

# MFCC의 단구간 시간 평균을 이용한 음성/음악 판별 파라미터 성능 향상\*

최무열(부산대), 김형순(부산대)

## <차 례>

- |   |            |
|---|------------|
| 1. 서 론  | 4. 실험 및 결과 |
| 2. Mean of Minimum Cepstral Distances (MMCD) 방법 | 4.1. 실험환경  |
| 3. MFCC의 단구간 시간 평균을 이용한 MMCD                    | 4.2. 실험결과  |
|   | 5. 결 론     |

## <Abstract>

### Improving Speech/Music Discrimination Parameter Using Time-Averaged MFCC

Mu Yeol Choi, Hyung Soon Kim

Discrimination between speech and music is important in many multimedia applications. In our previous work, focusing on the spectral change characteristics of speech and music, we presented a method using the mean of minimum cepstral distances (MMCD), and it showed a very high discrimination performance. In this paper, to further improve the performance, we propose to employ time-averaged MFCC in computing the MMCD. Our experimental results show that the proposed method enhances the discrimination between speech and music. Moreover, the proposed method overcomes the weakness of the conventional MMCD method whose performance is relatively sensitive to the choice of the frame interval to compute the MMCD.

\* Keywords : Speech/music discrimination, Cepstral distance, Time-averaged MFCC.

\* 이 논문은 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활 지원 로봇 기술개발사업)의 일환으로 수행됨.

## 1. 서론

오디오 신호로부터 음성과 음악을 자동으로 구별하는 기술은 디지털 음향매체의 발전과 더불어 오디오 신호를 저장하거나 다루는 일이 늘어남에 따라 그 중요성이 높아지고 있다. 음성/음악 판별 시스템은 음성인식 전처리 과정에서 인식기로 들어오는 음악 신호를 제외시킴으로써 음성인식 성능을 향상시킬 수 있다. 또한 음성과 음악을 구분하여 각각의 방식에 적합한 압축방식을 적용할 수 있다. 그 외에도 멀티미디어 정보검색 등에 유용하게 사용될 수 있다.

우수한 음성/음악 판별 성능을 얻기 위해서는 적절한 특징 파라미터를 선택하는 것이 매우 중요하다. 지금까지의 연구에 사용된 특징 파라미터로는 시간 영역에서의 특징을 이용한 영교차율(zero crossing rate(ZCR))과 그 변화 특성을 이용한 high ZCR ratio(HZCRR)[1][2], 신호의 단구간 에너지의 변화를 측정한 low short-time energy ratio(LSTER)[3] 등이 있다. 스펙트럼 영역 특징을 이용한 파라미터로 스펙트럼의 무게중심을 이용한 spectral centroid, 스펙트럼의 변화의 차이를 이용한 spectral flux(SF)[1]와 캡스트럼 거리를 이용한 cepstral distance(CD)[4], 음성의 스펙트럼 포락선에 대한 정보를 잘 나타내는 line spectrum pair(LSP)[5] 거리 등이 있으며, 인접한 여러 프레임들 사이의 캡스트럼 거리의 최소값의 평균을 이용하는 mean of minimum cepstral distances (MMCD)[6] 파라미터가 있다. 그 외에도 음성의 특징을 잘 반영하는 4 Hz modulation energy[1], 음악의 리듬을 이용한 pulse metric[1], 그리고 음소인식 결과를 기반으로 하는 entropy와 dynamism[7] 등의 다양한 특징 파라미터들이 제안되었다.

선행연구에 따르면 이들 파라미터 중에서 MMCD가 성능 면에서 매우 우수한 것으로 보고되었다[6]. 그러나 MMCD는 인접한 여러 프레임들과 비교하여 구한 여러 개의 캡스트럼 거리 중에서 최소값을 그 프레임에서의 캡스트럼 거리로 선택하여 이들의 평균을 구하는 것으로서, 캡스트럼 거리의 최소값을 구하는 프레임 범위가 성능에 상당한 영향을 미치는데 이를 실험을 통해 최적화해야 한다는 단점이 있다. 이는 훈련용 데이터에서 관찰되지 않는 특성을 가지는 다양한 음성과 음악 데이터에 대해서 안정적인 성능을 보장하기 어려울 수도 있다는 문제를 초래한다. 이 문제의 해결을 위해 복수의 프레임 범위에 대한 MMCD들을 함께 적용하는 다차원 MMCD[8]가 제안되었다. 다차원 MMCD는 기존의 MMCD의 단점을 극복할 수 있는 장점이 있으나, 계산량이 상당히 증가하고 복수의 프레임 범위를 선택해야 하는 문제가 여전히 존재한다.

