

오디오 인덱싱을 위한 음성/음악 분류 특징 비교

A Comparison of Speech/Music Discrimination Features for Audio Indexing

이 경 록*, 서 봉 수*, 김 진 영*

(Kyong Rok Lee*, Bong Su Seo*, Jin Young Kim*)

*전남대학교 전자공학과

(접수일자: 2000년 6월 27일; 수정일자: 2000년 12월 11일; 채택일자: 2001년 1월 18일)

본 논문에서 우리는 음향신호에서 음성과 음악을 분류하는 음성/음악 분류실험에 사용되는 특징들간의 상호조합을 비교하였다. 음향신호는 3가지 (음성, 음악, 음성+음악)와 2가지 (음성, 음악)로 분류하였다. 실험은 멜캡스트럼, 에너지, 영교차를 특징으로 사용하였고, 음성/음악 분류성능이 가장 좋은 특징간 상호조합을 모색하였다. 분류 알고리즘으로는 Gaussian Mixture Model (GMM)을 이용하였으며, GMM에 의한 데이터 모델링 전에 각기 다른 특징들을 하나의 특징공간에서 결합하였다. 실험결과 3가지 분류기준 적용시에는 멜캡스트럼, 영교차 조합이 가장 좋은 결과 (음성: 95.1%, 음악: 61.9%, 음성+음악: 55.5%)를 보였고, 2가지 분류기준 적용시에는 멜캡스트럼, 에너지 조합과 멜캡스트럼, 에너지, 영교차 조합이 가장 좋은 결과 (음성: 98.9%, 음악: 100%)를 보였다.

핵심용어: 오디오 인덱싱, 음성/음악 분류, GMM

투고분야: 음성처리 분야 (2.4)

In this paper, we describe the comparison between the combination of features using a speech and music discrimination, which is classifying between speech and music on audio signals. Audio signals are classified into 3classes (speech, music, speech and music) and 2classes (speech, music). Experiments carried out on three types of feature, Mel-cepstrum, energy, zero-crossings, and try to find a best combination between features to speech and music discrimination. We using a Gaussian Mixture Model (GMM) for discrimination algorithm and combine different features into a single vector prior to modeling the data with a GMM. In 3classes, the best result is achieved using Mel-cepstrum, energy and zero-crossings in a single feature vector (speech: 95.1%, music: 61.9%, speech & music: 55.5%). In 2classes, the best result is achieved using Mel-cepstrum, energy and Mel-cepstrum, energy, zero-crossings in a single feature vector (speech: 98.9%, music: 100%).

Key words: Audio indexing, Speech/music discriminate, GMM

Subject classification: Speech signal processing (2.4)

I. 서 론

정보화시대의 도래에 따른 정보의 형태변화는 대용량 멀티미디어 데이터베이스를 일반적인 추세로 만들었다. 요즘 이러한 데이터베이스에서의 멀티미디어 인덱싱 기반 정보검색이 활발히 연구되고 있다.

멀티미디어 인덱싱은 크게 비디오 인덱싱과 오디오 인덱싱으로 나눌 수 있다. 비디오 인덱싱은 영상을 커트로 구분하고 목표물을 추출함으로써 내용을 구분하는 방식이다. 오디오 인덱싱은 음향 신호에서 정보를 가진 부분 (음성, 폭발음, 합성 등)을 분리하고 내용을 구분하는 방식이다.

현재 실험 중인 오디오 인덱싱은 방송뉴스를 대상으로

하여 입력 음향 신호에서 음성부분만을 추출하고, 화자별로 분류한 다음, 특정화자 (앵커, 아나운서 등)의 발성부분에 대하여 음성인식기술을 적용하여 내용을 분석하고자 한다.

본 논문에서는 차후 음성인식을 통한 내용 분석에 필요한 음성부분만을 추출하는 오디오 인덱싱의 전처리부격인 음성/음악 분류기에 사용되는 특징들의 상호조합에 대하여 실험하였다.

