음악과 음성 판별을 위한 웨이브렛 영역에서의 특징 파라미터

김정민(삼성전자), 배건성(경북대)

<차 례>

1. 서론
2. 음악과 음성 판별을 위한 웨이브렛 영역에서의 특징파라미터 추출
3. 실험 및 고찰
4. 결론

<Abstract>

Feature Parameter Extraction and Analysis in the Wavelet Domain for Discrimination of Music and Speech

Jung-Min Kim, Keun-Sung Bae

Discrimination of music and speech from the multimedia signal is an important task in audio coding and broadcast monitoring systems. This paper deals with the problem of feature parameter extraction for discrimination of music and speech. The wavelet transform is a multi-resolution analysis method that is useful for analysis of temporal and spectral properties of non-stationary signals such as speech and audio signals. We propose new feature parameters extracted from the wavelet transformed signal for discrimination of music and speech. First, wavelet coefficients are obtained on the frame-by-frame basis. The analysis frame size is set to 20 ms. A parameter E_{sum} is then defined by adding the difference of magnitude between adjacent wavelet coefficients in each scale. The maximum and minimum values of E_{sum} for period of 2 seconds, which corresponds to the discrimination duration, are used as feature parameters for discrimination of music and speech. To evaluate the performance of the proposed feature parameters for music and speech discrimination, the accuracy of music and speech discrimination is measured for various types of music and speech signals. In the experiment every 2-second data is discriminated as music or speech, and about 93% of music and speech segments have been successfully detected.

* Keywords: Discrimination of music and speech, Wavelet transform

1. 서 론

음향신호를 식별하는 일은 소나, 레이더, 음성/오디오 신호처리 등에서 중요한 연구 분야의 하나로 활발한 연구 활동이 이루어지고 있다. 음악과 음성은 우리가 가장 쉽게 접할 수 있는 음향신호로서 음악과 음성의 판별은 방송 모니터링 시스템, 음성/오디오 코딩 등의 분야에서 응용될 수 있다. 음악과 음성이 섞여있는 멀티미디어 신호의 효율적인 검색 기능이나 자동 판별 기능을 요구하는 방송 모니터링 시스템에서 음악과 음성의 판별이 필요하며, 특히 고음질 저 전송률의 음향부호화 알고리듬 연구에 있어서 음성부호화기와 오디오부호화기를 스위칭 하는 방식으로 음향부호화기를 구현하고자 할 경우, 음악과 음성의 판별은 음향부호화기의 음질 성능에 매우 중요한 요소이므로 정확한 음악과 음성의 판별이 요구된다. 음악과 음성을 판별하기 위한 기존의 특징파라미터로는 스펙트럴 센트로이드(spectral centroid), 스펙트럴 플럭스(spectral flux), 스펙트럴 롤-오프 포인트(spectral roll-off point), 영교차율(zero crossing rate), 프레임 에너지(frame energy), 피치 강도(pitch strength) 등이 있다[1-5]. 이러한 특징파라미터들은 음악과 음성신호의 일반적인 특성을 어느 정도 반영하지만 다양한 음악과 음성을 판별하는데 있어서는 뛰어난 성능을 나타내지 못하고 있다[6].

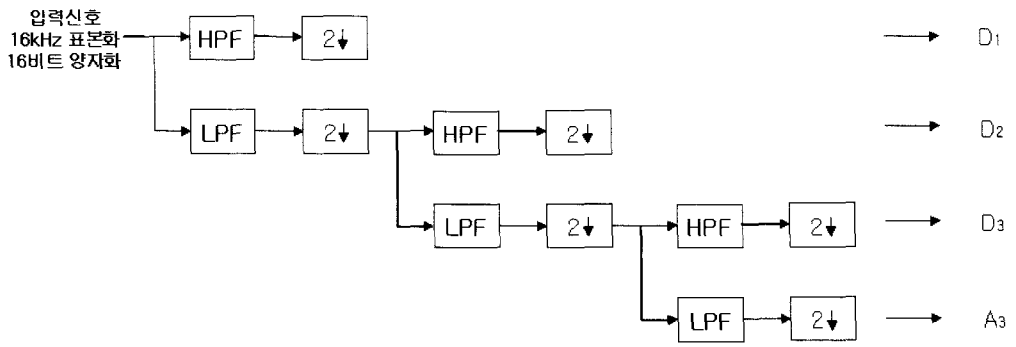
본 연구에서는 웨이브렛 변환을 이용하여 음악과 음성의 판별성능을 향상시킬 수 있는 특징파라미터의 추출에 관한 연구를 수행하였다. 일반적으로 웨이브렛 변환은 신호를 주파수 대역에 따라 다른 창 함수로 분해하여 고주파 대역에서는 높은 시간 해상도를 갖고 저주파 대역에서는 높은 주파수 해상도를 갖게 한다. 음악신호와 같이 저역과 고역의 주파수 성분을 함께 가지고 있으며, 특히 신호가 짧은 지속시간을 가지는 고주파 성분이나 긴 저주파 성분 또는 이의 합성으로 구성되어 있는 경우 다중 해상도를 갖는 웨이브렛 변환이 고정된 시간-주파수 해상도를 갖는 단 구간 푸리에 변환에 비해서 더욱 유효하다. 따라서 웨이브렛 영역에서 다양한 종류의 음악신호와 음성신호의 특성을 분석한 후 음악과 음성을 판별하기 위한 특징파라미터를 제안하였으며, 추출된 특징파라미터를 이용하여 음악/음성 판별 실험을 수행하였다. 한 프레임(16kHz 표본화 된 신호의 256 샘플) 분석구간에 대하여 다우비치(Daubechies) 4차 필터를 이용하여 구한 각 스케일 별 웨이브렛 계수로부터 파라미터 E_{sum}을 정의하고 2초 판별구간(125 프레임) 단위로 E_{sum}의 최소값과 최대값을 갖는 E_{min}과 E_{max}을 음악과 음성 판별을 위한 특징파라미터로 사용하였다. 제안한 특징파라미터는 휴지구간 또는 묵음구간의 음성신호의 특성을 잘 반영하며, 기존에 제안된 특징파라미터들에 비해 향상된 판별 정확도를 보였다.