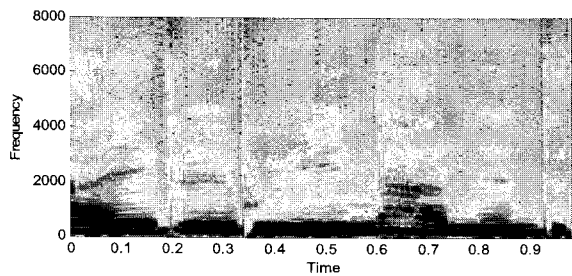
본 논문에서는 기존 MMCD의 성능 향상을 위해 MMCD를 구하기 전에 사용되는 MFCC에 단구간 시간 평균을 취하는 방법을 통해 MMCD를 추출하는 구간에 상관없이 높은 성능을 얻는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 MMCD 파라미터에 대해 설명하고,

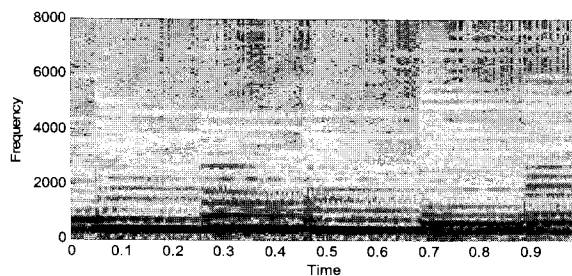
3장에서는 MMCD의 문제점을 개선하기 위한 방안으로 MFCC의 단구간 시간 평균을 사용한 MMCD 파라미터를 제안한다. 4장에서는 실험 방법 및 결과를 기술하며, 마지막으로 5장에서 결론을 맺는다.

## 2. Mean of Minimum Cepstral Distances (MMCD) 방법

음성은 다양한 음소들이 번갈아 나타나므로 짧은 구간에서도 스펙트럼 포락선의 변화가 자주 일어난다. 이에 비해 음악의 경우 동일한 악기 그룹의 연주가 계속되는 동안에는 빠른 템포로 음정 변화가 진행되더라도 스펙트럼 포락선은 비교적 유사한 형태를 유지한다. 이와 같은 특성은 <그림 1>의 음성과 음악의 스펙트로그램에 잘 나타나있다.



(a) 음성



(b) 음악

<그림 1> 음성과 음악의 스펙트로그램 예

<그림 1>은 각각 1초 동안 음성과 음악의 전형적인 스펙트로그램을 나타낸 것이다. <그림 1>(a)에서 보듯이 음성은 짧은 구간에서도 스펙트럼 포락선의 변화가 자주 일어난다. 동일한 시간 동안에 음악의 스펙트럼 변화를 나타내는 <그림 1>(b)에서는 여러 프레임을 이동하여 스펙트럼 포락선을 비교해 보아도 음성에 비

해 상대적으로 비슷한 모양을 가진다. 이와 같이 음성과 음악의 스펙트로그램 특성을 이용한 파라미터가 캡스트럼 거리를 이용한 파라미터이다.

신호의 스펙트럼 변화 특성을 이용한 MMCD 파라미터를 설명하기 위해, 우선 스펙트럼 포락선(spectral envelope)을 표현하는 캡스트럼의 저차 성분들을 사용하여 다음 식과 같은 캡스트럼 거리를 정의한다.

$$CD(n, d) = \sqrt{\sum_{k=1}^K (c(n+d, k) - c(n, k))^2} \quad (1)$$

여기서  $CD(n, d)$ 는  $n$ 번째 프레임에서의 캡스트럼 거리이고,  $K$ 는 캡스트럼 차수,  $c(n, k)$ 는  $n$ 번째 입력 프레임에 대한  $k$ 차 캡스트럼 계수 값을 나타낸다.

그리고 캡스트럼 거리를 바탕으로 인접한 여러 프레임들과 비교하여 구한 여러 개의 캡스트럼 거리 중에서 최소값을 그 프레임에서의 캡스트럼 거리로 선택한다. 캡스트럼 거리의 최소값을 구하는 식은 다음과 같다.

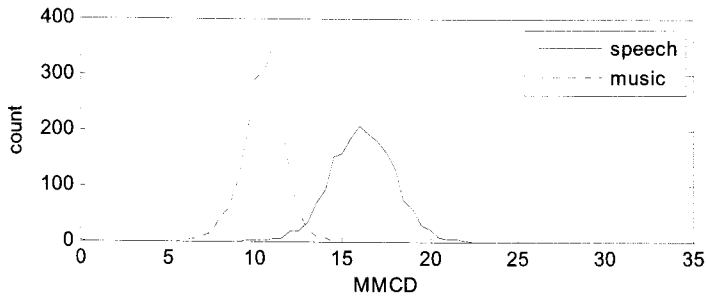
$$MCD(n, d_1, d_2) = \min_{d_1 \leq d \leq d_2} CD(n, d) \quad (2)$$

여기서  $MCD(n, d_1, d_2)$ 는  $n$ 번째 프레임과 인접한 프레임들과의 캡스트럼 거리 중에서 최소값을 나타내며,  $d_1$ 과  $d_2$ 는 비교대상 프레임의 범위를 나타낸다. 이들 최소값의 평균이 MMCD이며 MCD의 평균이라는 의미로  $\mu_{MCD}$ 라고 표현하고, 다음 식과 같이 정의한다.