음성/음악 분류기에 사용되는 특징으로서는 멜캡스트럼, 피치, 에너지, 영교차가 가장 널리 사용되고 있다[1]. 실험에서는 이들 중 3가지 특징 (멜캡스트럼, 에너지, 영교차)들을 선정하고, 특징들간의 상호조합을 통해 가장 양호한 결과를 생성하는 조합을 찾아내는데 중점을 두었다.

음성/음악 분류기에서 사용되는 분류 알고리즘으로는 Gaussian mixture model, k-nearest-neighbors classification, 벡터 양자화, 결정 트리 등이 있으나 본 논문에서는 GMM을 사용하였다[1-3].

책임저자: 이경록 (Krllee@dsp.chonnam.ac.kr)
500-757 광주광역시 북구 용봉동 300번지 전남대학교
전자공학과 신호처리실험실
(전화: 062-530-0472; 팩스: 062-530-0472)

II. 특징 파라미터

각 특징 파라미터의 분석은 아래와 같은 조건하에서 실시하였다. 음성 신호는 8kHz로 표본추출되었다. 한 프레임의 크기는 25ms이며 10ms씩 이동시켜가면서 특징 파라미터를 추출하였다. 델타 파라미터는 연속적인 5개 이상의 프레임 구간에서 추출하였다.

2.1. 멜캡스트럼

특징 파라미터로는 12개의 멜캡스트럼 계수들과 12개의 델타 멜캡스트럼 계수들을 사용하였다.

분석을 위해 사용된 캡스트럼 수식은 다음과 같다.

$$C_x(m) = \frac{1}{N} \sum_{k=0}^{N-1} \log |X(k)| e^{2\pi \frac{mk}{N}} \quad (1)$$

식 (1)의 대수전폭은 아래와 같다.

$$\log |X(k)| = \sum_{m=0}^{N-1} C_x(m) e^{-2\pi \frac{mk}{N}} \quad (2)$$

캡스트럼 파라미터는 멜 스케일 특성을 가진 26개의 필터로 구성된 필터뱅크로 필터링을 실시하였다.

멜캡스트럼은 차후 화자변화탐색 분야에서 다사 활용 가능하며, 효율면에서나 분류능력면에서도 가장 양호한 것으로 알려져 있다[1].

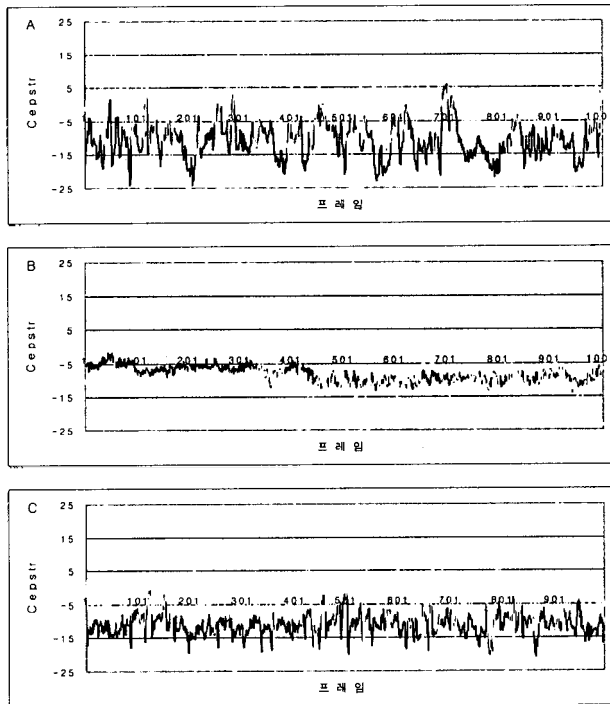


그림 1. 각 분류항목별 멜캡스트럼 변화 (1,2번째 계수값)
A : 음성 B : 음악 C : 음성+음악

Fig. 1. The variance of Mel Cepstrum in each class (the value of first, second Mel Cepstrum coefficients),
A : speech, B : music, C : speech and music.

그림 1은 멜캡스트럼 계수값의 변화를 나타낸 것이다. 멜캡스트럼의 12개의 계수값 중 각 분류항목의 특징을 가장 잘 나타낼 수 있는 2개의 계수값만을 선별하여 나타내었다 (1, 2번째 계수값).

