한국음향학회지 제21권 제2호 pp. 182~189 (2002)

오디오 데이터의 특징 파라메터 구성에 따른 내용기반 분석

The Content Based Analysis According to the Composition of the Feature Parameters for the Auditory Data

한 학 용*, 김 수 훈**, 허 강 인*
(Hag Yong Han*, Soo Hoon Kim**, Kang In Hur*)

*동아대학교 전자공학과, **부천대학 정보통신계열

(접수일자: 2001년 8월 30일; 수정일자: 2001년 12월 6일; 채택일자: 2001년 12월 27일)

본 논문은 오디오 색인·검색 시스템을 구현하기 위하여 오디오 신호에 대한 특징 파라메터 풀 (pool)을 구성하고 이에 따른 오디오 데이터의 내용분석 및 분류에 관한 연구이다. 오디오 데이터는 기본적인 다양한 오디오 형태로 분류되어진다. 본 논문에서는 오디오 데이터의 분류에 이용 가능한 특징 파라메터를 분석하고 추출방법에 대하여 논한다. 그리고 특징 파라메터 풀을 색인 그룹 단위로 구성하여 오디오 카테고리에 대한 설정된 특징들의 포함 정도와 색인기준을 오디오 데이터의 내용을 중심으로 비교·분석한다. 그리고 위의 결과를 바탕으로 분류절차를 구성하여 오디오 신호를 분류하는 모의실험을 행하였다.

핵심용어: 오디오 데이터, 색인검색
투고분야: 음성처리 분야 (2,4)

In this paper, we research the content-based analysis and classification according to the composition of the feature parameters pool for the auditory signals to implement the auditory indexing and searching system. Auditory data is classified to the primitive various auditory types. we described the analysis and feature extraction method for the feature parameters available to the auditory data classification. And we compose the feature parameters pool in the indexing group unit, then compare and analysis the auditory data centering around the including level and indexing criterion into the audio categories. Based on this result, we composed the classification procedure and simulate the auditory data classification.

Keywords: *Audio, Auditory data, Indexing searching*
ASK subject classification: *Speech signal processing (2,4)*

I. 개 요

오디오는 미디어의 중요한 형태로 멀티미디어 데이터의 핵심 요소 중의 하나이다. 그럼에도 불구하고 오디오 신호가 가지는 동적인 특성과 다양성으로 인해 멀티미디어 스트림의 오디오 색인·검색에 관한 연구는 내용기반 이미지나 비디오 데이터베이스에 대해 행해지는 연구에 비교하여 매우 미흡한 실정이다. 그러나, 최근 들어 디지털 오디오 데이터 베이스가 다양해지고 오디오의 내용 분석에 의한 오디오 데이터 베이스의 효과적인 관리에 대한 중요성이 인식되기 시작하고 있다. 실제 오디오 데이터의 색인·검색은 전문 미디어 제품에서 뿐만 아니라 많은 응용영역을 가지고 있으며, 오디오 내용 검색은 비디오 검색에도 중요한 역할을 할 것이다. 현재 비디오 분할과 색인에 대한 접근은 영상정보에 대부분 초점

책임저자: 한학용 (hyhan@electro.donga.ac.kr)
604-714 부산광역시 사하구 하단동
동아대학교 전자공학과 패턴연구실
(전화: 051-200-6773; 팩스: 051-200-7712)

이 맞추어져 있다. 멀티미디어 데이터에 대한 영상에 기반한 처리는 데이터의 의미 측면에서 분할이 용이한 이점을 가지고 있다. 그러나 다양한 멀티미디어 요소들의 통합적인 색인·검색에 대한 연구가 비디오를 완전하게 파싱(parsing)하기 위해서 해결해야 될 필수적인 과제이다[1].