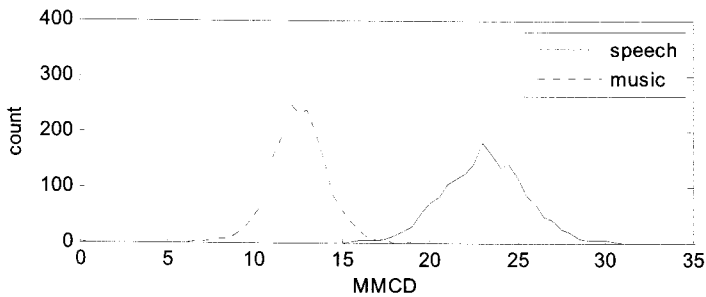
$$\mu_{MCD}(d_1, d_2) = \frac{1}{N-d_2} \sum_{n=1}^{N-d_2} MCD(n, d_1, d_2) \quad (3)$$

<그림 2>는  $d_1$ 과  $d_2$ 에 따른 음성과 음악의 MMCD 히스토그램을 나타낸 것이다. <그림 2>(a)와 <그림 2>(b)에서 보는 바와 같이 MMCD는 캡스트럼 거리의 최소값을 구하는  $d_1$ 과  $d_2$  구간의 선택에 따라 음성/음악 판별 성능의 차이가 크게 나타난다.

<그림 3>은  $d_1$ 과  $d_2$  구간을 다양하게 변화 시킨 경우의 음성/음악 판별 성능을 나타낸 것이다. <그림 3>(a)에서  $d_1=40$  ms,  $d_2=150$  ms 인 경우 음성과 음악의 판별 성능이 99.69%에 이르며, <그림 3>(b)에서  $d_1=30$  ms 또는  $d_1=40$  ms,  $d_2=200$  ms 인 경우 99.61%인 것을 볼 수 있다. 그러나 <그림 3>(b)에서  $d_1=10$  ms,  $d_2=200$  ms 인 경우 판별 성능은 97.86%까지 떨어진다. 이와 같이 MMCD 파라미터는 최



(a)  $d_1=10$  ms,  $d_2=150$  ms



(b)  $d_1=40$  ms,  $d_2=150$  ms

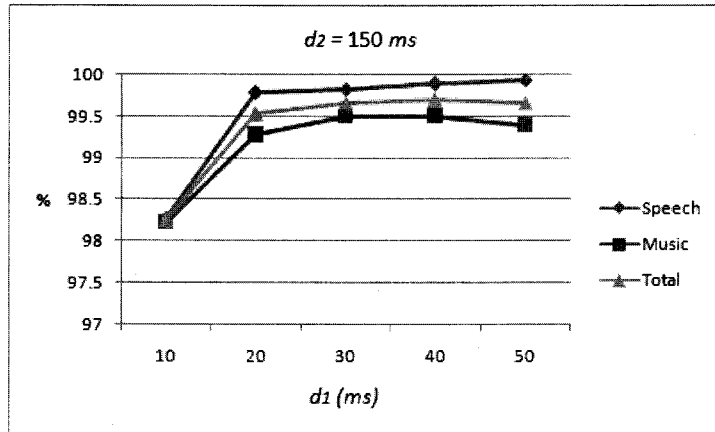
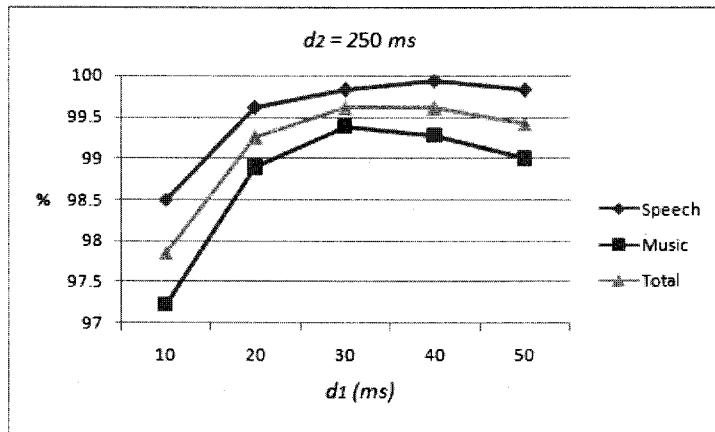
<그림 2> 음성과 음악의 구간별 히스토그램

소값을 구하는 최적의  $d_1$ 과  $d_2$  구간을 선택하였을 경우 매우 높은 성능을 나타내는 장점이 있으나 그렇지 못한 경우 구간별 성능차이를 보이며, 특히  $d_1=10$  ms처럼 인접한 구간에서 성능 저하가 크게 나타난다.

이와 같이 기존의 MMCD 파라미터는 캡스트럼 거리의 최소값을 구하는 프레임 범위를 실험을 통해 최적화해야 하는 단점이 있을 뿐 아니라, 그렇게 하더라도 훈련용 데이터에서 관찰되지 않는 특성을 가지는 다양한 음성과 음악 데이터에 대해서 안정적인 성능을 보장할 수 없다는 문제가 있다. 따라서 본 논문에서는 이러한 문제를 해결하기 위해 MMCD를 구하기 전에 사용되는 MFCC에 단구간 시간 평균을 취하는 방법을 통해 MMCD를 추출하는 구간에 상관없이 높은 성능을 보이는 방법을 제안한다.

