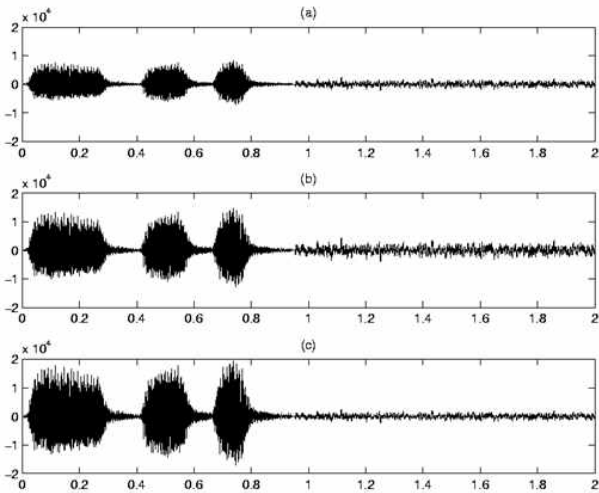


그림 2. 음향강화기술과 상황인지 기반의 음향강화 기술의 비교 (a) 입력 음향신호, (b) 음향 신호 강화 (c) 상황인지기반의 음향강화

Fig. 2. The comparison between reinforcement and context awareness based reinforcement (a) input sound signal (b) reinforced sound signal (c) reinforced sound signal based on context awareness.

대한 특징을 잘 표현하는 지와 데이터 양의 크기나 적절한 위치 등의 데이터 수집의 외부요인을 고려하였다. 또한 음향학적 환경 정보를 모을 때 한 방향이 아닌 여러 방향을 음향 데이터를 받아들일 수 있도록 일반적인 마이크가 아니라 고성능의 콘덴서 마이크를 이용하여 울트라 모바일 PC에 연결하여 데이터를 수집하였다. 수집된 데이터의 상황선정은 여론조사를 바탕으로 장애인 및 고령 난청자에게 위험이 될 수 있는 상황으로 선정하였으며 총 6가지로서 오토바이, 경적음, 천둥, 대형 자동차, 공사장, Universal Background Model (UBM) 이다. 환경 음향 데이터들은 8kHz로 샘플링되고 16bit로 양자화되어 있으며 수집된 데이터는 위에서 언급한 실제 환경의 특징 등과 같은 사항을 잘 만족시키는 20분 길이의 데이터로 구성하였다. 데이터들은 훈련을 위

해서 10분이 사용되었으며 테스트를 위해서 역시 10분의 데이터가 사용되었다. 상황인지를 위한 특징벡터로는 I장에서 기술된 13차 MFCC 특징벡터를 사용하였으며 16개의 혼합성분의 GMM을 구성하여 실험하였다.

표 1은 MFCC 특징벡터에 대한 1초 단위의 파일에 대한 상황인지를 한 결과이다. 실험결과 모든 상황에 대해 30% 이상의 인지 결과가 나타났으며 위험 상황인지 검출 확률과 UBM 검출 확률은 각각 94.1%, 96.4%이다. 천둥을 제외한 다른 위험상황에 대한 인지결과의 성능이 떨어진 것은 인식 결과가 1초 단위이므로 음향 부재구간에서의 오류 때문일 것으로 예상된다. 특히 경적음의 경우 짧은 시간동안 음향신호가 존재하며 음향 부재구간이 많아 UBM으로 오검출확률이 가장 높았다. 그림 2는 입력 음향신호파형, 음향강화기술을 적용한 음향 파형 그리고 상황인지기반의 음향강화기술을 적용한 음향 파형을 보여준다. 1초 이전의 파형은 경적음이며 1초 이후의 파형은 UBM이다. 단순히 음향강화기술을 적용하였을 경우에는 두 상황 모두 증폭하지만 상황인지기반의 음향강화기술의 경우 청각장애인 및 난청자에게 위험하다고 판단되는 상황에서만 음향을 증폭하는 것을 알 수 있었다.

