

3GPP2 SMV의 실시간 음성/음악 분류 성능 향상을 위한 Gaussian Mixture Model의 적용

Analysis and Implementation of Speech/Music Classification for 3GPP2 SMV Based on GMM

송지현*, 이계환*, 장준혁*

(Ji-Hyun Song*, Kye-Hwan Lee*, Joon-Hyuk Chang*)

*인하대학교 전자전기공학부

(접수일자: 2007년 3월 9일; 수정일자: 2007년 5월 2일; 채택일자: 2007년 11월 6일)

본 논문에서는 음성 인식과 음악 인식에서 뛰어난 성능을 보이는 Expectation-Maximization (EM) 알고리즘 기반의 패턴인식기법인 가우시안 혼합모델 (Gaussian Mixture Model, GMM)을 이용하여 기존의 3GPP2 Selectable Mode Vocoder (SMV)의 실시간 음성/음악 분류 성능을 향상 시키는 방법을 제안한다. SMV의 음성/음악 실시간 분류 알고리즘에서 사용된 특징벡터와 분류방법을 분석하고, 이를 기반으로 분류성능향상을 위해 패턴인식 알고리즘인 GMM을 도입한다. 구체적으로, SMV의 음성/음악 분류알고리즘에서 사용되어진 특징벡터만을 선택적으로 사용하여 효과적인 GMM을 구성한 실시간 분류기법이 제시되었다. SMV의 음성/음악 분류에 적용한 GMM의 성능 평가 들 위해 SMV 원래의 분류알고리즘과 비교하였으며, 다양한 음악장르에 대해 시스템의 성능을 평가한 결과 GMM을 이용하였을 때 기존의 SMV의 방법보다 우수한 음성/음악 분류 성능을 보였다.

핵심어: 음성/음악 분류 알고리즘, Selectable Mode Vocoder (SMV), Gaussian Mixture Model (GMM)

투고분야: 음향처리 분야 (2)

In this letter, we propose a novel approach to improve the performance of speech/music classification for the selectable mode vocoder (SMV) of 3GPP2 using the Gaussian mixture model (GMM) which is based on the expectation-maximization (EM) algorithm. We first present an effective analysis of the features and the classification method adopted in the conventional SMV. And then feature vectors which are applied to the GMM are selected from relevant parameters of the SMV for the efficient speech/music classification. The performance of the proposed algorithm is evaluated under various conditions and yields better results compared with the conventional scheme of the SMV.

Key words: Speech/Music classification algorithm, Selectable mode vocoder (SMV), Gaussian mixture model (GMM)

ASK subject classification: Speech Signal Processing (2)

I. 서론

최근 이동통신의 발전으로 무선통신기기를 이용한 멀티미디어 서비스가 보편화 되면서 제한적인 주파수 대역에서 효과적으로 음성을 전송하는 연구가 지속적으로 이루어지고 있다. 현재 제한된 통신망을 효율적으로 사용하기 위해 가변적인 전송률을 갖는 다양한 음성 코덱이 개

발 되었다 [1, 2]. 실제로 입력 음성신호의 유형에 따라서 다른 비트를 할당하는 것은 바로 최종 음성의 음질에 영향을 미치기 때문에 정확한 신호분류기술의 핵심기술로서 다루어지고 있다. 특히, 단순히 음성통신을 다루던 것에서 벗어나 음악신호를 이동통신망을 통해 효과적으로 전송하기 위한 음성/음악 분류의 중요성이 증가하여 관련된 연구가 활발히 진행되고 있다 [3, 4].

본 논문에서는 실시간 음성/음악 분류기법을 기반으로 가변 전송률 알고리즘을 채택하고 있는 ETSI의 3GPP2 표준코덱인 Selected Mode Vocoder (SMV)의 기존 방법을

책임저자: 장준혁 (changjh@inha.ac.kr)
420-751 인천시 남구 용현동 253 인하대학교 전자전기공학부
(전화: 032-860-7423; 팩스: 032-868-3654)

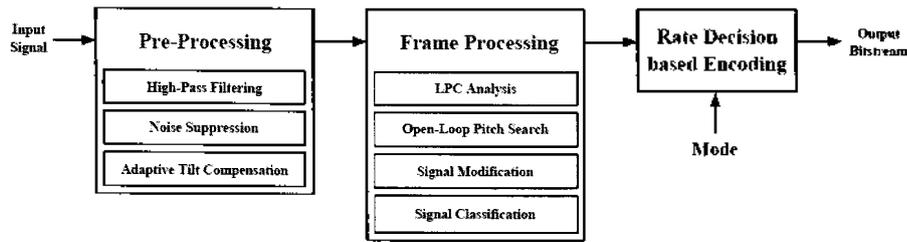


그림 1. SMV의 안코딩 과정에 대한 블록도
Fig. 1. Block diagram of the SMV encoding part.