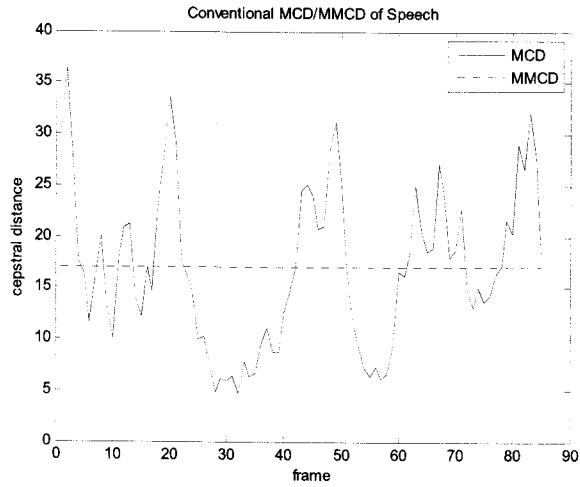
### 3. MFCC의 단구간 시간 평균을 이용한 MMCD

신호의 스펙트럼 포락선의 변화 특성을 이용한 MMCD는 스펙트럼 포락선 정보인 MFCC를 이용하여 계산하게 되는데 이 때 MFCC의 시간에 따른 변화의 정

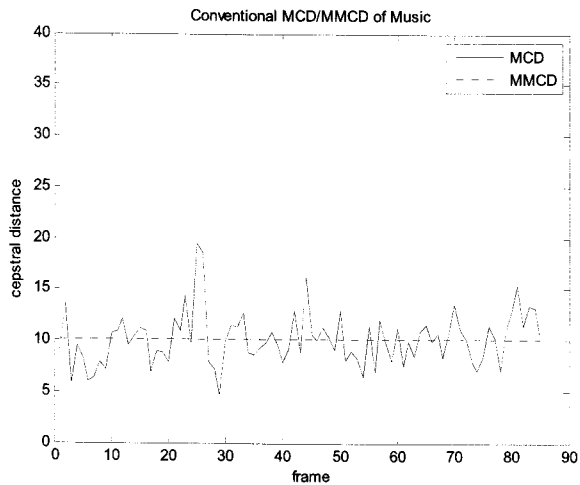
(a)  $d_1=10\sim 50$  ms,  $d_2=150$  ms(b)  $d_1=10\sim 50$  ms,  $d_2=250$  ms

&lt;그림 3&gt; 기존 MMCD의 구간별 음성/음악 판별 성능

도가 MMCD의 분포를 결정한다. 일반적으로 다양한 음소들이 교대로 나타나는 음성 신호의 시간에 따른 MFCC 변화량이 음악의 변화량보다 크기 때문에 최소 켈스트럼 거리의 시간에 따른 변화도 음성이 음악보다 크며 그에 따른 MMCD의 분포를 나타내면, <그림 2>(a) 및 <그림 2>(b)와 같이, 그 판별력이 대체로 높다는 것을 볼 수 있다. <그림 4>(a)는 1초간의 음성의 최소 켈스트럼 변화 곡선과 그에 해당하는 MMCD 값이며, <그림 4>(b)는 1초간의 음악의 최소 켈스트럼 변화 곡선과 그에 해당하는 MMCD 값을 나타낸 것이다. 음성의 최소 켈스트럼 변화 곡선은 일반적으로 변화가 크게 나타나면서 큰 값의 빈도가 높아 MMCD 값을 상승시킨다. 반면 음악의 최소 켈스트럼 변화 곡선은 일반적으로 변화가 작으며 작은 값의 빈도가 높아 낮은 MMCD 값을 갖는다.



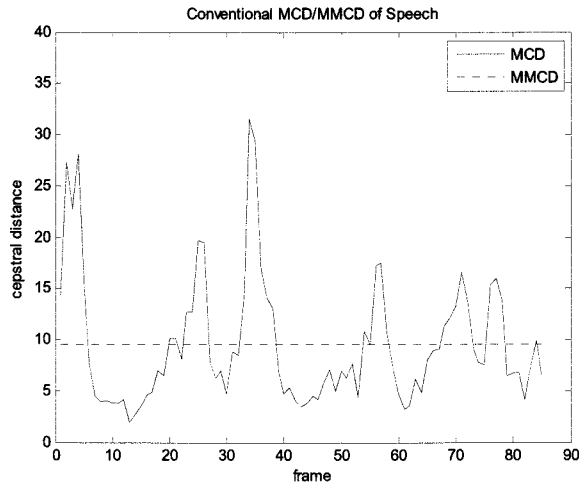
(a) 음성 (MMCD=16.99)



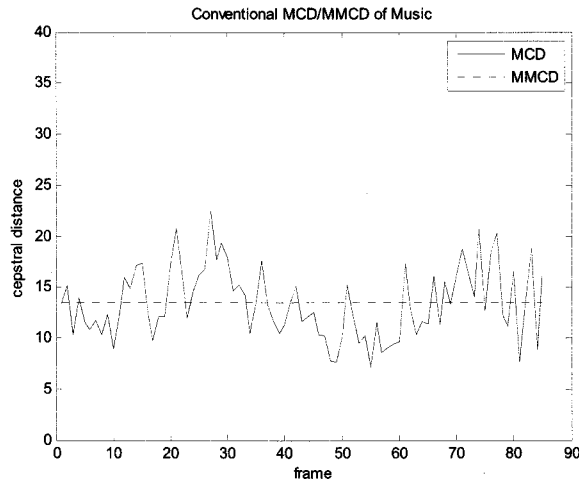
(b) 음악 (MMCD=10.13)

<그림 4> 1초 구간의 최소 캡스트럼 변화 곡선과 MMCD 값 (판별이 잘되는 경우)