선별된 멜캡스트럼 계수값의 변화를 살펴보면 음성, 음성+음악, 음악 순으로 격렬한 변화를 보인다.

2.2. 에너지

에너지의 특징 파라미터로는 정규화 로그 에너지와 델타 정규화 로그 에너지를 사용하였다.

사용된 에너지 파라미터의 수식은 다음과 같다.

$$E = \frac{1}{E_{max}} \left(\log \left(\frac{1}{N} \sum_{k=0}^{N-1} x(k)^2 \cdot h(m-k) \right) \right) \quad (3)$$

로그 에너지는 음성/음악 분류기가 크기 정보에 의존하여 분류하지 않도록 정규화를 실시하였다.

그림 2는 각 분류항목별 프레임 변화에 따른 정규화 로그 에너지 변화를 나타내고 있다. 음성의 변화가 음악과 음성+음악에 비해 훨씬 격렬하다.

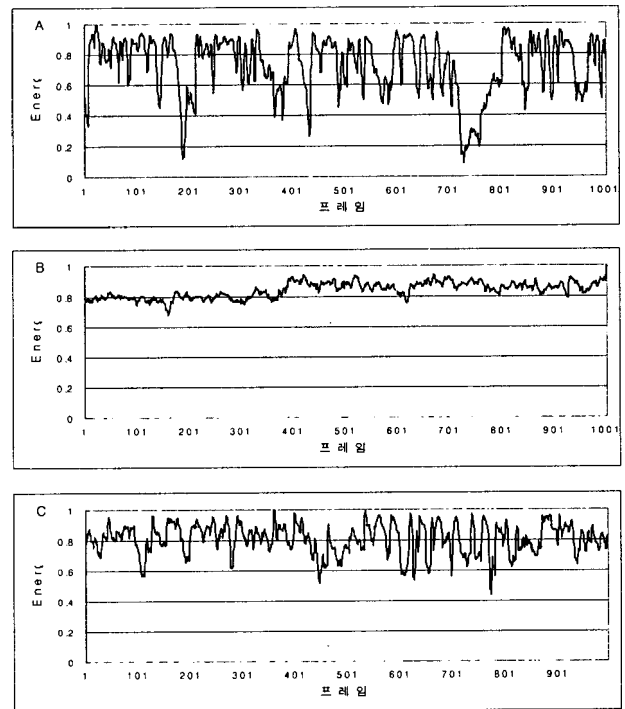


그림 2. 각 분류항목별 정규화 로그 에너지 변화

A : 음성 B : 음악 C : 음성+음악

Fig. 2. The variance of normalized log energy in each class,
A : speech, B : music, C : speech and music.

2.3. 영교차

영교차는 계산비용 대 성능비가 우수한 것으로 알려져 있다. 영교차를 기반으로 하는 4가지 특징은 영교차율, 영교차율 변화량의 표준편차, 영교차 평균보다 높은 값과 낮은 값들간의 차, 영교차율의 제삼 모멘트이다[2,4].

실험에 사용된 영교차의 수식은 다음과 같다.

$$Z_m = \sum_{n=m}^N |S[x(n)] - S[x(n-1)]| \cdot h(n-m) \quad (4)$$

이때, $S[x(n)]$ 은 $x(n)$ 이 0이상일때만 1이 된다.

영교차 특징들은 음성/음악 분류기가 크기 정보에 의존하여 분류하는 것을 방지하기 위해서 영교차율을 제외한 나머지 3가지 특징들에 대해서 정규화를 실시하였다.

[2]에서는 분석 윈도우를 이동함에 있어 중첩을 허용하지 않았지만 각 특징들간의 상호조합을 지원하기 위해서 동일하게 중첩을 적용하였다.

중첩 적용 전과 후의 분류능력의 변화를 살펴보면 표 1과 같다. 3가지 분류항목 적용시에는 음악에서의 분류능력의 향상된 반면, 음성과 음성+음악의 분류 결과가 나빠졌다. 2가지 분류항목 적용시에는 중첩 적용 전과 후의 성능변화가 없었다. 이는 상호조합을 지원하기 위해서 허용할 수 있는 범위내의 것이다.