본 논문의 구성은 다음과 같다. 2장에서 웨이브렛 변환을 이용하여 음악과 음성신호로부터 특징파라미터 E_{min}과 E_{max}을 추출하는 과정과 추출된 특징파라미터의 특성을 분석한다. 그리고 3장에서는 본 연구에서 제안한 특징파라미터를 이용하여

음악과 음성신호에 대한 판별실험을 수행한 결과를 제시한다. 마지막으로 4장에서 결론을 맺는다.

2. 음악과 음성 판별을 위한 웨이브렛 영역에서의 특징파라미터 추출

음악과 음성을 구분하기 위한 특징파라미터에 대한 다양한 연구가 이루어졌으나 기존에 제안된 다양한 특징 파라미터들은 1초 구간을 판별구간으로 했을 때 대략 80% 내외의 판별 정확도를 보이고 있으며, 각각의 특징파라미터는 판별성능에 있어서 상당한 차이를 보이고 있다[6]. 본 논문에서는 우수한 판별성능을 갖는 특징파라미터를 추출하기 위해서 다중해상도를 갖는 웨이브렛 변환[7]을 이용하였다. 각 스케일별로 웨이브렛 변환된 신호를 얻기 위해서 일반적으로 <그림 1>과 같이 트리 형태의 필터뱅크를 이용한다. <그림 1>은 웨이브렛 변환과정을 나타내는데 16kHz 표본화, 16비트 양자화 된 입력신호를 한 프레임(256샘플) 단위로 다우비치 4차 필터, 분석 레벨 3의 이산 웨이브렛 변환으로부터 4개의 웨이브렛 계수 D1, D2, D3, A3를 구할 수 있다. 이산 웨이브렛 변환을 통해서 가장 시간 해상도가 좋은 순서대로 웨이브렛 계수가 얻어지는데, 웨이브렛 변환과정을 반복 할 때마다 다운샘플링(downsampling)을 통해 웨이브렛 계수는 입력신호의 샘플 길이의 반이 된다. 또한 웨이브렛 변환을 반복할 때 마다 시간 해상도는 감소하고 주파수 해상도는 증가하게 된다. <그림 1>에서 구한 웨이브렛 계수들 중에서 D1 은 시간 해상도가 가장 좋으며 A3는 주파수 해상도가 가장 좋다. 이들 웨이브렛 계수들 중에서 시간 해상도가 좋은 순서로 웨이브렛 계수 D1, D2, D3을 사용하여 음악과 음성을 판별하기위한 특징파라미터를 추출하기 위해서 웨이브렛 각 스케일에서 계수 간 차이의 합으로써 E1, E2, E3를 정의하였다.



<그림 1> 웨이브렛 변환과정

각 스케일에서의 웨이브렛 계수는 각 대역통과필터의 출력신호에 대응되므로 계수간의 차이를 더해 줌으로써 시간영역에서의 신호특성 변화의 정도를 나타내는 측도가 된다. 다음의 식 (1)에서 식 (3)은 웨이브렛 영역의 각 스케일에서 계수 간 차이의 합으로서 E_1, E_2, E_3을 구하는 과정을 나타낸다. 웨이브렛 각 스케일에서 구한 E_1, E_2, E_3을 식 (4)와 같이 모두 더하여 한 프레임 분석구간에서 E_{sum}이라는 값으로 특징파라미터를 정의하였다.

