논문 2013-50-2-21

MFCC를 이용한 GMM 기반의 음성/혼합 신호 분류

(Speech/Mixed Content Signal Classification Based on GMM Using MFCC)

김 지 은*, 이 인 성**

(Ji-Eun Kim and In-Sung Lee)

요 약

본 논문에서는 MFCC를 이용한 GMM 기반의 음성과 혼합 신호 분류 알고리즘을 MPEG의 표준 코덱인 USAC에 적용하였다. 효과적인 패턴 인식을 위해 GMM을 이용하였고, EM알고리즘을 사용하여 최적의 GMM 파라미터를 추출하였다. 제안하는 분류 알고리즘은 두 가지 중요한 부분으로 나뉜다. 첫째는 GMM을 통해 최적의 파라미터를 추출하는 것 이고, 두 번째는 MFCC 값을 이용한 패턴인식을 통해 음성/혼합 신호를 분류하였다. 제안된 알고리즘의 성능을 평가한 결과 MFCC를 이용한 GMM 기반의 제안된 방법이 기존 USAC의 방법보다 우수한 음성/혼합 신호 분류 성능을 보였다.

Abstract

In this paper, proposed to improve the performance of speech and mixed content signal classification using MFCC based on GMM probability model used for the MPEG USAC(Unified Speech and Audio Coding) standard. For effective pattern recognition, the Gaussian mixture model (GMM) probability model is used. For the optimal GMM parameter extraction, we use the expectation maximization (EM) algorithm. The proposed classification algorithm is divided into two significant parts. The first one extracts the optimal parameters for the GMM. The second distinguishes between speech and mixed content signals using MFCC feature parameters. The performance of the proposed classification algorithm shows better results compared to the conventionally implemented USAC scheme.

Keywords : USAC, MFCC, GMM, Signal Classification

Ⅰ. 서 론

모바일 기기가 다양한 기능을 가지고, 다양한 기기를 하나의 모바일 기기로 통합하는 방향으로 기술이 발전하고, 디지털 라디오, 오디오 북 등 음성과 음악신호 모두를 이용하는 응용분야의 시장이 커지면서, 음성과 오디오 신호 모두에 대해 우수한 품질을 제공하는 새로운 부호화 기술에 대해 시장의 요구가 증대되고 있다[1].

인간의 음성 생성 모델을 기반으로 하고 있는 음성 부호화 기술과, 인간의 청각 모델을 기반으로 하고 있는 오디오부호화 기술은, 음성 통신과 음악 방송 등 각각의 독립적인 서비스 영역에서 독자적으로 기술 발전을 이루어 왔다. 하지만 최근 방송과 통신이 융합하는 방향으로 기술이 발전하면서, 음성 통신과 음악 방송으로 구분되던 서비스 구조가 깨어지고, 더 이상 음성과 오디오 신호를 별도의 콘텐츠로 분리하는 것이 어렵게 되었다[2].

따라서 통합된 방식으로부터 음성과 음악을 자동으

* 학생회원, *** 정회원, 충북대학교 전파통신 공학과
(Department of Radio Engineering, ChungBuk University)
※ 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2012-0004611)
접수일자: 2012년8월24일, 수정완료일: 2013년1월18일

로 구별하는 기술의 중요성이 점점 더 높아지고 있다.

이러한 우수한 음성/혼합 신호판별 성능을 얻기 위해서는 적절한 특징 파라미터를 선택하는 것이 매우 중요하다.

시간영역에서의 특징을 이용한 ZCR(Zero Crossing Rate)와 단 구간 에너지 변화를 측정한 LSTER(Low Short-Time Energy Rate) 등이 있으며 스펙트럼 영역 특징을 이용한 파라미터로 스펙트럼의 무게중심을 이용한 SC(Spectral Centroid), 스펙트럼의 변화의 차이를 이용한 SF(Spectral Flux)와 켑스트럼 거리를 이용한 CD(Cepstral Distance), 인간의 귀가 가지는 비선형적인 주파수 특성을 이용한 MFCC(Mel Frequency Cepstral Coefficients)등이 있다. 또한 음악 장르 분류기에 사용되는 분류 알고리즘에는 SVM(Support Vector Machines), LDA(Linear Discriminant Analysis), GMM(Gaussian Mixture Model), k-NN(k-Nearest Neighbor) 등이 사용 된다[3~4].