내용기반 오디오 데이터에 대한 색인·검색 연구에는 아직까지 많은 제한이 있다. 본 논문에서는 이미 연구되어 왔던 음성과 음악과 같은 오디오 형태의 기본적인 분류 카테고리를 확장한다[1-3]. 이전의 연구에서는 소홀히 다루어져 왔던 노래를 포함시키고 배경 오디오의 구별을 강조하여 음악과 음성만으로 구성된 카테고리뿐만 아니라 배경음이 음악인 음성, 배경음이 음악인 노래 그리고 배경음이 음악인 사운드 등과 같이 다양한 사운드를 분류 대상으로 설정한다. 설정된 대상에 따라 최적의 색인·검색을 위하여 일차적으로 오디오 신호 분석을 통하여 각 카테고리의 특징을 비교·분석한 후, 이를 바탕으로 몇 가지 특징을 제안하고 이를 기반으로 특징 파라메터 풀을 구성하여 오디오 데이터에 대한 분류실험을 행하였다.

본 논문은 다음과 같이 구성되어진다. 2장에서는 오디오 특징 파라메터 풀에 포함 가능한 여러가지 파라메터들을 소개하고 그 특징의 계산과 추출방법을 소개한다. 3장에서는 몇가지 오디오 데이터의 카테고리를 확장하여 설정하고 각각에 대하여 포함된 특징들을 비교·분석한다. 4장에서는 분석환경과 언급한 특징들 중에서 유효한 특징들로 이루어진 특징 파라메터 풀을 구성하여 정량적인 임계값에 의한 분류절차 구성에 대하여 설명하고 5장에서 모의실험 결과, 그리고 6장에서 결론을 맺는다.

II. 오디오 신호의 특징 파라메터들

본 논문에서의 특징 파라메터는 오디오 신호의 속성에 기반하여 다양한 특징 파라메터들을 유기적으로 조합하여 파라메터 풀을 구성하여 이용한다. 본 장에서는 신호처리에서 일반적으로 사용되는 특징 파라메터뿐만 아니라 "배음도", "FuF 지속도", "Freq. 집중도"라는 특징을 제안하고 오디오 데이터 분류 파라메터로 적용가능한가를 분석한다.

2.1. 단구간 평균 에너지

오디오 신호의 에너지는 일반적으로 식 (1)과 같이 자승평균대수 에너지로 정의하는데, 신호 파형의 변화정도를 크게 하여 임계값에 여유를 줄 경우에는 식 (2)를 사용한다[1-3,6].

$$E_n = \frac{1}{N}\sum_{m=0}^{N-1} x^2(m) \tag{1}$$

$$E_n(dB) = 10\log E_n \tag{2}$$

2.2. 단구간 평균 영교차율

영교차는 이산신호의 인접 샘플이 다른 부호를 가질 경우에 나타나며 영교차들이 발생하는 비율은 신호의 주파수 내용을 측정하는 가장 간단한 특징이다. 단구간 평균 영교차율 (이후 "ZCR")은 식 (3)과 같이 정의한다[1-3,6].

$$L_n = \frac{1}{2}\sum_{m=0}^{N-1}|\mathrm{sgn}[x(n-m)] - \mathrm{sgn}[x(n-m-1)]|$$

$$\text{where,}\quad \mathrm{sgn}[s(n)] = 1, s(n) \geq 0 \tag{3}$$

오디오 신호는 다양한 소스 (Source)로 구성되므로 ZCR 커버들은 매우 다른 성질들을 가지고 있다. ZCR 커버 특성의 규칙성, 주기성, 안정성 그리고 진폭의 범위와 분산값이 특징 파라메터로 이용가능하다.

2.3. 단구간 주파수

사운드는 기본 주파수와 그것의 정수배의 주파수인 배음을 가진 사운드 (그림 1)와 그렇지 않는 사운드 (그림 2)로 나눌 수 있다. 음성의 경우, 구성요소들이 고조파적인 유성음과 비고조파적인 무성음의 혼합된 형태의 사운드이다. 반면에, 악기에 의한 음악적 요소를 가진 사운드는 대부분 배음을 가진 사운드에 해당한다.