$$E_1 = \sum_{n=1}^{(256/2)-1} |D_1(n) - D_1(n+1)| \quad (1)$$

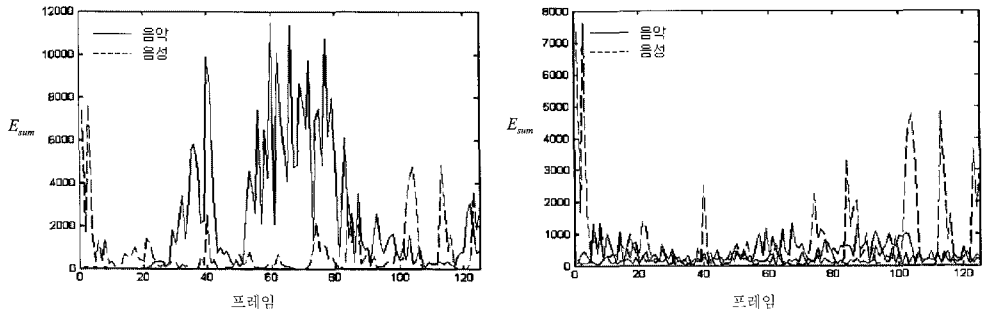
$$E_2 = \sum_{n=1}^{(256/4)-1} |D_2(n) - D_2(n+1)| \quad (2)$$

$$E_3 = \sum_{n=1}^{(256/8)-1} |D_3(n) - D_3(n+1)| \quad (3)$$

$$E_{sum} = E_1 + E_2 + E_3 \quad (4)$$

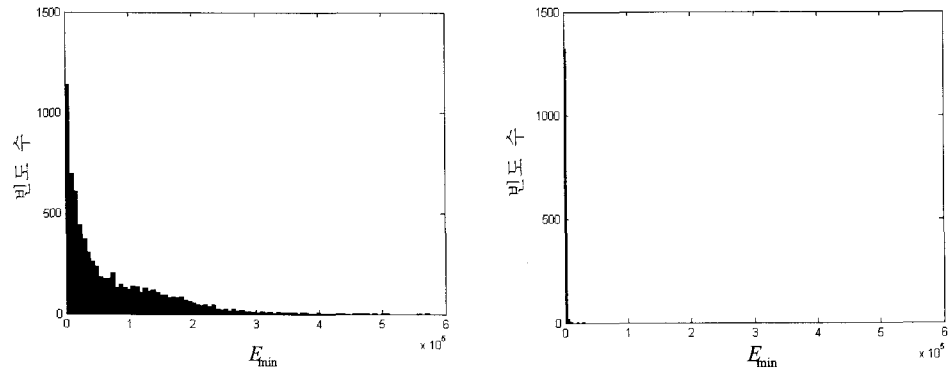
음성은 일반적으로 단일 음원에 의한 신호 즉, 한 사람의 발성에 의한 신호이므로 특히 휴지구간이나 묵음구간에 있어서는 웨이브렛 각 스케일에서 구한 E_1, E_2, E_3가 모두 유성음 발성시기의 분석구간과 비교해서 유사한 형태의 낮은 값을 나타낸다. 그러나 음악은 다양한 악기 연주로 이루어진 복합신호이기 때문에 휴지구간이나 묵음구간 발생이 드물다. 따라서 음악신호에 대해서 웨이브렛 각 스케일에서 구한 E_1, E_2, E_3은 다른 시점의 분석구간과 비교해서 유사한 형태를 나타내는 경우가 그리 많지 않다. 그러므로 음악과 음성을 판별하는 데 있어서 웨이브렛 각 스케일에서 구한 E 값을 모두 합한 E_{sum}을 사용하는 것은 좀더 안정적인 결과를 나타낼 수 있다. 음악과 음성신호에 대하여 한 프레임 분석구간에서 구한 E_{sum}을 2초(125 프레임) 판별구간에 대해서 나타내면 <그림 2>에서와 같이 크게 두 가지 형태를 보인다. 일반적으로 팝(pop), 락(rock), 힙합(hip-hop) 등의 강렬한 사운드의 음악은 <그림 2>의 (a)에서처럼 E_{sum}이 음성보다 더 큰 값을 나타내며 시간적인 변화 또한 음성처럼 상당히 크게 나타난다. 그리고 재즈(jazz), 클래식(classic) 등과 같이 전체적으로 잔잔한 사운드의 음악의 경우에 있어서는 <그림 2>의 (a)에서와 같이 E_{sum}의 변화량이 크게 나타나는 경우가 많지만, <그림 2>의 (b)에서 처럼 E_{sum}의 변화가 매우 작게 나타나는 경우도 상당함을 볼 수 있다. 따라서 판별구간에서의 E_{sum}의 변화율을 통해서 음악과 음성을 판별하는 것은 어렵고 절대적인 E_{sum} 값이 음악과 음성의 판별에 좀 더 효과적으로 적용될 수 있음을 알 수 있다. 이는