본 논문에서는 기존의 연구 결과 잡음의 영향을 덜 받고 판별 성능이 효과적인 것으로 나타난 MFCC의 파라미터 값을 이용한 GMM 기반의 음성과 혼합 신호 분류 알고리즘을 제안하였다. GMM을 사용하게 된 동기는 MFCC 특징 벡터의 통계적 분포를 다른 평균과 공분산 행렬을 갖는 복수개의 가우시안 함수에 의해서 효과적으로 표현할 수 있기 때문이다[5~7].

따라서 MPEG에서 표준화된 음성 및 오디오 코딩 구조를 사용한 USAC(Unified Speech and Audio Coding)의 신호 분류방법에서 단점인 복잡성과 Close-loop AbS 구조 방식의 많은 연산 량의 문제점을 개선고자 하였다. 또한 기존 USAC에서는 신호가 가지는 연속적인 특성을 고려하지 않은 현재 프레임만을 가지고 신호를 분류하기 때문에 이점을 보안하고자 과거 프레임과의 상관성을 이용한 GMM 분류 방법을 통해 더욱더 정확성을 높였다.

본 논문의 Ⅱ장에서는 USAC의 신호 분류방법과 음성과 혼합 신호를 분류하기 위해 사용한 특징 값인 MFCC, 신호 분류를 위한 분류기법인 GMM에 대해서 각각 설명한다. Ⅲ장에서는 기존의 USAC과 제안하는 알고리즘의 성능을 비교 평가하고 실험 결과를 살펴보며, 마지막으로 Ⅳ장에서 결론을 맺는다.

Ⅱ. MFCC 특징 파라미터를 이용한 GMM기반의 신호분류

1. USAC의 신호 분류

USAC은 선택적으로 입력 신호의 특성에 따라 입력 신호가 음성이면 AMR-WB+ 방식[8] 기반의 선형 예측 도메인 코더가 선택되며, 입력신호가 오디오 신호인 경우에는 HE-AAC V2 방식을 기반으로 하는 주파수 도메인 코더가 선택 된다[9~10].

그 후 선형 예측 영역 부호화코더의 경우 한 번 더 스위칭 하게 되어 Closed-loop AbS구조를 이용하여 MDCT를 기반으로 하는 TCX(Transform Coded eXcitation)모드와 ACELP(Algebraic Code Excited Linear Predictor)모드로 분류된다. 그림 1에서는 간략한 USAC의 신호 분류방법을 나타내며, 표 1에서는 USAC에서의 신호 특징에 따른 3가지 압축 방법을 보여준다.

그림 1과 표 1에서와 같이 입력된 신호가 처음 오디오 신호로 판단되는 신호는 FD(Frequency Domain)모드로 음성 신호 및 혼합신호로 판단되는 신호는

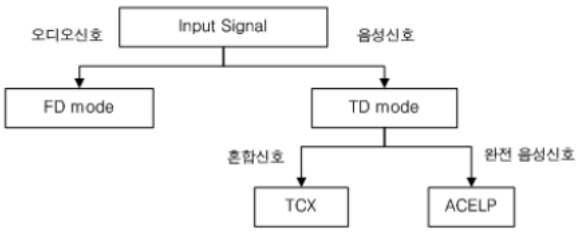

그림 1. USAC의 신호 분류 블록 다이어그램
Fig. 1. Block diagram of signal classification of USAC.

표 1. USAC Core Code의 3가지 압축 방법
Table 1. Three compression methods of USAC core code

명칭	특징	주요대상신호
AAC	MDCT 기반의 변환 부호화, 새로운 무손실 부호화 방법사용	고 전송률 오디오 신호
ACELP	LPC 기반으로 시간영역에서 부호화, 기존 AMR-WB의 ACELP모듈과 동일	음성 신호
TCX	LPC Residual 신호에 대해 MDCT 기반의 변환 부호화, Noise Fill사용	혼합 신호 및 저 전송률 오디오 신호

TD(Time Domain) 모드로 음성과 오디오 신호로 처음 분류 된 후에 음성 신호로 판단되어 TD 모드로 분류된 신호를 또 다시 완전한 음성인 경우는 ACELP 모드로 혼합 신호인 경우는 TCX 모드로 분류 되는 것을 알 수 있고, 다음과 같은 신호 분류는 전반적인 품질을 보장하기 위해서 매우 중요한 부분이 된다.