충실히 분석하고 이를 기반으로 음성/음악 분류성능을 향상시키기 위한 기법을 제안한다. 구체적으로, 기존의 SMV의 인코딩부분의 전처리과정에서 자동적으로 추출되는 파라미터 중 통계적인 분류특성이 우수한 것들을 모아 별도의 계산과정 없이 특징벡터로 사용하였다. 선정된 특징벡터를 이용하여 실시간 Gaussian Mixture Model (GMM) 기반의 음성/음악 분류 알고리즘을 개발하고 이것을 기존의 SMV 방법과 다양한 환경에서 비교하였다 [5, 6].

본 논문의 II 장에서는 SMV 코덱에 대해서 기술하고, III 장에서 SMV 음성/음악 분류 방법과 특징 벡터에 대해서 알아본다. IV 장에서는 실시간 GMM 알고리즘에 대한 소개와 SMV의 특징 벡터를 이용한 GMM 기반의 음성/음악 분류 알고리즘을 제시한다. V 장에서 다양한 실험 환경에서 제안된 알고리즘과 기존의 알고리즘의 결과를 비교 검토한 뒤 VI 장에서 결론을 맺는다.

II. SMV (Selectable Mode Vocoder)

SMV 음성부호화기는 3GPP2의 표준화된 저전송율 음성 코덱이다 [7, 8]. SMV 음성 부호화기는 4가지 전송률 Rate 1 (8.55kbps), Rate 1/2 (4.0kbps), Rate 1/4 (2.0kbps), Rate 1/8 (0.8kbps)과 4개의 동작모드 Mode 0 (premium), Mode 1 (standard), Mode 2 (economy), Mode 3 (supereconomy)을 갖는다. 다양한 평균 전송률을 갖는 특성 때문에 CDMA 시스템의 성능과 음질간의 관계에서 선택적으로 성능을 조절 할 수 있다 [9].

그림 1은 SMV 입력 신호의 처리 과정을 보여주는데, 8 kHz로 샘플링 된 입력 신호를 160개의 샘플 (20 ms)마다 프레임의 모드와 전송률을 결정한다 [10]. 입력된 신호는 전처리 과정을 거치게 되며 묵음 증가, 고대역 통과 필터, 노이즈 억제, 적응 틸트 보상 등을 통해서 백그라운드 노이즈를 제거한다. 프레임 처리 과정에는 선형 예측

분석, 개회로 (open-loop) 피치 검출, 신호 수정 및 분류 등이 포함된다.

선형 예측 분석과 개회로 피치 검출을 통해서 음성과 음악 분류에 대한 파라미터들이 추출된다. 신호 분류를 통해서 프레임은 잡음, 묵음, 부성음, 비정상적 유성음, 정상적 유성음중 한 개로 분류 된다. 프레임의 전송률은 현재 프레임의 분류된 종류와 통신 상태에 따라서 정해지는 모드를 바탕으로 전송률 결정 알고리즘 (Rate Determination Algorithm, RDA)에 의해 결정된다. 프레임이 음악으로 판별된 경우 프레임의 전송률은 모드에 상관없이 Rate 1로 결정되어 최고의 전송률을 할당하며 그 외의 경우에는 정해진 한계 값에 의해서 결정하여 인코딩한다 [10].

III. SMV 음악 분류 특징 방법 및 특징 벡터

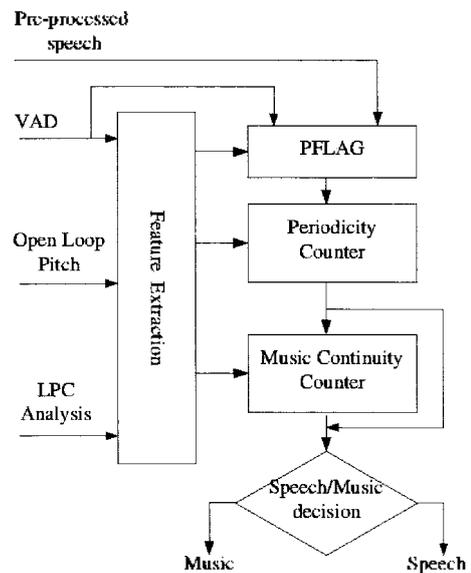


그림 2. 기존의 SMV 음성/음악 분류 알고리즘 블록도
Fig. 2. Block diagram of the speech/music classification algorithm of the SMV.