표 1. 영교차의 중첩 적용여부에 따른 분류결과 (백분율)

S : 음성 M : 음악 S+M : 음성+음악

Table 1. The result of discrimination when it use zero-crossings on 3 classes (unit:%).

S : speech, M : music, S+M : speech+music.

구분		영교차(중첩적용전)			영교차(중첩적용후)		
		16Mix	32Mix	64Mix	16Mix	32Mix	64Mix
3가지 분류 항목	S	87.8	79.2	82.9	85.3	80.4	81.7
	M	95.2	71.4	61.9	95.2	95.2	95.2
	S+M	11.1	11.1	16.6	5.5	5.5	5.5
2가지 분류 항목	S	90.1	90.1	90.1	90.1	90.1	90.1
	M	90.0	90.0	90.0	90.0	90.0	90.0

그림 3은 중첩이 적용된 상태의 영교차 특징 파라미터 중 영교차율의 각 분류항목별 변화를 나타낸 것이다.

음성과 음성+음악의 영교차율이 음악의 영교차율보다 적절하게 변화하는 것을 알 수 있으며, 특히 음성의 영교차율은 음성+음악의 영교차율보다 큰 차이로 변화하는 것을 볼 수 있다.

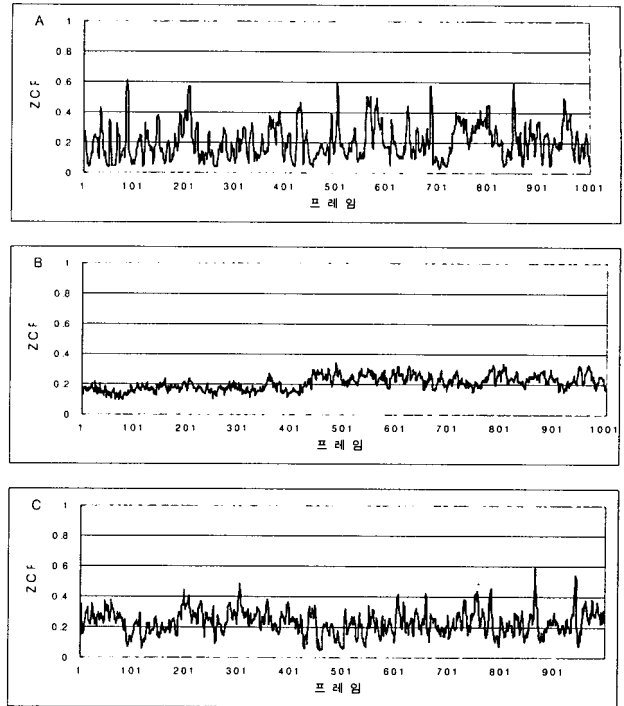


그림 3. 각 분류항목별 영교차율 변화

A : 음성 B : 음악 C : 음성+음악

Fig. 3. The variance of zero-crossing rate in each class,

A : speech, B : music, C : speech+music.

III. 실험 구성

3.1. 데이터베이스 구축

공중파 방송의 50분 분량의 뉴스 3회분, 대중가요, 기타 음악과일들로 데이터베이스를 구성하고 음성, 음악, 음성+음악의 3가지로 분류하였다.

각각의 데이터베이스는 적용한 분류기준별로 균등한 비율로 구성되어 있다. 음성 데이터베이스는 방송 뉴스를 대상으로 3가지 분류기준 (무소음시, 배경소음존재시, 다른 화자의 음성 혼재시)을 적용하였다. 음악 데이터베이스는 방송뉴스와 대중가요를 대상으로 3가지 분류기준 (발라드, 락, 클래식)을 적용하였다. 음성+음악 데이터베이스는 방송뉴스와 대중가요를 대상으로 3가지 분류 기준 (발라드+음성, 락+음성, 클래식+음성)을 적용하였다.

음성 트레인 데이터베이스는 3개 뉴스에서 각각 평균 1분 정도의 30개 신호씩 총 90개 신호, 음악 트레인 데이터베이스는 평균 45초 정도의 36개 신호, 음성+음악 트레인 데이터베이스는 평균 30초 정도의 37개 신호로 구성되어 있다.