오디오 데이터에 대한 고조파적인 배음의 포함 유무 검출은 AR모델로 신호의 스펙트럼을 추정함으로 가능하다. AR 모델을 통하여 생성된 계수로부터 추정된 스펙트럼은 주파수 스펙트럼의 포락선에 해당하고 두드러진 피크부분을 가지게 된다. AR모델로 스펙트럼을 추정하는 방법에는 Durbin, Berg 그리고 Yule-Waker가 제안한 알고리즘을 이용한다. 그림 1, 2, 3은 Durbin법으로 추정한 주파수 스펙트럼을 보여준다. 단구간에서의 고조파적인 특성을 찾기 위해서는 AR모델로 스펙트럼의 포락선을 추정하여 피크부분을 검출하는 방법이 직접 신호의 스펙트럼을 계산하는 경우보다 피크부분의 추출이 용이하다. 이 때 추정된 기본주파수의 정밀도를 충분히 높이기 위해서 AR 모델의 계수가 어느 정도 높아야 하지만 고조파의 포함 유무만을 알기 위해서는 비교적 피크의 검출이 용이한 낮은 차수로도 충분하다.(그림 1, 그림 3)

실제 피크의 형태적인 특성은 그림에서 알 수 있듯이 비고조파보다 고조파의 경우에 두드러진 피크를 가진다. 만약 AR모델로 추정한 오디오 신호의 세그먼트 내에 날카롭고 주기적인 피크를 가지면 이 세그먼트는 배음을 가진 사운드로 음악적 요소를 가진 것으로 분류가능하다.

또한, 고조파를 가지는 사운드의 기본 주파수를 단구간 기본 주파수 (short-time fundamental frequency: FuF)로 정의하여 사용한다. 근사적 기본주파수 결정법에는 자기상관함수와 상호상관함수를 이용하는 방법 외 많은 방법들이 제안되어져 있으나 본 연구에서 사용한 기본 주파수 결정법은 Medan이 제안한 상호상관함수를 이용한 알고리즘을 사용한다[4,5]. 사운드가 고조파를 가지지 않을 경우 기본주파수는 0이 된다. 그림 4, 5는 음악신호와 음성신호의 단구간 기본 주파수를 보여준다.

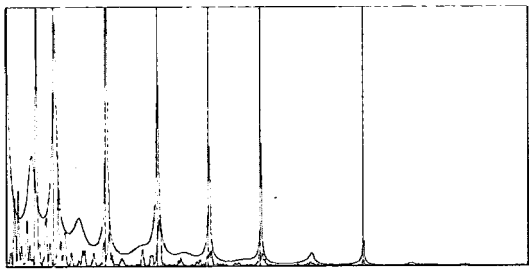


그림 1. 고조파를 가지는 음악신호
Fig. 1. Music signal with harmonics (AR model order: 40).

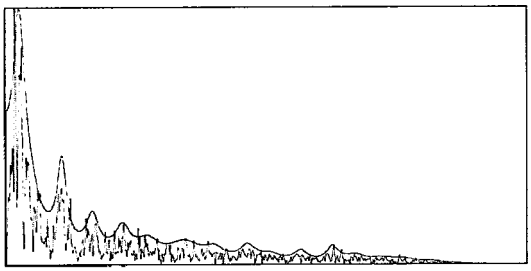


그림 2. 고조파를 가지지 않는 사운드
Fig. 2. Sound with non-harmonics (AR model order: 40).

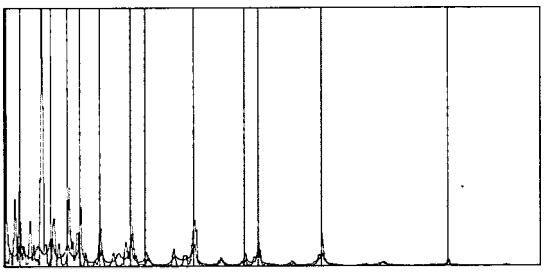


그림 3. 고조파를 가지는 음악신호
Fig. 3. Music signal with harmonics (AR model order : 80)