SMV는 8 kHz로 샘플링된 입력 신호를 20 ms 마다 부호화하는데, 음성 검출 (Voice Activity Detection, VAD)에서 부음과 음성으로 분류 한 뒤 음악 분류를 한다. 그림 2는 SMV 음악 분류 알고리즘의 블록을 나타내는데 VAD, 개회로 피치 검출, 선형 예측 부호화 (Linear Prediction Coding, LPC)에서 추출된 특징 벡터를 입력 받고, 각 특징 벡터의 이동 평균 값을 음악 분류 알고리즘의 특징 벡터로 이용한다. 이러한 특징 벡터들을 경험적으로 설정되어 설정된 문턱 값과 비교하여 음성/음악 분류를 한다. 구체적으로, 음악 분류 알고리즘에 입력되는 특징 벡터에 대해서 다음 세부 절에서 살펴본다.

3.1. LPC 분석에서 추출된 특징 벡터

- 1) 반사 계수 : i번째 반사 계수 (Reflection Coefficients, $k_1(i)$)가 Levinson-Durbin 알고리즘에 의해서 구해진다 [11].
- 2) Line spectral frequencies (LSF) : LSF는 인코딩 프레임의 마지막 1/4에 가중치를 갖고 구해진 10차 LPC값을 변화한 값으로 $lsf(i)$ 로 나타낸다 [12].

3.2. VAD에서 추출된 특징 벡터

- 1) 프레임 에너지 : 프레임에 대한 에너지 값으로 최소 10 이상의 값을 갖는다.

$$E = \max\left(10, 10 \cdot \log_{10}\left(\frac{R_1(0)}{L_{LPC}}\right)\right) \quad (1)$$

여기서 $R_1(0)$ 값은 신호의 파워, L_{LPC} 값은 LPC 윈도우의 길이로 240을 갖는다.

- 2) VAD decision flag : VAD에서 추출된 파라미터들과 설정된 문턱값의 비교를 통해서 현재 프레임이 음성인지를 판단한다.

3.3. 음악 분류 알고리즘 특징 벡터

- 1) LSF의 이동평균 :

$$\overline{lsf}(1) = 0.75 \cdot \overline{lsf}(1) + 0.25 \cdot lsf_1(1) \quad (2)$$

여기서 $lsf_1(1)$ 는 LPC 분석을 통해서 구해진 10차의 lsf 값 중 첫 번째 값을 나타낸다.

- 2) 에너지의 이동 평균 : 식 (1)에 의해서 구해진 에너지의 이동 평균 값이다.

$$\overline{E} = 0.75 \cdot \overline{E} + 0.25 \cdot E \quad (3)$$

- 3) 스펙트럼 차이 : 프레임의 반사 계수와 노이즈 프레임의 반사 계수의 차이에 의해서 구해진다.

$$SD_1 = \sum_{i=1}^{10} (k_1(i) - \bar{k}_N(i))^2 \quad (4)$$

- 4) 정규화 된 피치 상관도의 이동 평균 :

$$\overline{corr}_P = 0.8 \cdot \overline{corr}_P + 0.2 \cdot \left(\frac{1}{5} \cdot \sum_{i=1}^{i=5} corr_P^B(i)\right) \quad (5)$$

여기서 $corr_P^B(i)$ 는 이전 프레임의 피치 상관도 이다.

- 5) 주기적 계수 (periodicity counter)의 이동 평균 : (2) ~ (5)의 특징 벡터와 설정된 문턱 값의 비교를 통해서 증가하거나 감소하고, $\bar{c}_{pr} \geq 18$ 이면 프레임은 음악으로 분류한다.

$$\overline{c}_{pr} = \alpha \cdot \overline{c}_{pr} + (1 - \alpha) \cdot c_{pr} \quad (6)$$

여기서 α 값은 고정된 가중치 값을 나타낸다.

- 6) 음악 연속성 계수 (music continuity counter)의 이동 평균 : (2) ~ (6)의 특징 벡터와 설정된 문턱 값의 비교를 통해서 증가하거나 감소하고, $c_M \geq 200$ 이면 프레임을 음악으로 분류한다.

$$\overline{c}_M = 0.9 \cdot \overline{c}_M + 0.1 \cdot c_M \quad (7)$$

IV. 제안된 SMV 코딩 특징 벡터를 이용한 GMM 기반의 음성/음악 분류 시스템

본 논문에서는 별도의 연산 과정 없이 SMV의 인코딩 과정에서 자동 생성되는 파라미터들을 특징 벡터로 이용 하되 효과적인 GMM을 구성하기 위해 SMV 코딩 파라미터를 선택적으로 사용하여 실시간 음성/음악 분류를 제안

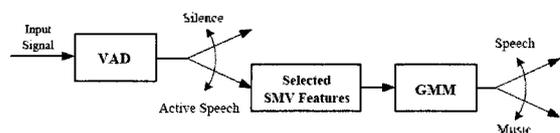


그림 3. 제안된 방법의 음성/음악 분류 블록도
Fig. 3. Block diagram of the proposed method.