그러나 음악의 경우 다양한 장르의 음악이 존재하며, 템포와 음정에 의해서 짧은 기간 갑작스런 스펙트럼 변화에 의해 MMCD 값이 커지는 경우가 있다. 또한 음성의 경우에도 지속시간이 긴 유성음들이 연속되는 경우 음성의 최소 캡스트럼 거리가 음악의 최소 캡스트럼 거리보다 작게 나타나는 경우가 발생하여 음성의 MMCD 분포가 음악의 분포와 겹치는 현상이 나타난다. <그림 5>(a)는 낮은 값의 빈도가 많아 음악의 MMCD보다 낮은 MMCD 값을 나타내고 있으며 이는 음성 구간이 음악 구간으로 오인식 되는 주된 요인이 된다. 또한 <그림 5>(b)는 변화가



(a) 음성 (MMCD=9.55)



(b) 음악 (MMCD=14.10)

<그림 5> 1초 구간의 최소 켈스트럼 변화 곡선과 MMCD 값 (판별 오류가 발생하는 경우)

심한 음악의 일부로 최소 켈스트럼 변화가 크게 나타나면서 높은 MMCD 값을 나타내 음성으로 오인식 되는 경우를 발생시키는 요인이 된다.

이런 경우 MMCD의 판별력이 떨어지는 문제를 해결하기 위해 본 논문에서는 MMCD를 구하는 과정에서 계산하는 MFCC에 단구간 시간 평균을 취하여 음성과 음악의 변화 특성의 변별력을 높일 수 있도록 하였다. MFCC에 단구간 시간 평균을 취하는 과정은 식 (4)와 같으며 수정된 켈스트럼을 이용한 켈스트럼 거리를 구



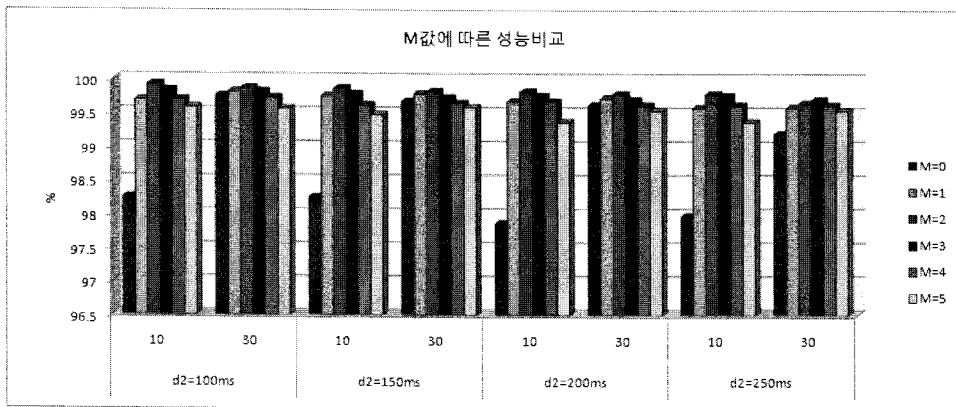
하는 것은 식 (5)와 같다. 본 논문에서 제안한 MMCD는 식 (5)를 이용하여 식 (2)와 식 (3)을 수행하여 얻는다.

$$c_{MA}(n, k) = \frac{1}{2M+1} \sum_{m=-M}^M c(n+m, k) \quad (4)$$

$$CD(n, d) = \sqrt{\sum_{k=1}^K (c_{MA}(n+d, k) - c_{MA}(n, k))^2} \quad (5)$$

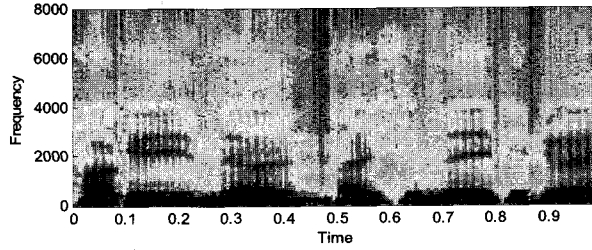
여기서  $2M+1$ 은 단구간 시간 평균을 취하는 윈도우 크기이다.

본 논문에서는 윈도우 크기를 변화시켜 본 결과  $M=2$ 에서 최고 성능을 보였다.  $M$ 의 변화에 따른 실험 결과는 <그림 6>에 나타났다.