2. MFCC를 이용한 GMM 최적 파라미터 추출

가. Mel Frequency Cepstral Coefficients

음성인식에 쓰이는 특징 값으로 LPC(Linear Prediction Coefficients)나 LPS(Liner Prediction Spectrum) 등과 같은 많은 방법이 존재하지만 주파수를 피치로 이용하였을 때 잡음의 영향을 덜 받고 효과적인 것으로 나타났다[11]. 그 중 MFCC는 음성 인식에 널리 쓰이는 유효한 특징 값으로 스펙트럼 기반을 특징으로 하며 인간의 귀가 가지는 비선형적인 주파수 특성을 이용한다.

그림 2의 블록다이어그램은 MFCC를 추출하는 과정을 나타낸다. 입력 신호를 윈도우를 씌워서 프레임 단위로 나눈 후에 FFT(Fast Fourier Transform)을 이용하여 주파수 영역으로 변환하게 된다. 그 후 주파수 대역을 여러 개의 필터 뱅크로 나누고 각 뱅크에서의 에너지를 구한다.

$$X(n,w_k)=\sum_{m=-\infty}^{\infty}x[m]w[n-m]e^{-jw_k m},\quad w_k=\frac{2\pi}{N}k \tag{1}$$

식 (1)에서 나타내는 주파수로 변환된$X(n,w_k)$의 크기는 필터 시퀀스의 주파수 응답에 의해 가중화되고 이런 필터 시퀀스는 저주파수(1000Hz이하)에서는 중심주파수와 대역폭이 선형적이지만 고주파수(1000Hz이상)로 갈수록 로그 스케일로 증가하는 특성을 가지고 있다. 이것은 인간의 귀가 가지는 비선형적인 특성으로 저주파 영역의 신호에서는 민감한 반면 고주파로 갈수

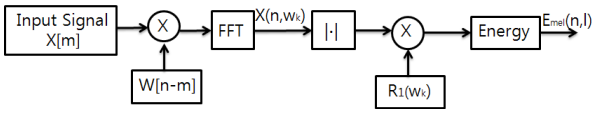

그림 2. MFCC 특징 파라미터 추출 과정
Fig. 2. Block diagram to extract MFCC.

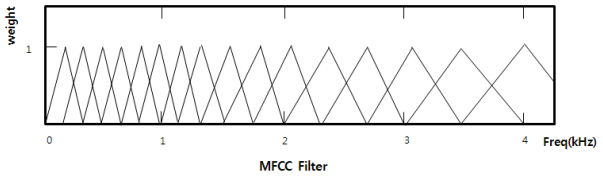

그림 3. Mel-scale 필터 뱅크
Fig. 3. Mel-scale filter bank.

록 민감하지 않은 특성을 적용한 것이다[12].

이런 특성을 가지고 있는 멜 스케일(mel-scale) 필터 뱅크를 그림 3에서 보여주고 있다. 식 (2)를 이용해서 멜 스케일을 계산하며 중심주파수는 멜 스케일로 존재하게 된다. 각 필터의 대역폭 역시 식 (3)의 critical bandwidth에 의해 결정된다.

$$Mel(f)=2595\times\log_{10}(1+\frac{f}{700}) \tag{2}$$

$$BW=\begin{cases}1000, & f<1000\\ 25+75[1+1.4(\frac{f}{1000})^2]^{0.69}, & f>1000\end{cases} \tag{3}$$

Mel 스케일 필터 뱅크의 l번째 필터의 주파수 응답을 $R_l(w_k)$라고 하면 n번 째 음성 프레임에 대한 Mel 에너지는 식 (4)로 표현 할 수 있다. L_l,H_l는 l번째 필터에서 0이 아닌 주파수 영 역의 상한, 하한 값을 의미한다.

$$E_{mel}(n,l)=\frac{1}{A_l}\sum_{k=L_l}^{H_l}\left|R_{l(w_k)}X(n,w_k)\right|^2 \tag{4}$$

$$A_l=\sum_{k=L_l}^{H_l}|R_l(w_k)|^2 \tag{5}$$

식 (5)는 다양한 대역폭을 갖는 필터들의 균일한 스펙트럼을 위한 정규화 과정이다. 따라서 Mel 에너지를 DCT(Discrete Cosine Transformation)를 적용을 통하여 멜 스케일 에너지를 무상관된 M차의 차수로 변환할 수 있다. 식 (6)을 이용하여 R개의 필터로 구성된 필터 뱅크 중 n번째 음성 프레임에 대한 m번째 계수를 계산하게 된다.