테스트 데이터베이스는 3개 뉴스에서 평균 1분 정도의 음성신호 82개, 평균 45초 정도의 음악신호 21개, 평균

45초 정도의 음성+음악신호 18개로 구성되었다. 이 중 음성+음악신호 데이터베이스는 총 18개 중 9개는 대중가요에서 나머지 9개는 실제 뉴스에서 추출하였다.

음향 신호의 시작과 끝 부분의 무음구간은 제거하였으나 음향 신호의 중간에 존재하는 발성간의 무음구간에 대해서는 3초 이내의 것에 한하여 허용하였다[6].

3.2. 실험 시스템

음성/음악을 분류하기 위한 알고리즘은 GMM을 이용하였다.

$$p(x) = \frac{1}{\sqrt{\det \Sigma} (2\pi)^d} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}} \quad (5)$$

이 때 d는 특징 파라미터의 차수이고, Σ 은 공분산 행렬이다.

음성/음악 분류기에서는 입력 음향 신호를 3가지 (음성, 음악, 음성+음악)와 2가지 (음성, 음악)로 분류하였다.

타 논문에서는 2가지 분류항목 (음성, 음악)이 많이 적용되고 있으나, 오디오 인덱싱에서 배경음악 속에서의 음성 신호 인식의 필요성 (광고방송, 방송뉴스의 시작부분 등)이 제기됨에 따라 3가지 분류항목 (음성, 음악, 음성+음악)을 도입하였다[1-3].

음향 신호의 분류 결정은 각 트레인 데이터베이스에 대해 학습시킨 분류항목별 GMM의 모델을 입력 음향 신호에 대응시켜 그 중 최대의 확률을 가지는 것을 선택하였다. 테스트는 각각의 특징에 대해 동일한 3가지 mixture (16, 32, 64)를 적용하여 상호비교 하였다.

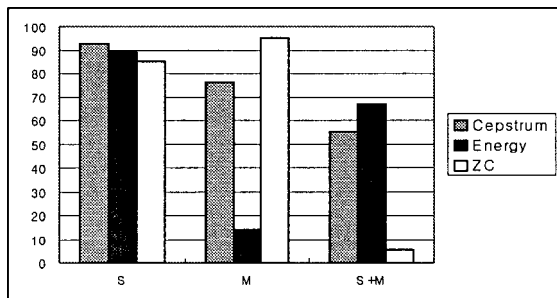


그림 4. 각 특징별 음성/음악 분류결과 (음성, 음악, 음성+음악)
Fig. 4. The result of classification by each features (speech, music, speech&music).

IV. 실험 결과

실험은 우선 각 특징들을 독립적으로 실험한 결과를 분석하여 상호조합에 반영함으로써 분류능력향상을 도모하였다. 3가지 분류항목 적용시에는 음향 신호를 음성, 음악, 음성+음악으로 분류하였으며, 2가지 분류항목 적용시에는 음성, 음악으로 분류하였다.

4.1. 멜캡스트림

실험결과 3가지 분류항목 적용시 16mixture에서, 2가지 분류항목 적용시 모든 mixture에서 결과가 양호하였다. 이 결과로 볼 때 멜캡스트림은 2가지 분류항목 적용시 mixture 변화에 큰 영향을 받지 않는다고 볼 수 있다.

3가지 분류항목 적용시에는 음악 구분에 있어서 영교차 특징에 비해 좋지 않았다. 2가지 분류항목 적용시에는 음성에 대해서는 모든 테스트 파일을 정확히 분류하였고, 음악에 대해서는 타 특징에 비해 가장 좋은 결과를 나타내었다.

4.2. 에너지

에너지는 3가지 분류항목을 적용하였을 경우 64mixture에서, 2가지 분류항목을 적용하였을 경우 32mixture에서 가장 좋은 결과를 나타내었다. 분류 결과에 의하면 타 특징에 비해 음악에 대한 분류능력이 낮았다.