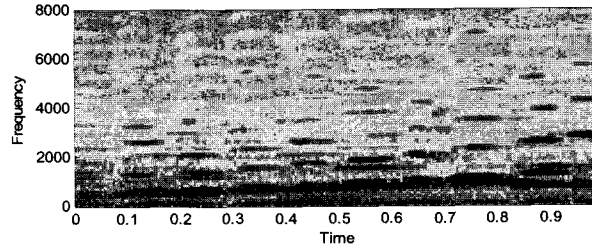


<그림 6> 윈도우 크기에 따른 인식 성능 비교

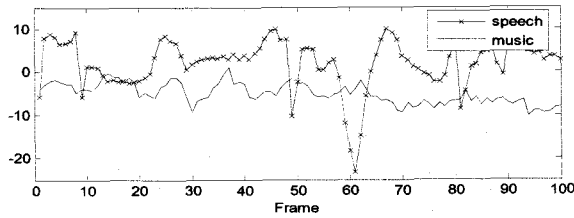
<그림 7>에서는 음성과 음악의 1초 구간의 스펙트로그램과 각각의 신호에서 추출한 MFCC의 평균을 취하기 전과 후의 MFCC 1차원 값의 시간에 따른 변화특성을 나타냈다. <그림 7>(c)에 나타난 MFCC 1차원 값의 시간에 따른 변화특성을 살펴보면 음성과 음악 모두 프레임 간의 변동이 매우 심하게 나타나는 것을 볼 수 있다. 이런 프레임 간의 잦은 변동은 프레임 간의 최소값을 구하는 경우에 음성과 음악의 특성을 혼동시키는 경우를 발생시킨다. 본 논문에서는 프레임 간의 변동이 매우 심한 MFCC의 캡스트럼에 단구간 시간 평균을 취하였다. <그림 7>(d)에서 보는 바와 같이 음성 음악 모두 프레임 간의 변동이 감소된 것을 볼 수 있는데 이로 인한 MMCD의 분포 변화는 <그림 8>에서 확인할 수 있다. <그림 8>은 <그림 2>(a)와 동일한 조건에서 제안한 방법의 MMCD를 취한 결과와 비교한 것이다.



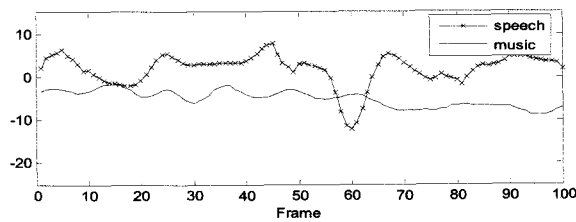
(a) 음성의 스펙트로그램



(b) 음악의 스펙트로그램



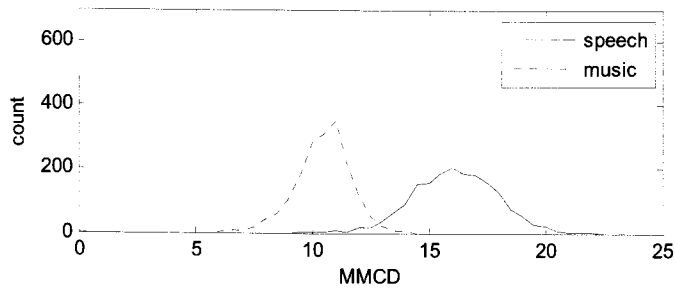
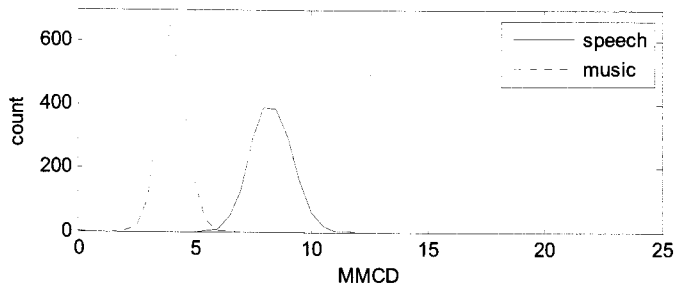
(c) 1차 MFCC의 시간에 따른 변화 곡선



(d) 1차 MFCC의 단구간 시간 평균 곡선

<그림 7> 음성 음악 신호의 스펙트로그램과 시간에 따른 MFCC 1차 변화곡선

제안한 방법의 MMCD 분포는 <그림 8>(b)에서 보는 바와 같이 기존의 MMCD와 비교할 때 음성과 음악의 분별력이 더 높아진 것을 확인할 수 있는데 이는 음성과 음악에서 나타나는 MMCD의 문제점이 모두 개선된 것에 기인한다고 할 수 있다. 즉, 음악에서 프레임 간의 변동이 감소하면서 음악의 천천히 변화하는 특성이 MMCD 계산시 더 뚜렷해진 결과이며, 음성의 경우 프레임 간의 변동이 동일하

(a)  $d_1=10\text{ms}$ ,  $d_2=150\text{ms}$  에서의 기존 MMCD 히스토그램(b)  $d_1=10\text{ms}$ ,  $d_2=150\text{ms}$  에서의 제안한 방법의 MMCD 히스토그램

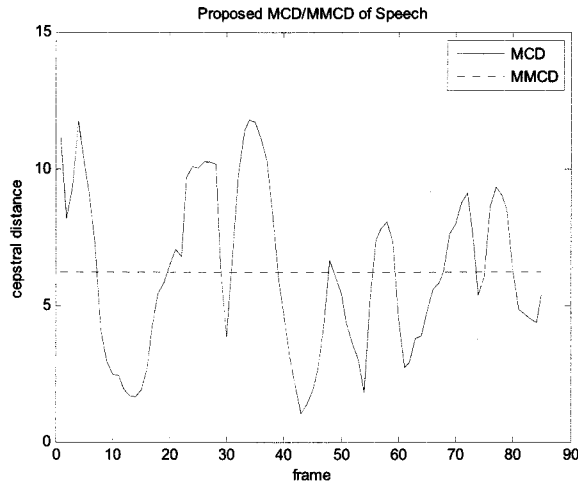
&lt;그림 8&gt; 동일 구간에서의 기존 MMCD와 제안한 방법의 MMCD 히스토그램 비교

게 감소되면서 음성의 MMCD 값이 줄어들기는 했으나 음소의 변화특성은 유지되므로 음악의 MMCD 값과 구별 되게 된다. 이로 인해 음성/음악의 판별성능이 향상되었다. <그림 9>는 기존의 MMCD에 의해 음성/음악 판별에 오류가 발생한 <그림 4>에 제안한 방법을 적용한 결과이다. 기존의 음성 MMCD 값은 음악의 MMCD 값보다 작게 나타났으나 제안한 방법에 의한 MMCD 값은 음악의 MMCD 보다 커진 것을 확인할 수 있다.