$$C_{mel}[n,m]=\frac{1}{R}\sum_{l=0}^{R-1}\log\{E_{mel}(n,l)\}\cos(\frac{2\pi}{R}lm) \tag{6}$$

나. Gaussian Mixture Model(GMM)

GMM은 그림 4와 같이 M개의 가우시안을 합하여 만들어진 모델로 음향학적인 분포를 표현함에 있어서 매우 뛰어난 것으로 나타났다. 다수의 음성/혼합 신호로부터 추출된 MFCC 특징 값을 이용하여 GMM 분류기를 훈련시킨 후 음성과 혼합 신호 분류기에 적용할 수 있다.

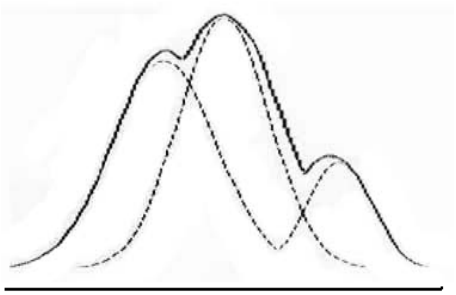

그림 4. 가우시안 혼합 모델
Fig. 4. Gaussian Mixture Model(GMM).

GMM은 식 (7)로 표현되며, M개의 요소 분포를 가중치와 함께 합산된 것이다.

$$p(x|\lambda)=\sum_{i=1}^{M}p_i b_i(x) \tag{7}$$

x는 D차 랜덤 벡터, $b_i(x)$는 요소 분포, p_i는 i번째 요소 분포에 대한 가중치를 의미한다. 이때 가중 p_i는 $\sum_{i=1}^{M}p_i=1$을 만족해야 한다. 각 요소 분포 $b_i(x)$는 식 (8)에서와 같은 μ_i의 평균 벡터와 $\sum i$의 공분산 행렬을 갖는 D차원 가우시안 분포를 갖는다고 가정 한다[13].

$$b_i(x)=\frac{1}{(2\pi)^{D/2}|\sum i|^{1/2}}exp\left\{-\frac{1}{2}(x-\mu_i)\sum i^{-1}(x-\mu_i)\right\} \tag{8}$$

M개의 가우시안 확률밀도함수의 선형 결합으로 정의되는 $p(x|\lambda)$는 각 모드에 대한 평균, 공분산, 가중치에 관한 함수이며 식 (9)에서처럼 3개의 매개 변수를 훈련 과정에서 모델링한다. 즉, 훈련 샘플을 이용하여 각 λ를 추정한다.

$$\lambda=\{w_i,\mu_i,\Sigma_i\} \quad , i=1,....M \tag{9}$$

GMM의 훈련은 Maximum Likehood 추정 방법을 이용하여 식(8)에 있는 GMM의 우도함수를 최대화하는 매개변수 λ를 추정한다. 이러한 우도함수를 최대화 할 수 있는 매개 변수를 추정하기 위해 EM(Expectation Maximization) 반복 알고리즘을 통해서 매개변수를 추정한다.

$$p(x|\lambda)=\prod_{t=1}^{T}p(x_t|\lambda) \tag{10}$$

EM 반복 알고리즘은 모델파라미터의 이전 초기 값을 설정한 후에 파라미터에 대한 현재 값을 얻기 위해서 계산을 하게 되고 이전 값과 현재의 차가 0에 가까워지면 알고리즘이 멈추게 되는 방식으로 초기 모델 λ인 혼합 모델에 $P(x|\lambda)$에 대해서 식 (10)을 만족하는 새로운 모델 $\overline{\lambda}$을 추정하며 다음번의 순환 과정에서 새로운 모델은 초기 모델이 되며 특정 오차 수준에 수렴하거나 최대 순환 횟수를 만족 할 때 까지 반복하게 된다[14].

$$p(x|\overline{\lambda})\geq p(x|\lambda) \tag{11}$$

이렇게 EM 과정을 통해 훈련된 GMM을 이용하여 새로운 신호는 확률이 최대가 되는 클래스로 분류하게 된다. 또한 입력된 신호는 다수의 프레임을 가지고 있으므로 프레임들에 대한 확률 평균값이 최대가 되는 클래스로 신호를 분류 하게 된다.