실험결과 3가지 분류항목 적용시에는 음성+음악 분류능력은 타 특징들에 비해 양호하였으나 음악의 분류능력은 가장 낮았다. 또한 음악 분류에서 다른 특징들과 달리 음성으로 오분류되는 현상이 나타났다.

2가지 분류기준 적용시에는 음성 분류능력에서는 양호한 결과를 나타내고 있으나, 음악 분류능력은 타 특징들에 비해 현저히 낮았다.

음악이 오분류되는 양상을 살펴보면 평균적으로 오분류된 것 중 38.2%가 음성으로 47.6%가 음성+음악으로 분류되었다.

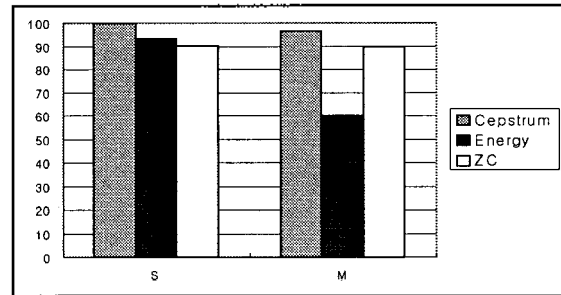


그림 5. 각 특징별 음성/음악 분류결과 (음성, 음악)
Fig. 5. The result of classification by each features (speech, music).

4.3. 영교차

GMM의 mixture 수는 16, 32, 64mixture를 적용하여 실시하였다.

표 4를 보면, 3가지 분류항목 적용시에는 타 특징에 비해 음악의 분류능력이 가장 우수한 것으로 나타났다. 이에 반해 음성+음악 분류능력은 타 특징에 비해 가장 떨어지는 것으로 나타났다. 또한 음성 분류에서 다른 특징과는 달리 음악으로 오분류되는 현상이 나타났다.

2가지 분류항목 적용시에는 음성과 음악 모두에서 양호한 분류결과를 얻을 수 있었다. 또한 mixture의 변화에 영향을 받지 않았다.

영교차 특징 적용시 음성+음악이 오분류되는 양상을 살펴보면 평균적으로 오분류된 것 중 44.5%가 음성으로 분류되었고, 50%가 음악으로 분류되었다.

4.4. 특징간의 상호조합

특징들간의 상호조합은 다음과 같이 실시하였다.

에너지+영교차 조합은 음악 분류능력이 좋지 않은 에너지와 음성+음악 분류능력이 좋지 않은 영교차를 상호 보완함으로써 전체적인 분류능력의 향상을 도모하는데 목적이 있다.

표 2를 보면 3가지 분류항목 적용시에는 64mixture에서 좋은 결과를 나타내었고, 2가지 분류항목 적용시에는 16mixture에서 좋은 결과를 나타내었다.

3가지 분류항목 적용시에는 에너지 단독 적용시에 비해서 음성과 음성+음악 분류에서는 성능저하가 관찰되었고, 음악 분류에서는 큰 성능향상이 관찰되었다. 영교차 단독 적용시에 비해서는 음성과 음성+음악 분류 결과는 향상되었지만, 음악 분류 결과가 크게 저하되었다.

2가지 분류항목 적용시에는 에너지 단독 적용시에 비해서 음악의 분류능력이 향상되었으며, 영교차 단독 적용시에 비해서 모든 분류항목의 분류능력이 향상되었다.

멜캡스트럼+영교차 조합은 모든 분류항목에 대해 양호한 결과를 나타내고 있는 멜캡스트럼에 음악 분류에서 좋은 결과를 나타내고 있는 영교차를 조합하였다.

표 2. GMM 분류결과 (에너지+영교차, 백분율)

S : 음성 M : 음악 S+M : 음성+음악

Table 2. Result of GMM (energy and zero-crossings, unit: %),

S : speech, M : music, S+M : speech&music.

구 분		16 mixture	32 mixture	64 mixture
3가지 분류항목	S	89.0	85.3	86.5
	M	76.1	85.7	80.9
	S&M	22.2	22.2	22.2
2가지 분류항목	S	92.3	92.3	90.1
	M	96.6	93.3	90.0

표 3을 보면 3가지 분류항목 적용시에는 64mixture에서 가장 좋은 결과를 나타내었고, 2가지 분류항목 적용시에는 32, 64mixture에서 가장 좋은 결과를 나타내었다.