2.4. Spectral Peak Tracks (SPT)

오디오 신호의 시간에 연속적인 스펙트럼상에서의 피크 트랙은 사운드의 중요한 특징들을 나타낸다. 예를 들어 악기에서 발생되는 사운드는 보통 같은 주파수 레벨에서 일정 박자동안 지속되는 SPT를 가지며, 음성은 보통 빗과 같은 모양의 배음을 갖는 SPT를 가진다. 노래의 SPT는 넓은 주파수 밴드에 존재하며 기본 주파수는 87 Hz~784 Hz 범위로 알려져 있다[6]. 노래는 상대적으로 길고 안정된 트랙을 갖는데 이는 음성이 일정 기간동안 특정 음정에 머무르기 때문이며, 간혹 성대의 진동으로 리플과 같은 형태의 모양을 보이기도 한다. 대개 음성의 SPT는 저주파대에 존재하고 기본 주파수가 100~300 Hz의 범위에 존재한다. 또한 유성음간의 천이 구간이 존재하므로 짧아질 수 있으며, 특정 음절의 억양이 있는 동안에는 피치가 변경될 수 있으므로 천천히 변한다.

본 연구에서 제안하는 SPT의 추출은 단구간 신호에 대한 주파수 분석 결과 얻어진 스펙트로그램 (그림 6(a))의 파워 스펙트럼상에서 적절한 임계값을 설정하여 피크 부분을 절단 (그림 6(b))하고 절단된 부분의 중앙점을 취하여 피크의 트랙을 추출 (그림 6(c))하는 방법이다. 최종적으로 추출된 피크 트랙 (그림 6(d))은 오디오 데이터의 피크부분의 분포 특징만을 가진 히스토그램으로 전체 히스토그램보다 오디오 데이터의 내용에 따른 우세 주파수 특징만을 반영한 보다 나은 분류 분포를 가진다. 이는 설정된 색인구간에서 오디오 데이터의 내용분석에 효과적인 인자로서 이용된다. 그림 7은 설정된 카테고리마다 전형적인 SPT의 분포를 보인다. SPT의 분포는 색인 구간마다 추출되며 실제 시스템 구현시에는 묵음 구간을 고려한 스케일링으로 정규화하여 카테고리 분류가 가능한 파라메터로 이용된다.

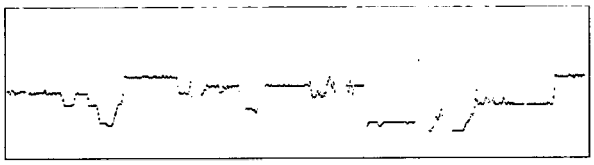


그림 4. 음악신호의 기본주파수
Fig. 4. Fundamental Freq. of misic signal

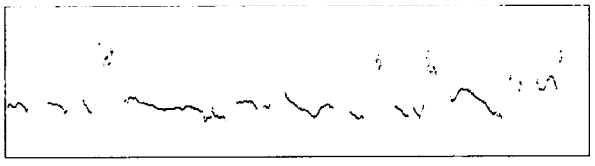
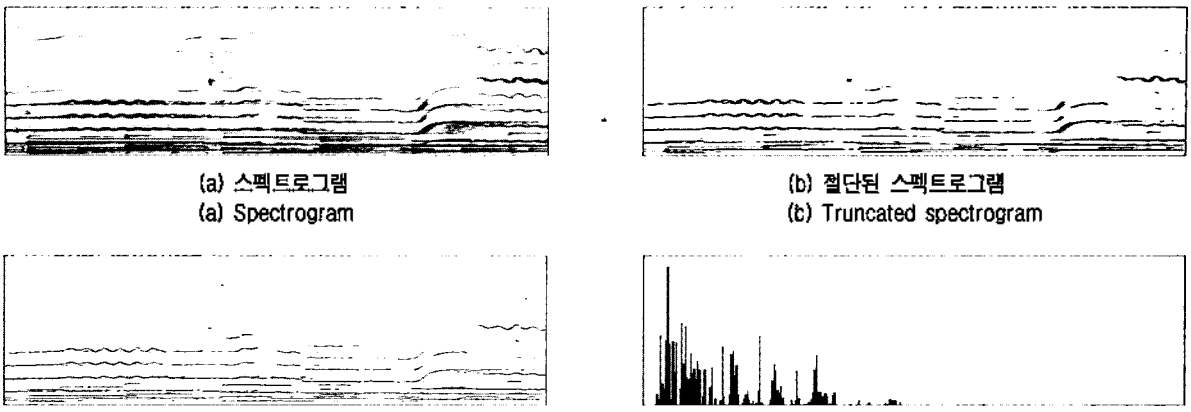


그림 5. 음성신호의 기본주파수
Fig. 5. Fundamental Freq. of speech signal.