3. 음성과 혼합 신호 분류

입력 신호를 프레임 단위로 식 (4)를 이용하여 MFCC특징 벡터를 구하게 된다. 현재 정의된 확률 모델을 따르는 시스템에서 데이터가 관찰될 가능성을 우도 값이라 하며, 이 우도 값을 바로 사용하는 대신 반복된 곱셈에 의한 오차를 줄이기 위해서 로그 함수를 적용한 것을 로그 우도 함수라고 한다. 즉, 이 특징벡터를 버퍼에 저장하고, 앞서 도출된 음성과 혼합 신호의 GMM과 버퍼에 저장된 MFCC특징의 패턴의 로그 우도를 비교하여 음성과 혼합 신호를 분류한다. 이와 같은 로그 우도함수를 통해서 식 (12)과 식 (13)과 같이 신호를 분류하게 된다. 음성 신호는λ_s, 혼합 신호는 λ_m로 나타내며, 로그 우도 함수가 음성이 클 경우에는 음성신호로 혼합신호가 큰 경우에는 혼합신호로 판단하게 된다. 식 (12)은 음성신호, 식 (13)은 혼합신호를 나타낸다.

$$\sum_{n=1}^{N} \log p(x_n|\lambda_s) > \sum_{n=1}^{N} \log p(x_n|\lambda_m) \quad (12)$$

$$\sum_{n=1}^{N} \log p(x_n|\lambda_s) < \sum_{n=1}^{N} \log p(x_n|\lambda_m) \quad (13)$$

Ⅲ. 성능 평가

본 논문에서는 제안된 MFCC 특징 파라미터를 이용한 GMM 기반의 음성, 혼합 신호 분류 성능을 평가하기 위하여 우선 음성과 혼합 신호의 가장 적합한 가우시안 혼합 모델을 찾는 실험을 하였다. 인간의 귀의 비선형적 특성을 이용한 MFCC 특징 파라미터를 사용함으로써 적합한 가우시안 혼합 모델을 찾게 되며 MFCC에서는 DCT를 수행한 후 멜 스케일 부 밴드 에너지의 평균값인 13번째 켑스트럼 원소까지 특징으로 포함 한다. 제안하는 신호 분류에서는 각 프레임 별로 신호 분류 결과를 내지 않고, 연속된 프레임들의 통계적 특징에 대해서 분류기를 구성하게 된다.

실험에는 Mono 채널이며 16 bits/sample, 16kHz로 샘플링된 입력 신호를 사용하고, 1024샘플을 한 프레임으로 사용하여 두 단계로 거쳐 실험이 진행된다. 첫 번째 단계에서는 음성과 혼합신호의 최적의 GMM을 찾기 위해 실시되었다. 음성 신호와 혼합 신호를 EM 알고리즘을 기반으로 GMM 모델을 구현하였다. 두 번째 단계로는 훈련된 GMM을 가지고 음성과 혼합 신호를 분류 한다. 버퍼에 저장된 13개의 MFCC계수는 GMM의 특징벡터로 사용되며, 입력 신호 분류는 각 프레임을 가지고 처리한다.

성능 평가에서 실험을 위하여 40개의 음원을 사용하여 35개는 GMM훈련을 위해 사용되었고 나머지 5개는 분류 테스트를 위해 사용하였다.

표 2에서는 USAC의 신호 분류기와 제안하는 신호 분류기의 결과를 비교하여 성능을 나타내었다. 표 2의 혼합 신호1에서와 같이 혼합 신호를 입력 신호로 사용하였을 때 USAC은 61.6%만이 TCX를 모드를 선택하여 혼합 신호로 판단하였지만, 제안한 알고리즘에서는 모든 프레임을 혼합 신호로 판단하였다. 또한, 영어 남성의 음성 신호 역시 USAC에서 81%만이 ACELP 모드를 선택한 음성 신호로 판단하였지만 제안하는 알고리즘에서는 모든 프레임을 음성 신호로 판단하였다.

표 2. USAC과 제안된 알고리즘의 신호 분류 성능
Table 2. Signal classification of USAC and proposed algorithm.