한다. 이번 연구에서 GMM을 사용하게 된 동기는 SMV 특징 벡터의 통계적 분포를 다른 평균과 공분산 행렬을 갖는 복수개의 가우시안 함수에 의해서 효과적으로 표현할 수 있기 때문이다. 그림 3은 제안된 GMM기반의 SMV 코딩파라미터를 이용한 음성/음악 분류 알고리즘의 블록 다이어그램을 보여주는데 SMV의 음성/음악 분류 알고리즘에서 VAD값이 신호가 있다고 판별할 경우 제안된 GMM은 생성된 특징 벡터와 GMM의 모델을 이용하여 우도를 생성하고, 평가를 통해서 음성/음악을 분류하는 과정을 이루게 된다.

먼저, GMM은 화자 인식과 음악 인식에서 뛰어난 성능을 보이는 패턴 인식기로서 가우시안 형태의 혼합 모델 기반 확률 밀도 함수로 다음과 같이 주어진다.

$$P(\vec{x}|\lambda) = \sum_{i=1}^M \alpha_i P_i(\vec{x}) \tag{8}$$

여기서,

$$\sum_{i=1}^M \alpha_i = 1, \tag{9}$$

$$P_i(\vec{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T (\Sigma_i)^{-1} (\vec{x}-\vec{\mu}_i)\right\}, \tag{10}$$

실제로 파라미터 α_i 는 혼합된 가우시안 밀도의 가중치를 나타내고, \vec{x} 는 D개의 특징 벡터, $\vec{\mu}$ 는 평균 벡터, Σ_i 는 공분산 행렬, λ 는 GMM의 모델 파라미터를 나타낸다.

$$\lambda = \left\{ \alpha_i, \mu_i, \Sigma_i \right\}, \quad i = 1, \dots, M \tag{11}$$

정밀한 확률 밀도함수를 사용하는 GMM에서 특징 벡터의 통계적 편차에 대한 구분이 뚜렷할수록 더욱 뛰어난 성능을 보이기 때문에 이를 위해서 SMV의 음악 분류에 사용되는 특징 벡터의 통계적인 분류 특성을 조사하였다. 실제로, 평균에너지, 스펙트럼 차이, 정규화 된 피치 상관도, 음악 연속성 계수가 통계적인 분류 특성이 우수하며, 그림 4는 4개의 특징 벡터의 통계적인 분류 특성을 비교하여 보여주고 있다. 또한, 음성/음악 분류에서 피치는 우수한 성능을 보여주기 때문에 SMV에서 정규화 되지 않은 피치 특징 벡터를 추출하여 총 5개의 특징 벡터를 GMM의 특징 벡터로 제안한다 [13]. 선택된 특징 벡터를 이용하여 EM 알고리즘 기반의 학습을 통해서 음성 모델 λ_s 과 음악 모델 λ_m 을 만들었다.