3가지 분류항목 적용시에는 멜캡스트럼 단독 적용시에 비해서 음성, 음악 분류능력의 향상이 관찰되었으며, 영교차 단독 적용시에 비해서 음성, 음성+음악 분류능력이 향상되었다.

2가지 분류항목 적용시에는 멜캡스트럼 단독 적용시에 비해서 음성 분류능력이 약간 저하되었으며, 영교차 단독 적용시에 비해서 모든 분류항목의 분류능력이 향상되었다.

멜캡스트럼+에너지 조합에서는 모든 분류항목에서 양호한 분류능력을 보여주고 있는 멜캡스트럼과 음성+음악에서 양호한 분류능력을 보이는 에너지를 결합시킴으로서

음성+음악 분류능력 향상을 도모하였다.

표 3. GMM 분류결과 (멜캡스트럼+영교차, 백분율)

S : 음성 M : 음악 S+M : 음성+음악

Table 3. Result of GMM (Mel-cepstrum and zero-crossings, unit: %),

S : speech, M : music, S&M : speech&music.

구 분		16 mixture	32 mixture	64 mixture
3가지 분류항목	S	96.3	93.9	95.1
	M	80.9	76.1	85.7
	S&M	44.4	55.5	44.4
2가지 분류항목	S	98.9	98.9	98.9
	M	93.3	96.6	96.6

표 4를 보면 3가지 분류항목 적용시에는 16mixture에서 좋은 결과를 나타내었고, 2가지 분류항목 적용시에는 16mixture에서 좋은 결과를 나타내었다.

3가지 분류항목 적용시에는 멜캡스트럼 단독 적용시에 비해서 음악, 음성+음악 분류능력이 저하되었으며, 에너지 단독 적용시에 비해서는 음성, 음악 분류능력이 향상되었다.

2가지 분류항목 적용시에는 멜캡스트럼 단독 적용시에 비해서 음악을 100% 분류하였으며, 에너지 단독 적용시에 비해서 전 분류항목의 분류성능이 향상되었다.

멜캡스트럼+에너지+영교차의 조합에서는 모든 분류 항목에서 양호한 특징인 멜캡스트럼에 영교차에 의한 음악의 분류능력 향상과 에너지에 의한 음성+음악의 분류능력 향상을 도모하였다.

표 4. GMM 분류 결과 (멜캡스트럼+에너지, 백분율)

S : 음성 M : 음악 S+M : 음성+음악

Table 4. Result of GMM (Mel-cepstrum and energy, unit: %),

S : speech, M : music, S+M : speech&music.

구 분		16 mixture	32 mixture	64 mixture
3가지 분류항목	S	96.3	93.9	95.1
	M	52.3	57.1	47.6
	S&M	61.1	61.1	66.6
2가지 분류항목	S	96.7	96.7	98.9
	M	100	100	100

표 5를 보면 3가지 분류항목 적용시에는 32mixture에서 가장 좋은 결과를 나타내고 있으며, 2가지 분류항목 적용시에는 16mixture에서 가장 좋은 결과를 나타내고 있다.

실험에 의하면 특징 상호간 조합은 특징 단독으로 사용하였을 때보다 더 좋은 결과를 나타내고 있다.

결과에서 음악이 음성으로 오분류되는 것보다 음성이 음악으로 오분류되는 것이 더욱 치명적이다. 음악으로 분류되면 차후 실험에서 제외되기 때문이다.

연산시간면에서는 에너지 (2초), 영교차 (3.4초), 멜캡스트럼 (7.4초)순으로 연산시간이 소요되었고, GMM의

mixture수의 증가에 비례하여 연산시간도 비례적으로 증가하였다. 이는 실험이 실시간성을 요구하지 않으므로 문제되지 않는다.

타 논문의 실험결과와 비교하면, [2]에서는 2가지 분류 기준 적용시 98.4%의 결과를 보였고, 본문에서는 음성 98.9%, 음악 100%의 결과를 보였다.