Test file	방법	분류 에러
영어(남성)	USAC	19%
	Proposed	0%
영어(여성)	USAC	18%
	Proposed	0%
한국어(남성)	USAC	22%
	Proposed	1%
한국어(여성)	USAC	19%
	Proposed	0%
혼합 신호1	USAC	38.4%
	Proposed	0%
혼합 신호2	USAC	13.8%
	Proposed	2.8%
혼합 신호3	USAC	33.9%
	Proposed	3.4%

표 3. USAC과 제안된 알고리즘 PESQ값
Table 3. PESQ of USAC and Proposed algorithm.

	PESQ	
영어(남성)	Proposed	3.8105
	USAC	3.6728
영어(여성)1	Proposed	3.5047
	USAC	3.3204
영어(여성)2	Proposed	3.0421
	USAC	2.8152
한국어(남성)	Proposed	3.9795
	USAC	3.6054
한국어(여성)	Proposed	3.7348
	USAC	3.5819
혼합 신호1	Proposed	3.1821
	USAC	2.7321
혼합 신호2	Proposed	2.9385
	USAC	2.6822
혼합 신호3	Proposed	2.6497
	USAC	2.4835

또한 혼합 신호2에서의 모드 선택도 13.8%의 오류를 가지는 반면 제안된 알고리즘의 오류는 2.8%로 모드 선택을 더 정확하게 하는 것을 알 수 있다.

또한 기존 USAC에서는 Closeloop AbS 방식을 사용함으로써 Inverse transform까지 포함되어있어 주파수와 시간의 변환으로 연산량이 상당이 많은 알고리즘이다. 하지만 제안하는 알고리즘에서는 MFCC의 특징파

라미터를 구하기 위한 FFT의 n(logn)의 연산량과 GMM의 음성과 혼합신호를 판별하는 우도함수에서 약 0.04 wmops연산량이 발생하며 Inverse transform을 포함하지 않기 때문에 기존의 USAC보다 적은 연산량이 사용된다는 것을 알 수 있었다.

실험과 같은 올바른 모드 선택이 음질의 어떠한 영향을 미치는지를 알아보기 위해 객관적 음질평가인 ITU-T 표준안으로 제정된 PESQ(Perceptual Evaluation of Speech Quality)척도를 이용하여 모드의 선택의 중요성을 표 3에 나타내었다[15].

표 3에서는 기존 USAC의 신호 분류와 제안된 알고리즘의 신호 분류에 따른 PESQ값을 나타낸다.

제안된 알고리즘의 PESQ값은 USAC에 비해 혼합신호에서 평균 0.3정도 높게 나타났고, 음성 신호에서는 평균 0.2정도의 음질 차이를 보였다. 또한 그림 5~7에서는 USAC의 신호 분류기와 제안하는 신호 분류기를 사용하였을 때의 결과를 비교하여 나타내었다. 그림 5의 (a)는 혼합 신호를 나타내고, (b)는 기존 USAC의 신호 분류기를 통해 음성신호로 잘못 판단하여 ACELP 모드로 잘못 부호화 되었을 때의 결과 이다. 그림 5의 (c)는 우리가 제안하는 신호 분류기를 통해 혼합신호로 판단하여 TCX 모드로 부호화 되었을 때의 결과 이다. 그림 5의 (b)와 (c)에서 보듯이, 원신호의 피치 성분이 제안하는 신호 분류기를 사용하였을 때 기존의 USAC보다 더 잘 나타나는 것을 알 수 있다.

그림 6 역시 혼합 신호를 기존의 USAC과 제안하는 알고리즘의 신호 분류 방법으로 각각 적용해 보았을 때 제안하는 알고리즘이 기존의 USAC보다 원 신호와 더 유사하다는 것을 알 수 있다. 그림 7의 (a)는 음성 신호를 나타내었고, (b)는 기존 USAC의 신호 분류기를 통해서 혼합신호로 잘못 판단하여 TCX 모드로 잘못 부호화 되었을 때의 결과이다. 그림 7의 (c)는 제안하는

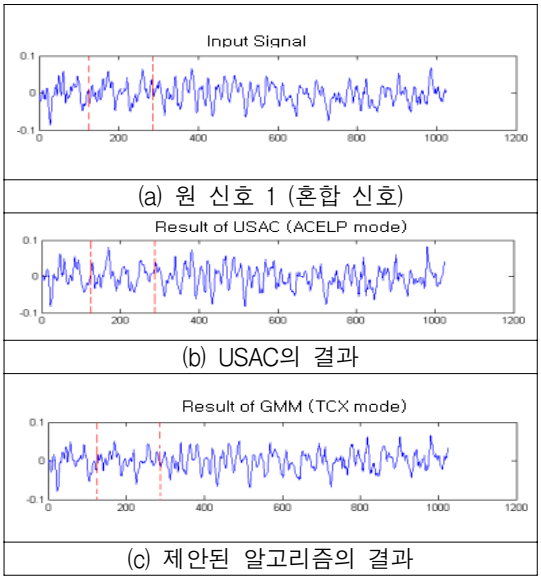

그림 5. USAC과 제안된 알고리즘의 결과 비교
Fig. 5. Result comparisons of USAC and proposed algorithm.