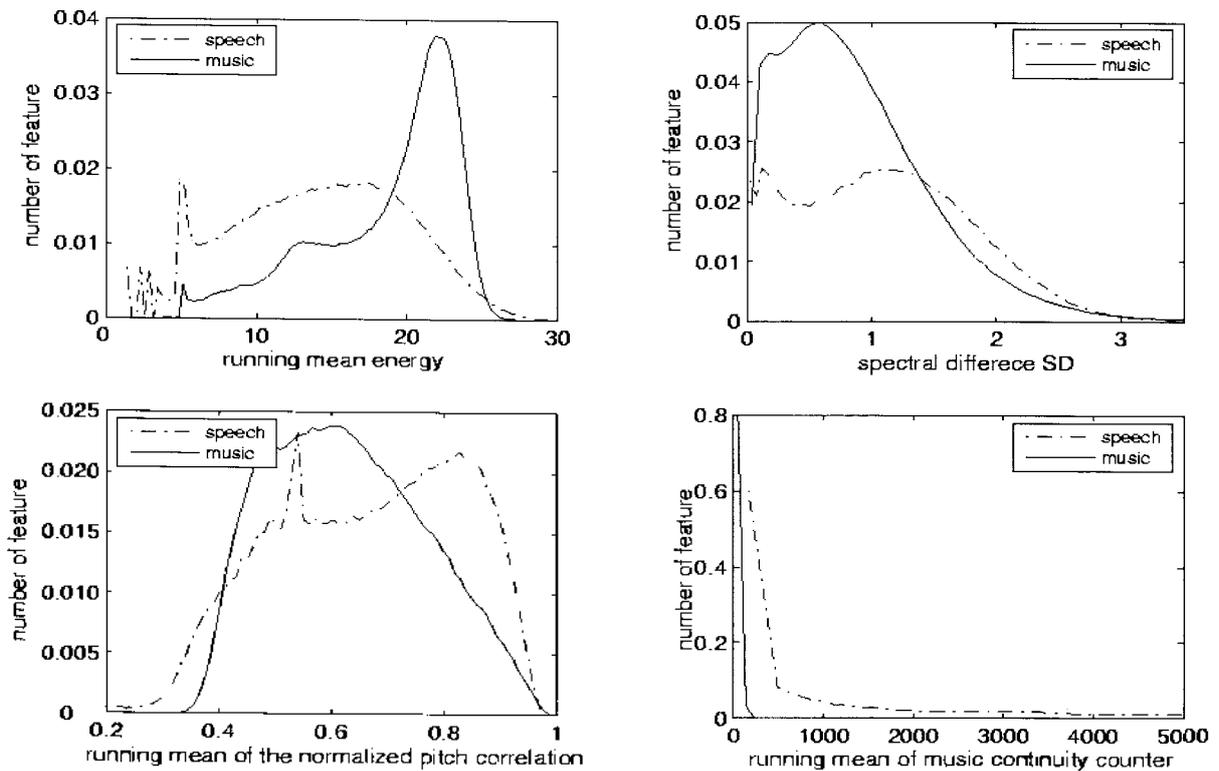


그림 4. SMV 특징 벡터에 대한 음성/음악 분포도
 Fig. 4. Normalized distributions of the SMV feature vectors for speech and music.

테스트 과정에서 일반적으로 기혼련된 음성과 음악에 대한 GMM 모델에 실제 데이터가 입력되어 구한 우도비 (Likelihood Ratio)를 계산하여 문턱값과 비교를 통해서 음성/음악 분류를 한다.

$$\log A_x^{(t)} = \log \frac{P(\vec{x}^{(t)} | \lambda_m)}{P(\vec{x}^{(t)} | \lambda_s)} \begin{matrix} \text{music} \\ > \\ \text{speech} \end{matrix} \quad \eta \quad (12)$$

여기서 η 는 음성/음악 분류의 문턱값이고, t 는 프레임 번호이다. 매 프레임 간 독립적인 처리과정에 대한 오류를 보상하고 상관성을 고려하여 실제로 Smoothed Likelihood Ratio (SLR)를 사용하여 음성 부호화에 적합한 실시간 분류 방법을 아래와 같이 제시한다 [14].

$$\psi_x^{(t)} = \exp\{\beta \log \psi_y^{(t-1)} + (1 - \beta) \log A_x^{(t)}\} \quad (13)$$

여기서 β 는 실험을 통해서 최적화된 가중치로 0.98로 설정한다.

V. 실험 결과

본 장에서는 SMV의 음성/음악 분류 알고리즘에 적용한 GMM 기반의 실시간 음성/음악 분류 성능을 알아보기 위해, 기존의 SMV의 알고리즘과 비교하였다. 본 실험을 위해서 음성 데이터베이스로 8 kHz로 샘플링 된 약 6 sec 정도의 깨끗한 음성으로 326명의 남자와 138명의 여자 화자에 의해서 각 10개의 파일이 발음 된 TIMIT 데이터베이스가 사용되었다 [15]. 음악 데이터베이스는 CD로부터 여러 장르의 음악을 모바일 폰을 통해서 녹음하였고, 8 kHz로 다운 샘플링 하여 사용하였으며, 5분 정도의 음악파일이 사용되었다.

제안된 음성/음악 분류 알고리즘의 모델은 음성 파일 4200개와 음악 파일 60개 (메탈 12개, 재즈 12개, 블루스 12개, 힙합 12개, 클래식 12개)를 이용하여 트레이닝 하였으며, 다양한 혼합 성분의 개수 (16, 32, 64, 128, 256, 512)를 사용하였다.