음성의 휴지구간 또는 묵음구간에서 추출된 E_{sum}이 일반적으로 음악보다 매우 작은 값을 나타내며 가령 음성의 휴지구간 또는 묵음구간에서와 비슷한 E_{sum} 값을 가지는 음악이라 할지라도 음성보다 E_{sum}의 변화가 적기 때문에 음악과 음성의 판별이 가능하다. 본 논문에서는 분석구간에서 구한 웨이브렛 영역에서의 E_{sum} 값을 음악과 음성을 판별 하는데 사용하기 위해서 판별구간에서의 E_{sum}의 최소값 E_{min}과 최대값 E_{max}를 음악과 음성 판별을 위한 특징파라미터로 추출하고 분석하였다. 그리고 음악과 음성의 판별 알고리듬으로 추출된 특징파라미터 E_{min}과 E_{max}의 임계값을 사용하기 위해서 다음에 설명하는 분석과정을 통해서 적절한 임계값을 설정하였다.



(a) 팝 음악과 음성의 E_{sum} (b) 클래식 음악과 음성의 E_{sum}

<그림 2> 판별구간에서의 음악과 음성신호의 E_{sum}

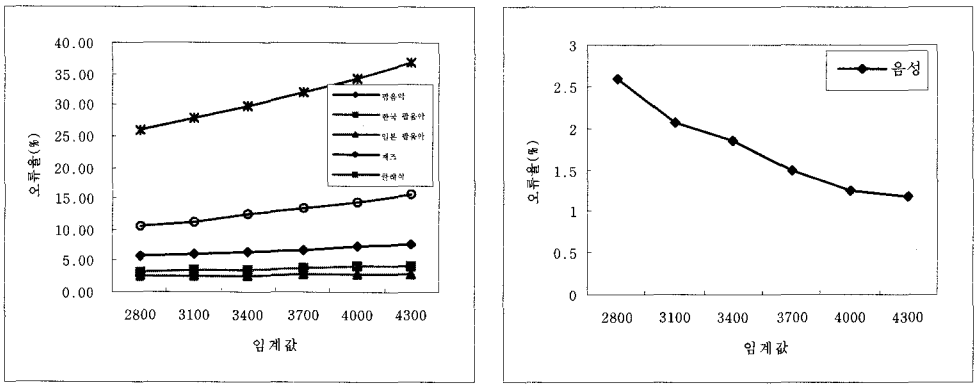


(a) 음악의 E_{min} 분포 (b) 음성의 E_{min} 분포

<그림 3> 판별구간에서의 음악과 음성신호의 E_{min} 분포

다양한 장르의 음악과 음성신호에 대하여 판별구간에서의 E_{min} 값의 분포를 히스토그램으로 나타낸 것은 <그림 3>과 같다. <그림 3>의 (a)는 판별구간에서의 음

악의 E_{min} 값의 분포를 나타내며 <그림 3>의 (b)는 음성의 E_{min} 값의 분포를 나타낸다. 음성의 E_{min} 값은 매우 작은 값을 가지는 반면 음악의 E_{min} 값은 아주 작은 값에서부터 매우 큰 값까지 전체적으로 넓게 분포한다. 따라서 적절한 E_{min} 값에 대한 기준을 설정한다면 대부분의 음악과 음성을 구분할 수 있다. 음악과 음성을 구분하기 위해서 몇 개의 E_{min}의 임계값을 적용하여 다양한 음악장르와 음성신호에 대하여 판별을 수행하였으며 그 결과는 <그림 4>에 나타낸 것과 같다. E_{min}이 2800일 때 음악의 판별오류가 가장 낮으며 음성은 판별오류의 변화가 매우 적어서 음악과 음성을 판별하기 위한 E_{min}의 임계값을 2800으로 설정하였다. 일반적으로 E_{min}의 임계값에 의해서 대부분의 음악은 제대로 음악으로 구별할 수 있지만 클래식, 재즈 등의 음악은 음성의 E_{min}과 비슷한 값을 나타내는 판별구간이 존재하기 때문에 상당수의 판별구간이 음성으로 판별된다. 따라서 E_{min}에 의해서 음성으로 구분된 실제 음악신호에 대해서는 다시 한번 음악과 음성신호를 구분하는 과정이 필요한데, 음성으로 구분된 음악신호들 대부분이 판별구간에서 E_{sum} 값의 변화가 매우 작다는 사실로부터 판별구간에서의 E_{max} 값을 사용하여 음악과 음성을 판별할 수 있다.