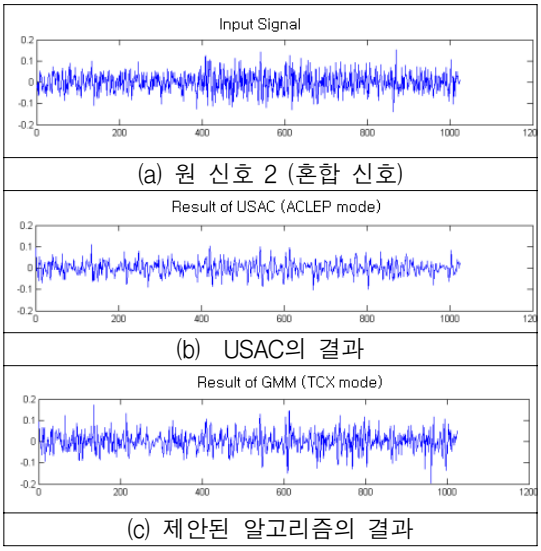

그림 6. USAC과 제안된 알고리즘의 결과 비교 2
Fig. 6. Result comparisons of USAC and proposed algorithm.

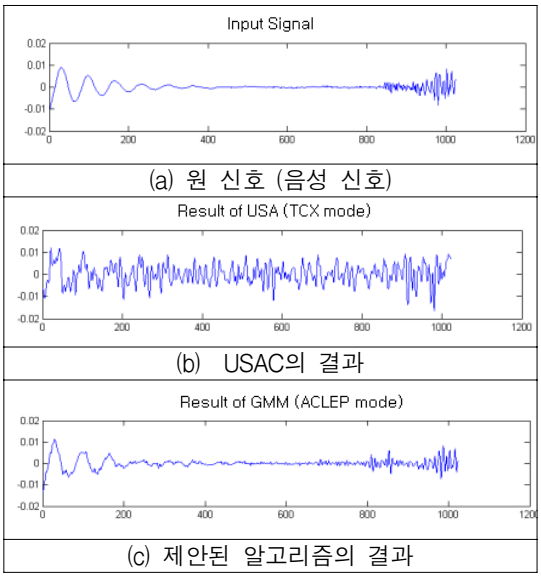

그림 7. USAC과 제안된 알고리즘의 결과 비교 3
Fig. 7. Result comparisons of USAC and proposed algorithm.

신호 분류기를 통해 음성신호를 판단하여 ACELP 모드로 부호화 되었을 때의 결과이다.

그림 7의 (b)에서 볼 수 있듯이, 음성신호의 묵음 구간 에서 노이즈가 발생하여, 음질에 안 좋은 영향을 미치는 것을 알 수 있었다. 그와 다르게 제안하는 분류기인 (c)에서는 묵음 구간에서도 노이즈가 적게 발생 하여 원 신호와 유사한 결과를 나타내는 것을 볼 수있다.

이처럼 신호를 분류함에 있어서 모드 선택이 원 신호와의 오류를 줄여줄 뿐 아니라 음질과도 영향을 주는 것을 알 수 있고, 올바른 모드 선택의 중요함을 알 수 있다.