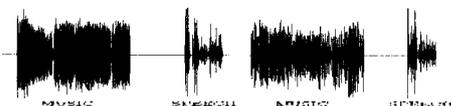


그림 5. 테스트 파일의 형태
Fig. 5. Shape of test files.

표 1. 제안된 알고리즘의 수행 시간 증가량 비교
Table 1. Comparison of computational time of the proposed method.

N	16	32	64	128	256	512
GMM	0.30%	0.53%	1.06%	2.05%	4.04%	9.00%

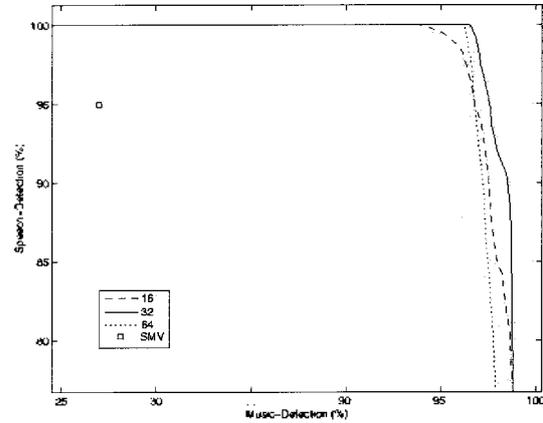


그림 6. SMV와 제안된 알고리즘의 음성/음악 분류에 대한 ROC 곡선
Fig. 6. ROC curve for the speech/audio classification using the SMV and the proposed method.

SMV와 제안된 알고리즘의 객관적인 성능을 평가하기 위해서 테스트 파일을 만들었다. 동일한 데이터에 의한 성능 향상을 피하기 위해서 트레이닝에 사용된 음성/음악 데이터는 테스트에 사용되지 않았다. 테스트 파일은 그림 5와 같은 형태로 5개 음성 파일 (6 ~ 12초), 5개 음악 파일 (28 ~ 32초), 10개 무음 (3 초 ~ 15초)을 사용하여 만들었다.

다양한 음악 장르에 대한 음성/음악 분류 성능을 확인하기 위해서 테스트 파일의 음악을 2가지 형태로 각 장르 별 (힙합, 메탈, 재즈, 블루스, 클래식)로 구성된 형태의 테스트 파일 60개, 음악 장르가 혼합된 형태의 테스트 파일 24개 총 84개의 테스트 파일을 만들었다. 두 시스템 실제 성능을 알아보기 위해서 테스트 파일의 20 ms 마다 실제로 결과를 0 (무음), 1 (음성), 2 (음악)로 수동으로 작성한 것과 비교하였다.

먼저, 최적의 혼합 성분 개수를 구하기 위해서 혼합된 형태의 테스트 파일에 대해서 각 혼합 성분 개수에 대한 Receiver Operation Characteristic (ROC) 곡선과 수행 시간 증가량을 구하여서 각각 그림 1과 표1에 도시하였다. ROC 곡선은 식 (12)의 우도비에 가중치를 주고 값을 변화 시키면서 음성/음악에 대한 검출 확률 (P_d)을 구하였다. 여기서 P_d 는 음악프레임 일 경우 음악으로 분류된 확률과 음성프레임 일 경우 음성으로 분류된 확률을 뜻한다. 실험결과 전체 구간에서 제안된 음성/음악 분류 알고리즘 성능이 우수함을 확인할 수 있었고, 혼합 성분 개수

표 2. SMV와 제안된 알고리즘의 음성/음악 분류 성능 비교
Table. 2. Comparison of speech/music detection probability P_d between the method of the SMV and the proposed technique.

TEST	Method	Music	Speech
Mixed	SMV	0.21	0.93
	Proposed	0.91	1.00
Metal	SMV	0.22	0.91
	Proposed	0.96	1.00
Blues	SMV	0.15	0.90
	Proposed	0.89	1.00
Hiphop	SMV	0.28	0.90
	Proposed	0.97	1.00
Classic	SMV	0.50	0.90
	Proposed	0.63	0.97
Jazz	SMV	0.27	0.92
	Proposed	0.91	0.99

64차 이상에서는 성능 변화를 볼 수 없었다. 제안된 알고리즘의 계산 증가량은 Intel (R) Core (TM)2 CPU 6300 1.86GHz 에서 비주얼 C++을 이용하여 각 알고리즘의 전체 수행 시간을 측정하여 다음과 같이 구하였다.

$$\text{수행시간증가량(\%)} = \frac{(\text{제안된 알고리즘의 수행시간} - \text{SMV의 수행시간})}{\text{SMV의 수행시간}} \times 100$$

표 1을 통해서 혼합 성분의 개수가 증가함에 따라서 수행 시간이 증가하는 것을 볼 수 있다. 두 실험을 통해서 성능과 계산 량을 고려했을 때 혼합 성분 개수 32차를 사용했을 때 가장 우수함을 알 수 있다.