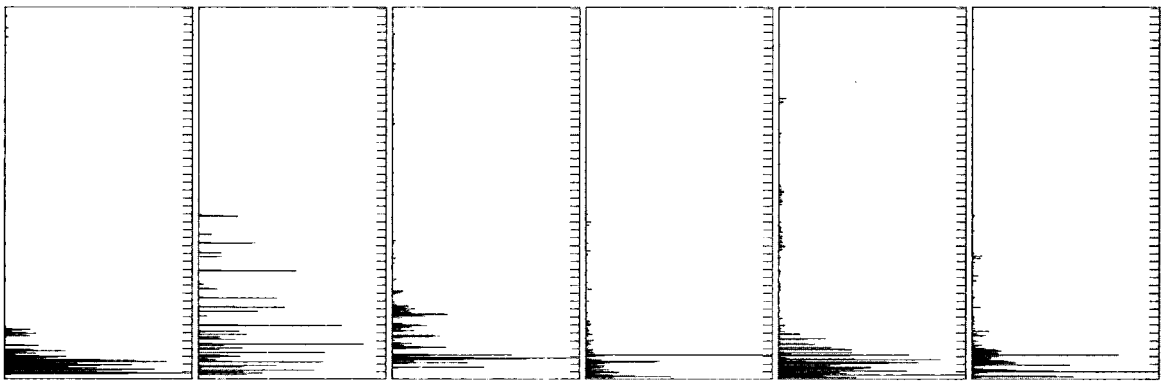
(a) 스펙트로그램
(a) Spectrogram

(b) 절단된 스펙트로그램
(b) Truncated spectrogram

(c) 추출된 SPT
(c) Extracted SPT

(d) SPT의 주파수 분포
(d) Freq. Distribution of SPT

그림 6. SPT의 추출 (이은미의 "서른즈음에" 중에서)
Fig. 6. SPT extraction.



(a) 음성
(a) Speech

(b) 음악
(b) Music

(c) 노래
(c) Song

(d) 효과음
(d) Effect sound

(e) 배경음악+노래
(e) Background music+Song

(f) 배경음악+음성
(f) Background music+Speech

그림 7. 카테고리에 따른 전형적인 SPT 분포
Fig. 7. Typical SPT distribution according to the categories.

2.5. 배음도 (Harmonic Degree)

사운드를 단구간 분석하여 해당 프레임이 배음을 가지는 경우에 음악적 요소를 가지는 것으로 인덱스 값을 1, 그렇지 않으면 0으로 설정한다. 인덱스 열에서 0의 수와 인덱스 총수의 비율을 "배음도 (이하 "HD")"라고 정의한다. 이 파라메터는 사운드에 포함되는 음악 성분이 적을수록 이 비율이 낮아지며 음악적 성분 포함 유무를 측정할 수 있는 중요한 파라메터가 된다. 색인오류는 장음, 음간의 천이, 저볼륨/저주파수 사운드에서 발생한다.

배음도의 계산은 그림 8에서와 같은 알고리즘으로 계산하며 0-1 사이의 값을 가린다.

2.6. 단구간 기본 주파수 지속도 (FuF Duration Degree)

단구간 기본 주파수 지속도 (이하 "FDD")은 일정 대역의 단구간 기본 주파수 값이 임계 프레임 이상 지속되는 구간의 총수로 정의한다. 단구간 기본 주파수의 지속도는 일정 피치가 지속되는 정도를 나타내므로 음악적 요소가 포함된 범주를 분류하는데 효과적으로 이용가능하다.

2.7. 주파수 집중도 (Freq. Convergence Degree)

주파수 집중도 (이하 "FCD")는 색인 그룹 내의 주파수의 누적치 평균값이 임계치 이상되는 주파수의 총수로 정의한다. 주파수 대역의 분포는 2장에서 언급한 SPT를 추출한 분포에서 취하는 방법, 그리고 직접 FFT하여 구하는 방법이 있다. 본 논문에서는 직접 FFT에 의한 방법을 사용하였다.