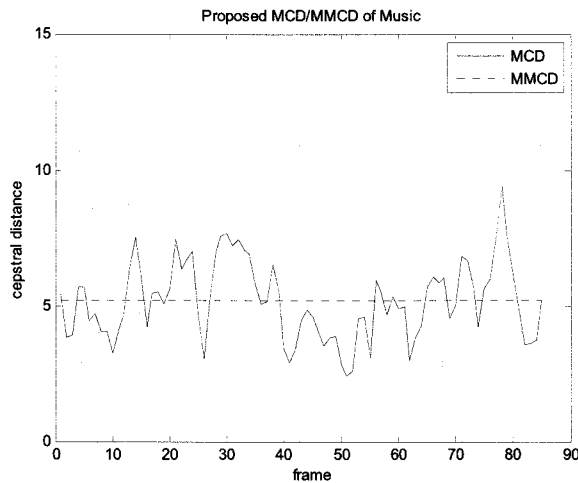
## 4. 실험 및 결과

### 4.1. 실험환경

본 논문에서 제안한 방법의 성능 평가를 위해 음성 데이터는 국어공학센터에서 구축한 phonetically balanced sentence(PBS) 589 문장에 대한 남녀 50명분의 발성 데이터 약 13시간 분량을 사용하였다. 그리고 음악 데이터는 다양한 클래식 음악과 장르 음악을 제공하는 RWC 음악 데이터베이스를 사용하였다. 음악의 훈련 데



(a) 음성 (MMCD=6.21)

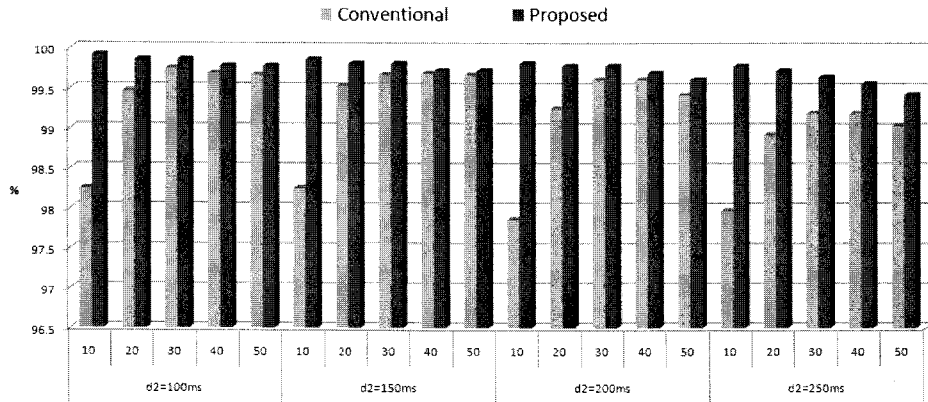


(b) 음악 (MMCD=5.34)

&lt;그림 9&gt; 1초 구간의 최소 캡스트럼 변화 곡선과 제안된 방법의 MMCD 값

이터는 5시간 분량의 클래식 음악(Symphonies, Concerti, Orchestral, Chamber, Solo, March)을 사용하였다. 음악 데이터는 16 kHz로 다운샘플링하고 16 bit로 양자화하여 음성 데이터와 동일한 조건이 되도록 만들었다.

테스트 데이터는 훈련 시 사용되지 않고 clean 환경에서 녹음한 음성 데이터와 다양한 장르(Jazz (Modern, Fusion), Latin (Bassa Nova, Samba, Reggae, Tango), Blues, Folk, Country, Indian, Flamenco)의 음악을 15초씩 번갈아 나타나도록 만든 1시간 분량의 데이터로 구성하였다. 테스트에 사용한 음악은 사람의 목소리 또는



<그림 10> 기존 파라미터와 제안한 파라미터와의 성능 비교

노래소리가 포함되지 않은 연주곡으로 구성하였다.

음성/음악 판별을 위한 분류기로는 특징 벡터의 분포를 몇 개의 가우시안 분포들의 가중합으로 표현하는 GMM 분류기를 사용하였다. 실험에 사용된 믹스처 수를 1, 2, 4, 8 로 증가시킨 경우 1개와 2개에서는 성능 차이가 미미 했으며 4, 8개로 증가 시킨 경우 근소한 차이로 감소하는 결과를 얻었다. 따라서 본 논문에서는 믹스처 1개만을 사용하는 가우시안 분류기를 사용하였는데 이는 믹스처에 의한 계산량의 증가 없이도 높은 성능을 나타내기 때문이다. 칩스트럼 거리 측정을 위해 MFCC를 이용하였고, 이 때 프레임의 크기는 20 ms, 프레임의 이동은 10 ms로 하였다. 이렇게 구한 MFCC를 이용하여 1초 구간마다 MMCD를 계산하여 파라미터로 사용하였다.