V. 결 론

본 논문에서는 청각장애인을 위한 상황인지기반의 음성강화 알고리즘을 제안하였다. 제안된 알고리즘은 입력된 음향신호를 GMM을 이용하여 상황인지를 한 후 청각장애인에게 위험한 상황으로 인지되는 음향신호만을 음성강화기술을 이용하여 증폭하였다. 실험 결과 MFCC 특징벡터를 사용한 상황인식이 다양한 음향학적 환경에서 우수한 성능을 보인 것을 알 수 있었다.

감사의 글

본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였음. [KI001824, 장애인 및 고령자를 위한 Digital Guardian 기술개발] 그리고, 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2009-0085162)

참고 문헌

[1] 김종태, 신석철, 원용명, 이상철, 김원우, “청각 및 언어장애인을 위한 유선통신 대화장치 및 보조기기 개발”, 정보통신연구진흥원, [IITA] 정보통신연구진흥원 학술기사

[2] L. Ma, B. P. Milner and D. Smith, “Acoustic environment classification,” *ACM Transactions on Speech and Language Processing*, Vol.3, No.2, Jul. 2006, pp. 1-22.

[3] Y. Toyoda, J. Huang, S. Ding and Y. Liu, “Environmental sound recognition by multilayered neural networks,” in *Proc. International Conference on Computer and Information Technology*, Sep. 2004, pp. 123-127.

[4] L. Couvreur and M. Laniray, “Automatic noise recognition in urban environments based on artificial neural networks and hidden Markov models,” *InterNoise*, Prague, Czech Republic, Aug. 2004.

[5] N. Sawhney, “Situational awareness from environmental sounds,” MIT Media Lab. Technical Report, June 1997.

[6] S. Sigurdsson, K. B. Petersen, and T. Lehn-Schioler, “Mel Frequency Cepstral Coefficients: An evaluation of Robustness of MP3 Encoded Music,” in *Proc. Int. Conf. Music Inf. Retrieval*, pp. 286-289, 2005.

[7] C. M. Bishop, *Neural networks for pattern recognition*, Oxford University Press, UK, 1995.

[8] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern classification*, John Wiley & Sons, INC., 2001.

[9] G. Xuan, W. Zhang and P. Chai, “EM algorithms of Gaussian mixture model and hidden Markov model,” in *Proc. International Conference on Image Processing*, Vol. 1, Oct. 2001, pp. 145-148.

[10] J. A. Bilmes, “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixed and hidden

Markov Models,” *International Computer Science Institute, Technical Report*, TR-97-02, Apr. 1998.

[11] A. P. Dempster, N. M. Land and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society, Series B*, Vol. 39, No. 1, 1977, pp. 1-38.

[12] J. W. Shin, N. S. Kim, W. Sung, “A statistical model-based voice activity detection,” *IEEE Signal Processing Letters*, vol. 6 no. 1, pp. 1-3, Jan. 1999.

저 자 소 개



최 재 훈(학생회원)
 2007년 인하대학교 전자전기 공학부 학사.
 2008년 삼성전자 정보통신 총괄 연구원.
 2009년 인하대학교 전자공학과 석사.
 2010년 인하대학교 전자공학과 박사과정.
 2011년 한양대학교 전자컴퓨터통신공학과 박사과정

<주관심분야 : 디지털 음성신호처리>



장 준 혁(정회원)
 1998년 경북대학교 전자공학과 학사.
 2000년 서울대학교 전기공학부 석사.
 2004년 서울대학교 전기컴퓨터 공학부 박사.
 2000년~2005년 (주)넷더스 연구소장
 2004년~2005년 캘리포니아 주립대학, 산타바바라(UCSB) 박사후연구원
 2005년 한국과학기술연구원(KIST) 연구원
 2005년~2011년 인하대학교 전자공학부 조교수
 2011년~현재 한양대학교 융합전자공학부 부교수
 <주관심분야 : 음성신호처리, 오디오 신호처리, 통신 신호처리, 휴먼/컴퓨터 인터페이스>