표 2는 SMV와 제안된 GMM 기반의 알고리즘의 음성/음악 검출 확률 (P_d)을 나타낸다. GMM의 혼합 성분 개수는 32차를 이용하였고, 검출 확률은 테스트 파일의 음악 형태에 대한 평균 수치로 나타내었다. 표 2을 통해서 기존의 SMV 알고리즘 보다 제안된 GMM 기반의 알고리즘이 전반적으로 우수함을 볼 수 있다. 특히, 음악 분류에서는 매우 향상된 성능을 보였다. 다양한 음악 장르에 대한 음악 분류 성능은 음악 장르와 시스템에 따라서 큰 차이를 보이지만 기존의 SMV와 제안된 GMM 기반의 제안된 방법을 비교하였을 때 모든 장르에서 제안된 알고리즘의 성능이 뛰어난을 알 수 있다.

코덱 음질의 객관적인 평가를 위해서 전체 테스트 파일에 대해서 Perceptual Evaluation of Speech Quality (PESQ) 테스트 실험을 하였고, PESQ값은 테스트 파일의 음악 형태에 대한 평균 수치로 나타내었다. 표3은 기존의 방법보다 GMM을 이용한 제안된 알고리즘 방법이 혼합형 (Mixed), 메탈 (Metal), 블루스 (Blues), 힙합 (Hiphop),

표 3. SMV와 제안된 알고리즘의 PESQ 수치 비교
Table.3. PESQ scores of the SMV and the proposed method.

TEST	Method	PESQ
Mixed	SMV	3.077
	Proposed	3.180
Metal	SMV	2.983
	Proposed	3.232
Blues	SMV	3.124
	Proposed	3.174
Hiphop	SMV	3.143
	Proposed	3.151
Classic	SMV	2.783
	Proposed	2.923
Jazz	SMV	3.155
	Proposed	3.167

클래식 (Classic), 재즈 (Jazz) 에서 각각 평균 0.103, 0.249, 0.05, 0.008, 0.140, 0.012 정도 향상된 수치를 보인다. 따라서 실험에 사용된 모든 음악 환경에서 기존의 SMV 알고리즘 보다 제안된 방법이 전체적으로 우수한 음질을 보여주고 있다.

마지막으로, 그림 7은 SMV와 제안된 알고리즘의 실시간 음성/음악 분류 성능을 평가하기 위해서 음성/음악의 파형과 두 시스템의 분류 결과를 시간 축 상에서 비교하여 도시하였다. 그림 7 (1)은 테스트 파일의 일부분으로 음성 - 음악 - 음성 - 음악 순서로 구성된 부분이고, 그림 7 (2) 는 테스트 파일의 매뉴얼을 나타내고, 음성일 때 1, 음악일 때 2를 나타낸다. 그림 7 (3)과 그림 7 (4) 는 SMV의 실시간 음성/음악 분류결과와 제안된 방법의 분류 결과를 나타내는데 두 시스템의 실시간 음악 분류 성능 뚜렷한 차이를 볼 수 있다.

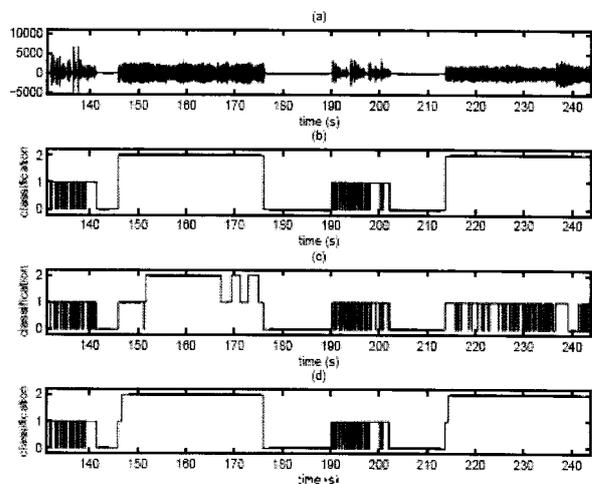


그림 7. (1) 테스트 파일의 파형 (2) 테스트 파일의 음성/음악 분류 매뉴얼 (3) SMV의 음성/음악 분류 결과 (4) 제안된 알고리즘의 음성/음악 분류 결과
Fig. 7. (1) Waveform of the test file (sequence : speech - silence - music - silence - speech - silence - music) (2) Manual of the test file (3) Results of the SMV (4) Results of the proposed method.