#### 4.2. 실험결과

기존의 MMCD와 제안한 방법의 성능을 비교하기 위해 다양한 구간에 따른 음성/음악 판별 성능을 <그림 10>에 나타내었다. 그림에서 보는 바와 같이 제안한 방법은 실험에 사용된 모든  $d_1$ ,  $d_2$  구간에 대해 일관적으로 기존 방법보다 높거나 동일한 성능을 나타내고 있다. 특히 기존 MMCD는  $d_2$  구간에 상관없이 모든  $d_1 = 10$  ms인 구간에서 다른 구간에 비해 매우 낮은 성능을 보인 반면 제안한 방법의 결과에서는  $d_1 = 10$  ms인 구간의 성능이 매우 높은 성능을 보인다. 제안한 방법의 최고 성능은  $d_2 = 100$  ms,  $d_1 = 10$  ms인 구간에서 99.92%를 보였으며, 최저 성능은

$d_2=250$  ms,  $d_1=50$  ms 구간에서 99.42%로서  $d_1$ ,  $d_2$  구간 선정에 따른 성능 편차가 매우 적음을 확인할 수 있다. 따라서 제안한 방법은 모든 구간에서 기존의 MMCD 보다 높은 성능을 나타낼 뿐만 아니라 실험적으로 결정한 구간에 따라 성능이 변하는 문제점도 동시에 극복하는 장점이 있는 것으로 평가된다.

## 5. 결론

본 논문에서는 기존의 MMCD 기반의 음성/음악 판별 성능을 개선하기 위하여 MMCD를 구하는 과정에서 사용되는 MFCC 계산에 있어서 시간에 따른 평균값을 사용하는 방법을 제안하였다. 기존 MMCD 파라미터는 최소 켈스트럼 거리를 구하는 구간에 따른 성능차이가 커서 최고 성능을 나타내는 최적 구간을 실험적으로 선정해야 하는 단점이 있었으나 제안된 방법에 의한 실험결과 기존 방법에 비해 성능이 개선되었을 뿐만 아니라 MMCD를 추출하는 최적 구간을 실험적으로 선정해야 하는 문제도 해결되는 결과를 얻었다.

앞으로 파라미터 추출 시간의 변화에 강인한 특징 파라미터의 개발을 통해 음성/음악 판별 시간을 단축하면서도 우수한 판별성능을 유지하는 방식에 대해 계속 연구할 계획이다. 또한 배경음악이 있는 경우 음성에 대한 판별 성능과 음악에서 노래에 해당하는 장르의 판별 성능을 제안한 방법을 통해 살펴봄으로써 음성과 음악이 겹쳐 있는 경우에 대한 문제점과 개선점에 대해 연구할 계획이다.

## 참 고 문 헌

- [1] E. Scheirer, M. Slaney, "Construction and evaluation of a robust multifeature music/speech discrimination", *Proc. ICASSP*, Vol. 2, pp. 1331-1334, 1997.
- [2] L. Lu, H. Jiang, H. J. Zhang, "A robust audio classification and segmentation method", *Proc. 9th ACM Multimedia*, pp. 203-211, 2001.
- [3] J. Saunders, "Real-time discrimination of broadcast speech/music", *Proc. ICASSP*, Vol. 2, pp. 993-996, 1996.
- [4] S. Takeuchi, T. Uchida, M. Yamashita, M. Sugiyama, "Optimization of voice/music detection in sound data", *Proc. Workshop on Consistent & Reliable Acoustic Cues for Sound Analysis*, <http://www.ee.columbia.edu/crac/>, 2001.
- [5] K. El-Maleh, M. Klein, G. Petrucci, P. Kabal, "Speech/music discrimination for multimedia application", *Proc. ICASSP*, Vol. 4, pp. 2445-2449, 2000.
- [6] 박슬한, 최무열, 김형순, "변형된 켈스트럼 거리를 이용한 음성/음악 판별", *말소리*, No. 56, pp. 195-206, 2005.
- [7] J. Ajmera, I. McCowan, H. Bourlard, "Speech/music discrimination using entropy and

dynamism features in a HMM classification framework”, *Speech Communication*, Vol. 40, No. 3, pp. 259-430, 2003.

- [8] 최무열, 송화전, 박슬한, 김형순, “다차원 MMCD를 이용한 음성/음악 판별”, *말소리*, No. 60, pp. 191-201, 2006.

접수일자 : 2007년 11월 23일

게재결정 : 2007년 12월 21일

▶ 최무열 (Mu Yeol Choi)

주소: 609-735 부산시 금정구 장전동 산30번지 부산대학교 공과대학 전자공학과

소속: 부산대학교 전자공학과 음성통신연구실

전화: 051) 516-4279

E-mail: mychois@pusan.ac.kr

▶ 김형순 (Hyung Soon Kim) : 교신저자

주소: 609-735 부산시 금정구 장전동 산30번지 부산대학교 공과대학 전자공학과

소속: 부산대학교 전자공학과 음성통신연구실

전화: 051) 510-2452

E-mail: kimhs@pusan.ac.kr