III. 오디오 데이터에 대한 내용 분석

3장에서는 묵음, 음성, 음악, 노래, 효과음, 배경음이 음악인 음성, 노래 그리고 효과음의 8가지 분류 카테고리를 설정하고 2장에서 설정한 특징 파라메터들의 기여도를 중심으로 각 색인 세그먼트마다 내용기반으로 분석한다.

3.1. 묵음

묵음은 잡음이나 단발성 click음 같은 것을 포함한 인간이 인지할 수 없는 사운드라고 정의한다. 묵음을 검출하는 일반적인 방법에는 에너지의 임계값을 설정함으로 행해진다. 그러나 잡음의 에너지 레벨은 음악과 비교하여 더 높은 경우도 있다. 잡음에 주의를 기울이지 않는 동안에도 음악을 들을 수 있는 이유는 잡음의 주파수 레벨이 아주 낮기 때문이다. 그러므로 묵음을 검출하기 위해서는 에너지와 ZCR을 사용한다. 만약 에너지가 설정 임계치보다 계속 낮거나, 세그먼트 내에서 ZCR이 특정 임계치보다 높으면 해당 세그먼트를 묵음으로 분류 가능하다.

3.2. 음성

순수한 음성은 5가지의 조건으로 분류 가능하다. 첫 번째 조건은 ZCR과 에너지의 순간적 커브 간의 상호 배타적인 관계이다.(그림 8)

음성 세그먼트에서 ZCR커브는 무성음 성분에 대하여 피크 모양, 유성음에 대하여 오목한 모양을 가진다. 반대로 에너지 커버는 유성음에서 피크를 가지고 무성음에서 오목한 형태를 나타낸다. 두 특성들 간에 상호 배타적인 관계를 이용하여 다음과 같이 음성신호를 구별한다. 즉, 최대진폭의 1/3 지점에서 ZCR과 에너지 커브를 크리핑하여 적은 부분을 제거하여 두 커버의 단일 피크부분만 남도록 하고 두 신호 잔차 커브의 내적을 계산한다. 음성의 경우, 이 내적은 대개 에너지와 ZCR의 피크가 다른 시간에 존재하므로 "0" 근방의 값을 가진다. 반면에 다른 오디오 형태에 대하여는 큰 값을 가진다. 두번째 조건은 ZCR 커버의 모양이다. 음성은 ZCR 커브의 범위가 크며 낮은 하한선을 가진다. 세번째와 네번째 조건은 각각 ZCR 커브 진폭의 범위와 분산이다. 음악 세그먼트의 경우에는 진폭의 범위와 분산이 대개 특정 임계값보다 적으며 반대로 음성의 경우에는 특정 임계값보다 크게 된다. 다섯번째 조건은 단구간 기본 주파수의 성질과 관련된다. 유성음의 성분은 고조파를 갖지만 무성음의 경우에는 고조파를 갖지 않는다. 음성은 특정 범위 내에서 일정한 고조파를 가진다. 단구간 기본 주파수 커브에서의 고조파 부분들은 에너지 커브의 피크와 관련되어 있고 반면에 단구간 기본 주파수에서의 0 부분은 에너지 커브의 볼록면에 대응한다. 다섯가지 조건에서 각각의 경우 0과 1로 나누어