VI. 결론

본 논문에서는 ETSI의 3GPP2 표준코덱인 SMV의 실시간 음성/음악 분류 성능을 향상시키기 위해 기존의 SMV 코딩 특징 벡터를 이용하여 GMM에 기반을 둔 음성/음악 분류 방법을 제안하였다. SMV의 음성/음악 분류 알고리즘에 사용되는 통계적 분류 특성이 우수한 특징 벡터를 GMM의 특징 벡터로 적용하였고, 기존의 SMV와 성능을 비교한 결과 제안된 GMM을 이용하여 SMV의 분류 성능 향상한 기법이 기존의 SMV의 분류성능에 비해서 향상된 실시간 음성/음악 분류 성능을 보여 주었다.

감사의 글

이 논문은 2007학년도 인하대학교의 지원에 의하여 연구되었음.

참고 문헌

1. Y. Gao, E. Shlomot, A. Benyassine, J. Thyssen, Huan yu Su, and C. Murgia, "The SMV Algorithm Selected by TIA and 3GPP2 for CDMA Applications," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 709-712, May 2001.
2. 3GPP2 Spec., "Source-controlled variable-rate multimedia wideband speech codec (VMR-WB), service option 62 and 63 for spread spectrum systems," 3GPP2-C.S0052-A, v.1.0, Apr. 2005.
3. J. Saunders, "Real-time discrimination of broadcast speech/music," Proc. IEEE International Conference on Acoustics, Speech, and Processing, 2, 993-996, May 1996.
4. W. Q. Wang, W. Gao, and D. W. Ying, "A fast and robust speech/music Discrimination Approach," Proc. International Conference on Information, Communications and Signal Processing, 3, 1325-1329, Dec. 2003.
5. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," Digital Signal Processing, 10, 19-41, Jan. 2000.
6. D. A. Reynolds, and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture models," IEEE Transactions on Speech and Audio Processing, 3, 72-83, Jan. 1995.
7. J. Makinen, P. Ojala, and H. Toukoma, "Performance comparison of source controlled GSM AMR and SMV vocoders," Proc. International Symposium on Intelligent Signal Processing and Communication Systems, 51-154, Nov. 2004.
8. C. V. Goudar, P. Rabha, M. Deshpande, and A. Rao, "SMVlite: Reduced Complexity Selectable Mode Vocoder," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 1, 701-704, May 2006.

9. S. Craig Greer, and A. Dejaco, "Standardization of the selectable mode vocoder," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 953-956, May 2001
10. 3GPP2 Spec., "Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems," 3GPP2-C.S0030-0, v3.0, Jan. 2004.
11. P. Kabal, R. Prakash and Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," IEEE Trans. Acoustics, speech and signal processing, ASSP-34, (6) 1419-1426, Dec. 1986.
12. P. Vary and R. Martin, *Digital Speech Transmission : enhancement, coding and error concealment*, (182-187, 2006)
13. A. R. Abu-El-Quran and R. A. Goubran, "Pitch-based feature extraction for audio classification," Proc. IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications, 43-47, Sep. 2003.
14. Y. D. Cho, and A. Kondoz, "Analysis and improvement of a statistical model-based voice activity detector," IEEE Signal Process. Lett., 8, 276 - 278, Oct. 2001.
15. W. M. Fisher, G. R. Doddington and K. M. Goudie-Marshall, "The DARPA speech recognition research database: Specifications and status," Proc. DARPA Workshop Speech Recognition, 93-99, Feb. 1986.

저자 약력

• 송지현 (Ji-Hyun Song)



2007년 2월: 인하대학교 전자공학과 학사
2007년 3월~현재: 인하대학교 전자 전자공학과 석사과정

• 이계환 (Kye-Hwan Lee)



2007년 2월: 인하대학교 전자공학과 학사
2007년 3월~현재: 인하대학교 전자 전자공학과 석사과정

• 장준혁 (Joon-Hyuk Chang)



1998년 2월: 경북대학교 전자공학과 학사
2000년 2월: 서울대학교 전기공학부 석사
2004년 2월: 서울대학교 전기컴퓨터공학부 박사
2000년 3월~2005년 4월 : 휴렛팩스 연구소장
2004년 5월~2005년 4월: 캘리포니아 주립대학, 산타바버라 (UCSB) 박사후연구원
2005년 5월 ~2005년 8월 한국과학기술연구원 (KIST) 연구원

2005년 9월~현재: 인하대학교 전자전자공학부 조교수