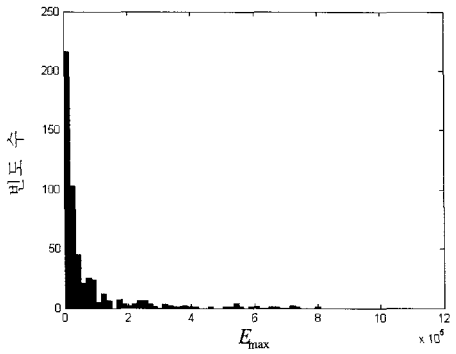


(a) 음악의 판별결과 (b) 음성의 판별결과

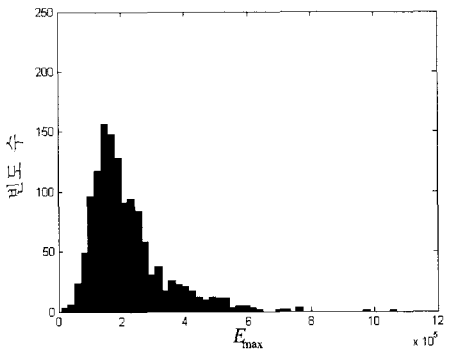
<그림 4> E_{min}의 임계값에 따른 음악과 음성의 판별결과

<그림 5>는 판별구간에서의 음악과 음성신호의 E_{max} 값의 분포 특성을 나타낸 것이다. 이 경우 음성의 E_{max} 값이 음악보다 더 높은 범위에 존재하는 것을 볼 수 있다. 따라서 E_{min}의 임계값에 의해서 음성으로 판별된 실제 음악신호를 음성신호와 구별하기 위해서는 음악과 음성신호의 E_{max} 값을 사용하면 음악과 음성의 구분이 용이해진다. <그림 6>은 E_{max}의 임계값에 따른 판별결과를 보인 것이다. 음악과 음성을 구별하기 위한 E_{max}의 임계값을 설정하기 위해서 50000에서 75000까지

5000단위로 증가시켰을 때 음악은 임계값이 증가할수록 판별오류가 감소하고 음성은 판별오류가 증가하였다. 실험결과로부터 음악과 음성신호 모두에서의 판별성능을 고려하여 E_{max}의 임계값으로 65000이라는 값을 설정하였다.