Ⅳ. 결 론

본 논문에서 음성과 오디오 신호의 통합 코덱인 MPEG의 표준 코덱 USAC의 신호 분류에서의 성능을 향상시키고자 MFCC의 특징 파라미터 값을 추출하여 GMM을 통한 음성 신호와 혼합 신호 분류 방법을 제안하였다. 기존 USAC의 신호 분류방법에서 복잡성과 Close-loop AbS 방식의 많은 연산 량의 문제점을 개선하고자 하였고, 기존 USAC에서는 신호가 가지는 연속적인 특성을 고려하지 않은 현재 프레임만을 가지고 신호를 분류하였으나, 본 논문에서는 과거 프레임과의 상관성을 이용한 GMM 분류 방법을 통해 정확성을 높였다. 실험 결과 혼합신호의 경우 기존의 USAC의 잘못된 신호 모드 선택이 평균 29%였지만 제안하는 알고리즘을 통해서 평균 2.4%로의 오류만을 나타냈다. 음성신호 역시 제안하는 알고리즘은 평균 99.75%로 ACELP 모드를 정확히 선택 하는 것을 볼 수 있다. 이러한 올바른 모드 선택으로 원 신호를 제안하는 알고리즘이 기존의 USAC 보다 피치 성분을 잘 나타나는 것을 알 수 있었고, 객관적 음질평가인 PESQ의 값 역시 높게 나오는 것을 알 수 있다.

이처럼 MFCC를 이용한 GMM 분류 방법을 통해 현저히 적은 계산 량 뿐 아니라, 모드 선택의 정확성으로 더 나은 음질을 보이는 것을 알 수 있다.

참 고 문 헌

[1] ISO/IEC SC29 WG11 N9519, Call for Proposals on Unified Speech and Audio Coding, 82nd MPEG Meeting, October, 2007.

[2] 송정욱, 오현오, 강홍구, “통합 음성/오디오 부호화를 위한 새로운 MPEG 참조 모델,” 전자공학회논문지, 제47권 SP편, 제5호, 74~80쪽, 2010년 9월

[3] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classificaion, Wiley-Interscience, 2001.

[4] N.Scaringella, G. Zoia, and D.G.Stork, *Pattern Classification* Wiley-interscience, 2001.

[5] J. Bergstra, N.Casagrande,D. Erhan, D. Eck, and B. Kegl, “Aggregate features and ADABOOST for music calssificatio.” *Machine Learning*, vol. 65, no. 2, pp. 474-484, Dec. 2006.

[6] Martin F. Mcknney, Jeroen Breebaart, “Features for audio and music calssification” in *Proc. lnt. Conf. on Music lnfo. Retrieval (ISMIR-03)*, 2003.

[7] K. West, S. Cox, “Features and classifiers for the automatic classification of musical audio signals,” in *Proc. lnt. Conf. on Music lnfo. Retrieval (ISMIR-08)*, 2004.

[8] Bernd Geiser et al, “Candidate Proposal for ITU-T Super-wideband Speech and Audio Coding”, *ICASSP*, pp.4121-4124. 2009.

[9] M. Neuendorf, et al. ,“A novel scheme for low bitrate unified speech and audio coding-MPEG RM0,” in Proceedings of the 126th AES Convention, Munich, Germany, May 2009.

[10] 원양희, 이형일, 강상원, “ARM Core(R)를 이용한 AMR-WB+오디오 부호화기의 실시간 구현,” 전자공학회논문지, 제46권 제 3호, 119~124쪽, 2009년 5월

[11] B.Atal, “Automatic recognition of speakers from their voices” proc.IEEE vol.64 pp 460~475 apr.1976

[12] Thomas F. Quantieri, *Discrete-Time Speech Signal Processing*, Prentice Hall, 2001

[13] J. Makinen, B. Bessette, S. Bruhn, P. Ojala, R. Salami, and A.Taleb, “AMR-WB+: a new audio coding standard for 3RD generation mobile audioservices,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’05), vol. 2, pp. 1109 - 1112, March 2005.

[14] A.P.Dempster; N.M.Laird, et al.,“Maximum Likelihood from Incomplete Data via the EM Algorithm”, Journal of the Royal Statistical Society. Series B (Methodological),Vol.39,No.1.

[15] ITU-T Recommendation (1996). “Methods for subjective determination of transmission quality”, P.800, 08.

저 자 소 개

김 지 은(학생회원)
2012년 충북대학교 정보통신공학과 학사 졸업
2012년 ~현재 충북대학교 전파통신공학과 석사과정
<주관심분야 : 음성/음악 신호처리, 영상 신호처리>

이 인 성(정회원)-교신저자
1983년 연세대대학교 전자공학 학사 졸업
1985년 연세대대학교 전자공학 석사 졸업
1992년 Texas A&M University 전기공학과 박사 졸업
1993년~1995년 한국전자 통신연구원 이동통신 기술연구단 선임연구원
1995년~현재 충북대학교 전기전자공학부 정교수
<주관심분야 : 음성/오디오 신호처리, 이동통신, 적응필터>