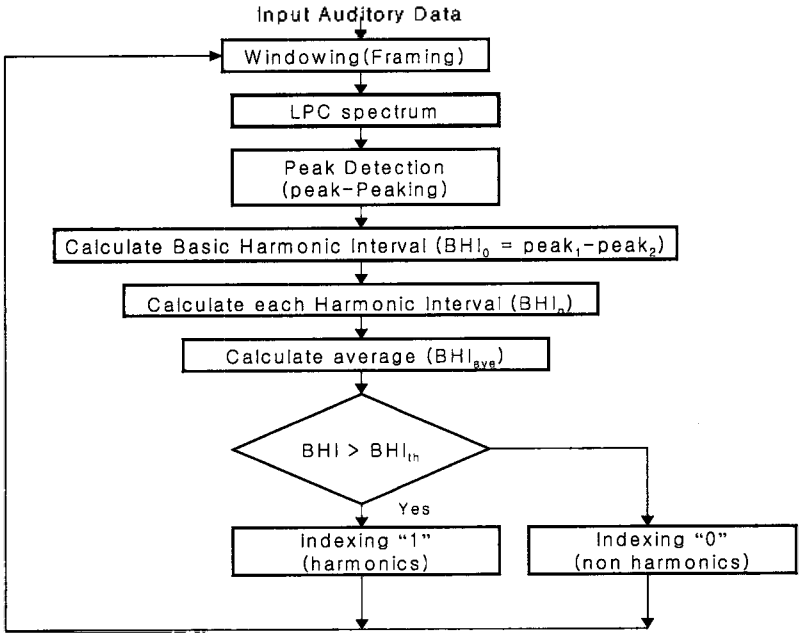


그림 8. 배음도 추출 알고리즘
Fig. 8. Harmonic degree extraction algorithm.

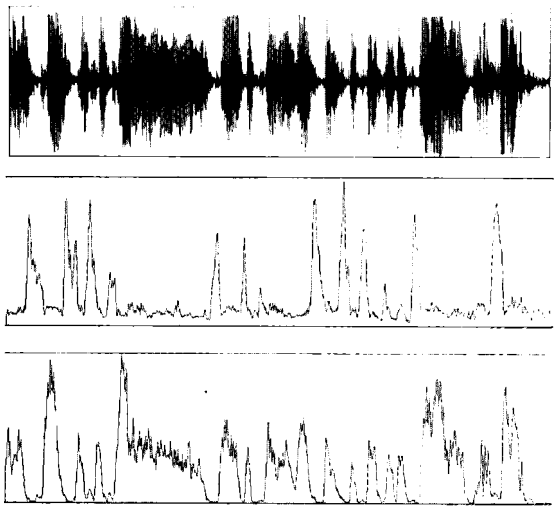


그림 9. 음성의 에너지와 ZCR의 상호 배타적 관계
Fig. 9. Exclusive relation between energy and ZCR for the speech.

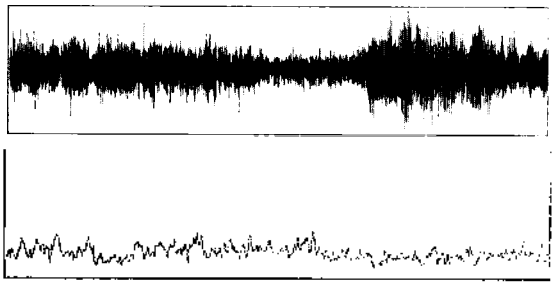


그림 10. 음악신호의 안정된 ZCR
Fig. 10. The stable ZCR of music signal.

결정값을 정하고 이들 결정값의 가중평균으로 해당 세그먼트가 음성인지 아닌지 분류가능하다.

3.3. 음악

순수한 음악은 배음도, 단구간 기본 주파수 지속도, Freq. 집중도, ZCR의 분산값과 진폭범위 등의 조건으로 조건집합을 구성하고 결정 기준값을 정하여 분류 가능하다. 음악신호의 경우, 평균 ZCR이 음성에 비해 특정 범위 내의 단구간에서 훨씬 안정된 형태이고 파형의 변화는 불규칙적이며 진폭이 음성에 비해 적은 범위를 가진다.

3.4. 노래

노래는 다음의 세 가지의 특징 중에 하나로 특징지워진다: (1) 배경음이 없이 노래만 존재하는 경우에는 성대의 진동에 기인하여 단구간 기본 주파수의 트랙은 리플모양의 고조파를 갖는 피크 트랙이 존재한다. (2) 노래는 음성에 비하여 지속시간이 길다. (3) 순수한 노래만의 데이터는 주파수가 300 Hz 이상의 높은 기본 주파수를 갖는다. (2)(3)의 특징은 SPT를 주파수 분포를 분석함으로 분류 가능하다.