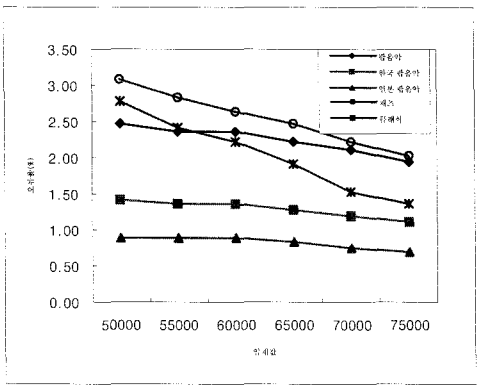


(a) 음악의 E_{max} 분포

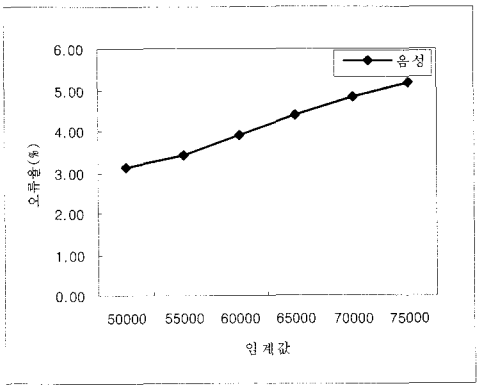


(b) 음성의 E_{max} 분포

<그림 5> 판별구간에서 음악과 음성신호의 E_{max} 분포



(a) 음악의 판별결과



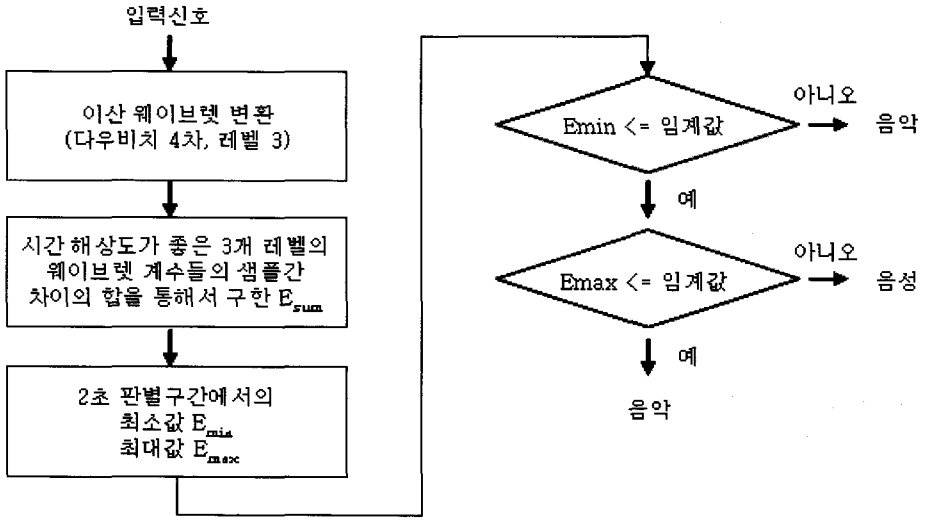
(b) 음성의 판별결과

<그림 6> E_{max}의 임계값에 따른 음악과 음성의 판별결과

3. 실험 및 고찰

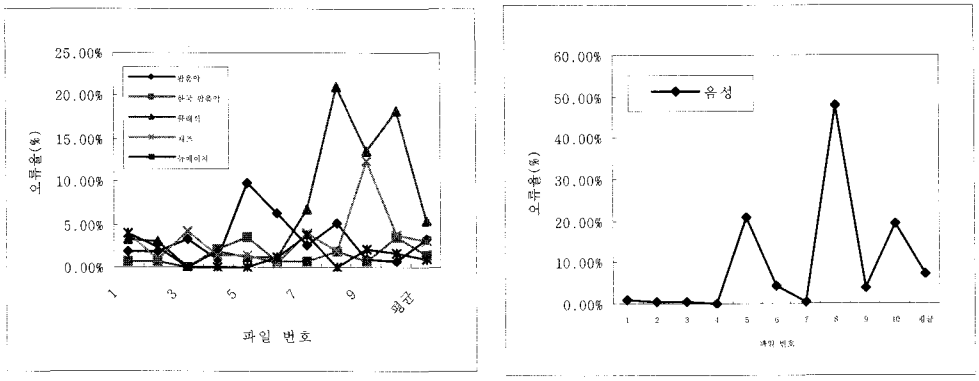
음악과 음성 판별을 위해서 프레임 단위의 분석구간에서 구한 E_{sum}으로부터 2초 판별구간 단위로 다양한 음악과 음성에 대한 예비실험에서 E_{min}과 E_{max}값을 추출하여 그 분포 및 임계값에 따른 판별오류를 분석하였다. 그리고 그 결과를 이용하여 음악과 음성을 판별할 수 있는 각각의 임계값을 구하였다. 음악과 음성을 구

별하기 위한 E_{min}의 임계값은 2800, E_{max}의 임계값은 65000으로 설정하였으며, 예비 실험에서 분석에 사용되지 않은 음악과 음성 신호에 대하여 판별실험을 수행하였다. <그림 7>은 본 논문에서 제안된 음악과 음성 판별 과정을 나타낸 것이다. 우선 신호가 입력되면 16kHz 표본화, 16 비트로 양자화한 후, 한 프레임(256 샘플) 단위로 다우비치 계수 4차, 레벨 3으로 이산 웨이브렛 변환을 수행하여 웨이브렛 계수 D1, D2, D3, A3를 구한다. 그 다음 세 개 레벨의 웨이브렛 계수 D1, D2, D3에 대하여 분석구간에서의 E_{sum}을 계산하고 2초 판별구간에 대해서 E_{min}과 E_{max} 값을 추출한다. 판별부에서는 우선 추출된 E_{min}을 설정된 임계값과 비교하여 일차적으로 음악과 음성을 구분한다. 1차 판별과정에서 음악으로 구분된 신호는 최종적으로 음악으로 판별된 것이며 음성으로 구분된 신호에 대해서는 추출된 E_{max}를 설정한 임계값과 비교하여 최종적으로 음악과 음성을 판별한다. 일반적으로 강렬한 사운드의 팝 계열의 음악과 주로 잔잔한 선율로 구성된 재즈 또는 클래식 계열의 음악은 판별구간에서의 E_{sum}의 시간적인 변화형태가 상당히 많은 차이를 나타낸다. 따라서 다양한 장르의 음악과 음성신호를 대상으로 음악과 음성 판별실험이 요구되는데, 본 실험에 사용한 음악 장르로는 팝, 한국 팝, 클래식, 재즈, 뉴에이지 등이 있다. 실험에 사용한 음악의 수는 장르별로 10곡으로 하여 모두 50곡의 음악을 사용하였으며 음성은 일본어 또는 영어로 발음되는 방송뉴스 상황의 음성과 어학 강좌시의 음성으로 이루어진 10개의 파일을 판별실험에 사용하였다. 음악은 모든 장르를 합쳐서 7956개의 판별구간에 해당하는 4시간 25분 12초 분량을 판별실험에 사용하였으며, 음성은 1199개의 판별구간에 해당하는 39분 58초 분량을 실험에 사용하였다. 실험에 사용된 전체 신호의 길이는 음악과 음성신호를 모두 합쳐서 9155개의 판별구간에 해당되는 5시간 5분 10초 분량이다.