표 5. GMM 분류결과 (멜캡스트럼+에너지+영교차, 백분율)
 S : 음성 M : 음악 S+M : 음성+음악
 Table 5. Result of GMM (Mel-cepstrum, energy and zero-crossings, unit:%),
 S : speech, M : music, S+M : speech+music.

구분		16 mixture	32 mixture	64 mixture
3가지 분류항목	S	96.3	95.1	95.1
	M	71.4	71.4	61.9
	S&M	44.4	55.5	55.5
2가지 분류항목	S	98.9	98.9	98.9
	M	100	96.6	96.6

V. 결론 및 향후계획

본 논문에서는 음성/음악 분류기에 대해 다루었다.

음성/음악 분류기에서 사용된 특징들은 멜캡스트럼, 에너지, 영교차간의 상호조합이다. 분류 알고리즘은 GMM을 사용하였으며, 3가지의 mixture (16, 32, 64)를 일괄적으로 적용하였다.

실험결과 각 특징들을 단독으로 적용한 분류결과에 비해 특징간 상호조합을 적용한 분류결과가 보다 양호하였다. 3가지 분류항목 적용시에는 멜캡스트럼+영교차 조합이 가장 양호한 결과를 보여주었고, 2가지 분류항목 적용시에는 멜캡스트럼+에너지 조합, 멜캡스트럼+에너지+영교차 조합이 가장 양호한 결과를 나타내었다.

본 논문에서는 특징조합의 구성요소에 대해 하나의 특징공간에서 GMM을 적용하였으나, 차후에는 특징조합의 구성요소 각각에 대해 특징공간을 부여한음성/음악 분류에 대해 연구할 계획이다.

감사의 글

본 연구는 정보통신분야 우수학교 지원사업 연구비 지원으로 이루어졌습니다.

참고 문헌

1. Michael J. Carey, "A Comparison of Features for Speech, Music Discrimination", Proc. ICASSP, vol. 1, 1999.
2. John Saunders, "Real-Time Discrimination of Broadcast Speech / Music", Proc. ICASSP, pp. 993-996, 1996.
3. E. Scheier and M. Slaney, "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator", Proc.

ICASSP, pp. 1331-1334, 1997.

4. B. Kedam, "Spectral Analysis and Discrimination by Zero-Crossing", Proc. IEEE, vol. 74, No. 11, pp. 1477-1493, Nov. 1986.
5. John D. Hoyt, "Detection of Human Speech in Structured Noise", Proc. IEEE, pp. 237-240, 1994.
6. T. Hain, "Segment Generation and Clustering in the HTK Broadcast News Transcription System", In Proceedings of the Broadcast News Transcription and Understanding Workshop, 1998.
7. Jean-Luc Gauvain, Lori Lamel, "Partitioning and Transcription of Broadcast News Data", Proc. ICSLP, Sydney, 1998.
8. 이경록, 서봉수, 김진영, "음성음악분류를 위한 특징 비교", 한국음향학회 학술발표대회 논문집 제19권, pp. 157-160, 2000.

▲ 이 경 록 (Kyong Rok Lee) 1974년 10월 29일생



1997년 3월 : 호남대학교 전자공학과 졸업
 1998년 8월 ~ 전남대학교 대학원 정보통신협동 석사과정
 * 주관심분야 : 멀티미디어 인태싱

▲ 서 봉 수 (Bong Su Seo) 1973년 5월 26일생

1999년 2월 : 조선대학교 제어계측공학과 졸업
 1998년 3월 ~ 전남대학교 대학원 전자공학과 석사과정
 * 주관심분야 : 음성인식

▲ 김 진 영 (Jin Young Kim)

1986년 : 서울대학교 전자공학과 졸업
 1988년 : 서울대학교 전자공학과 석사
 1994년 : 서울대학교 전자공학과 박사
 1994년 ~ 1995년 : 한국통신 소프트웨어 연구소
 1995년 ~ 전남대학교 전자공학과 조교수
 * 주관심분야 : 음성인식, 음성합성, 멀티모달 MMI