<그림 7> 음악과 음성 판별 과정의 순서도

본 논문에서 제안한 웨이브렛 영역에서의 특징파라미터 E_{min}과 E_{max}를 이용하여 다양한 장르의 음악과 음성신호에 대하여 판별실험을 한 결과를 <그림 8>에 나타내었다. 전체적인 판별결과는 모든 장르의 음악에 대해서 약 95% 이상의 판별 정확도를 나타내며 음성신호에 대해서는 약 93% 정도의 판별 정확도를 보인다. 음악의 경우는 음성과 비슷한 시간적인 변화를 나타내는 단일악기 연주가 포함된 구간에서 오류가 많이 발생하며, 음성은 배경 잡음이 강하거나 잔향이 많은 구간에서 대부분의 오류가 발생하는 것을 확인 할 수 있었다. 따라서 판별오류가 많이 발생하면서 음성신호와 비슷한 시간적인 변화를 보이는 단일악기 연주음악에 대해서는 판별오류를 줄이기 위한 보조적인 요소가 필요하다.

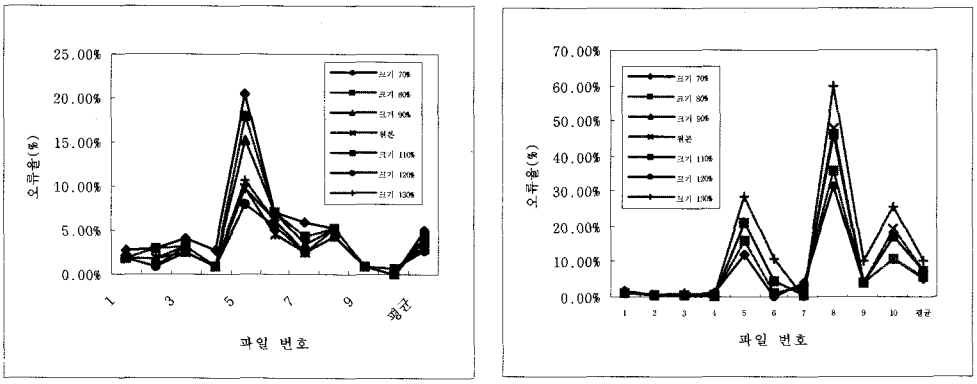


(a) 음악의 판별실험 결과 (b) 음성의 판별실험 결과

<그림 8> 음악과 음성 판별실험 결과

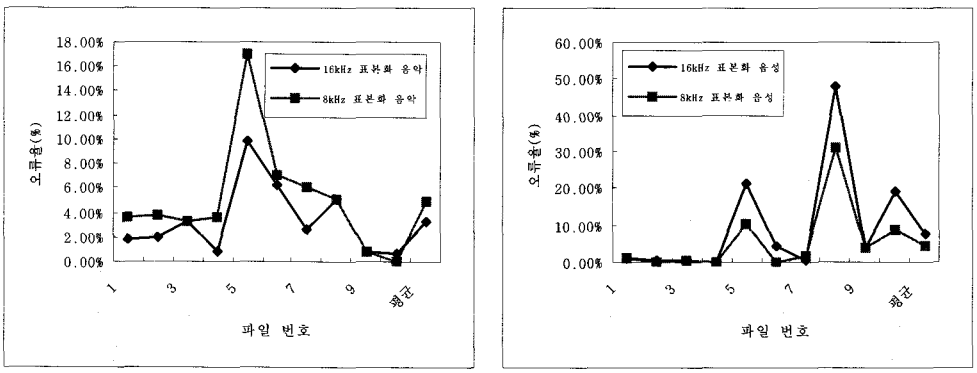
본 논문에서 제안한 음악과 음성 판별을 위한 특징파라미터 E_{min}과 E_{max}는 절대적인 값을 기준으로 음악과 음성의 판별에 사용되기 때문에 입력신호의 크기에 영향을 받을 수 있으므로 판별실험에 사용된 음악과 음성신호의 크기를 원래 크기의 70%에서 130%까지 변화시키면서 제안한 특징파라미터를 이용한 음악과 음성의 판별 성능을 살펴보았다. <그림 9>는 음악과 음성신호의 크기를 원래 크기에서 70%에서 130%까지 변화시켰을 경우 크기 변화에 따른 팝 음악과 음성의 판별 오류의 변화를 나타낸다. 팝 음악의 경우 신호의 크기가 감소하면 판별오류가 증가하고 신호의 크기를 증가하면 판별오류가 감소하는 경향을 보인다. 음성의 경우는 팝 음악의 경우와는 달리 신호의 크기가 감소하면 판별오류가 감소하고 신호의 크기를 증가하면 판별오류가 증가하는데 전체적으로는 판별성능에 큰 변화가 없지만 원래 신호크기에서 오류가 많이 발생했던 음악과 음성신호에 대해서는 판별성능의 변화가 크게 나타난다. 또한 제안한 특징파라미터가 좀 더 좁은 대역에

서도 임계값의 재설정 과정 없이 적용가능한지를 살펴보기 위해서 신호의 표본화 율을 16kHz에서 8kHz로 바꾼 다음 판별실험을 한 결과를 <그림 10>에 나타내었다. 표본화 율을 8kHz로 바꾸면 전체적인 판별성능이 음악은 조금 떨어지고 음성은 약간 향상된다. 또한 원래 판별오류가 많이 발생했던 단일악기 연주음악과 배경잡음이 강한 음성은 표본화 율 변화에 따른 판별성능 변화가 매우 크게 나타났다.



(a) 음악의 판별실험 결과 (b) 음성의 판별실험 결과

<그림 9> 크기 변화에 따른 음악과 음성의 판별실험 결과



(a) 음악의 판별오류 (b) 음성의 판별오류

<그림 10> 표본화 율 변화에 따른 음악과 음성의 판별오류

4. 결 론

본 논문에서는 음악과 음성을 구분하기 위해 기존에 사용되는 여러 특징파라미터의 조합 대신에, 웨이브렛 영역에서 2초 길이를 갖는 판별구간 단위로 새로운 특징파라미터를 제안하고 다양한 장르의 음악과 음성 신호에 대해 판별실험을 수행하였다. 다우비치 4차, 레벨 3의 이산 웨이브렛 변환을 이용하였고, 신호가 입력되면 프레임 단위로 웨이브렛 변환을 수행하여 웨이브렛 계수 D1, D2, D3, A3를 구한 다음 웨이브렛 계수 D1, D2, D3의 샘플 간 차이를 구하여 모두 더한 값인 E_{sum}을 특징벡터로 정의하였다. 일반적으로 음성이 휴지구간 또는 묵음구간에서 E_{sum}이 매우 작은 값을 가지므로 2초 판별구간에서 가장 작은 E_{sum}을 추출하여 E_{min}으로 정의하고 음악과 음성을 판별하기 위한 특징파라미터로 사용하였다. E_{min}의 임계값에 의해서 음성으로 구분되는 신호에는 실제로 음악인 신호가 상당 부분 포함되며 이러한 신호에 대해서 다시 음악과 음성을 판별하기 위해서 판별구간에서 가장 큰 E_{sum}을 추출하여 E_{max}으로 정의하고 특징파라미터로 사용하였다. 또한, 입력신호의 크기 변화 및 표본화 율 변화에 따른 판별 성능을 실험하였다.

제안한 특징파라미터 E_{min}과 E_{max}를 사용하여 다양한 장르의 음악과 음성신호에 대하여 판별실험을 수행한 결과, 단일 악기 연주음악에서 판별 오류가 다소 높게 나타났으며 음성은 배경 잡음 또는 잔향이 큰 음성에서 다소 높은 판별오류가 나타났다. 다양한 음악과 음성신호에 대한 판별실험에서 본 연구에서 제안된 특징파라미터는 93% 이상의 판별율을 보였다. 앞으로 제안한 특징파라미터를 이용하여 입력신호의 크기 변화 및 표본화 율 변화에 보다 강인한 알고리즘에 대한 연구가 필요하다.

참 고 문 헌

[1] M. J. Carey, E. S. Parris, H. Lloyd-Thomas, "A comparison of features for speech, music discrimination", *Proc. ICASSP*, Vol. 1, pp. 149-152, Mar. 1999.

[2] J. Saunders, "Real-time discrimination of broadcast speech/music", *Proc. ICASSP*, Vol. 2, pp. 993-996, May 1996.

[3] Y. Nakajima, Yang Lu, M. Sugano, A. Yoneyama, H. Yamagihara, A. Kurematsu, "A fast audio classification from MPEG coded data", *Proc. ICASSP*, Vol. 6, pp. 3005-3008, Mar. 1999.

[4] E. D. Scheirer, "Tempo and beat analysis of acoustic musical signals", *Journal of the Acoustical Society of America*, Vol. 103, No. 1, pp. 588-601, Jan. 1998.

[5] T. Zhang, J. Kuo, "Hierarchical classification of audio data for archiving and retrieving", *Proc. ICASSP*, Vol. 6, pp. 3001-3004, Mar. 1999.

[6] E. Sheirer, M. Slaney, "Construction and evaluation of a robust multifeature speech/music

discriminator", *Proc. ICASSP*, Vol. 2, pp. 1331-1334, Apr. 1997.
[7] I. Daubechies, "The wavelet transform time frequency localization and signal analysis", *IEEE Transactions on Information Theory*, Vol. 36, No. 5, pp. 961-1005, Sept. 1990.

접수일자: 2007년 2월 5일
게재결정: 2007년 3월 17일

▲ 김정민(Jung-Min Kim)
주소: 443-742 경기도 수원시 영통구 매탄 3동 삼성전자
소속: 삼성전자 디지털 멀티미디어 연구소 오디오연구실
전화: 031) 200-3683
E-mail: jungmin75.kim@samsung.com

▲ 배건성(Keun-Sung Bae) : 교신저자
주소: 702-701 대구광역시 북구 산격 3동 1370번지 경북대학교
소속: 경북대학교 전자전기컴퓨터학부 신호처리 연구실
전화: 053) 950-5527
E-mail: ksbae@ee.knu.ac.kr