3.5. 배경음이 음악인 음성, 노래

오디오 데이터의 배경음은 대부분 음악적 요소를 가진 것이므로 본 연구에서는 배경음이 음악인 경우로 제한한다. 일반적으로 음성이 강할 경우에는 배경 음악은 감추어져 검출할 수가 없다. 이 경우 HD는 1에 가까운 값을 가진다. 그러나 음성의 천이 부분이나 휴지 부분의 음악이 더 강하게 되는 영역에서는 음악 성분들이 검출되어 HD가 0에 가까운 값을 가지는 경우로 나뉘어진다. 또한 배경음이 포함된 경우 다른 카테고리에 비하여 상대적으로 높은 에너지 값을 가진다. 그러므로 에너지 값과 함께 HD의 임계값을 설정하여 일차적으로 해당 세그먼트에 대한 음악적 요소의 포함 유무를 검출하여 배경음악 유무를 분류하고 이차적으로 해당 세그먼트를 음성, 노래로 분류 가능하다.

3.6. 효과음

효과음은 비고조파 사운드 (박수소리, 발걸음소리, 폭발음)와 고조파 사운드 (차임벨, 전화의 touch-tone) 그리고 고조파와 비고조파가 혼합된 사운드 형태 (웃음소리나 개짖는 소리)의 3가지 형태로 분류 가능하다. 그러나 실제 효과음의 종류는 그 형태가 이것보다 더 다양하고 분류의 기준이 되는 정량적인 요소가 부재하므로 분류에 어려움이 있다.

IV. 분석환경과 분류 절차 구성

4.1. 실험 데이터

본 논문에서 오디오 신호의 분석을 위하여 사용한 DB는 "친구", "Sound of Music" 등의 영화 오리지널 사운드 트랙에서 16 kHz, 16비트로 샘플링하여 5초간의 오디오 클립 204개를 만들어 내용에 따라 7가지로 분류, 구축하고 분석하였다. 모의 분류 실험에 사용한 데이터는 이 중에서 카테고리마다 각각 40개의 색인단위를 취하여 실험하였다.

4.2. 실험 환경

분석방법은 프레임 사이즈 (256 points), 스텝 사이즈 (60 points)로 단구간 분석하고 2초간의 데이터 (530

표 1. 파라메터 pool의 분석결과
Table 1. Analysis result of parameter pool. SD: Standard Deviation

구분 / Parameter	Speech		Music		Song		Back+Speech		Back+Song	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Energy	161.8	137.6	704.8	601.2	435.3	537.5	295.9	180.1	1320.7	958.0
ZCR	20.4	6.9	25.4	8.3	27.5	11.2	21.9	10.4	27.3	10.8
ZCR Range	93.0	42.6	40.2	13.5	87.7	57.0	60.625	28.9	82.4	39.3
ZCR Var./10000	178	189	2	58	330	659	44	130	152	307
IP	248.7	615.8	12866.6	18028.1	2111.3	6659.2	1424.45	2512.9	8653.1	11648.9
FOD	90.9	44.7	122	85.9	123.1	102.0	90.95	42.5	60.1	35.8
HD	0.389	0.127	0.874	0.150	0.678	0.149	0.591	0.175	0.721	0.09

frame)가 기본 색인 단위가 되도록 분석하였다. 그러나, 실제 시스템의 구현에서는 사운드가 있는 부분을 검출하는 전처리를 거쳐 묵음을 제외한 사운드 세그먼트만을 분석하게 된다. 분석 도구는 윈도우 API함수를 이용하여 자체적으로 제작된 환경에서 행하였다.

4.3. 파라메터 풀의 구성

입력된 오디오는 2장과 3장에서 언급한 특징 파라메터 중에서 다음과 같은 파라메터들로 풀을 구성한다. 본 연구에서는 단구간 기본 주파수의 지속도는 실시간 처리에 어려움이 있고, 다른 파라메터로 이 특징의 분류 기준은 대체 가능하므로 제외하였다.

● 색인 그룹에 대한
① 평균 에너지 ② 평균 ZCR ③ 2/3 클리핑된 에너지와 ZCR의 내적값 ④ ZCR의 범위 (최대 ZCR - 최소 ZCR) ⑤ ZCR의 분산값 ⑥ Freq. 집중도 ⑦ 배음도 (HD)

4.4. 오디오 검색 절차의 구성

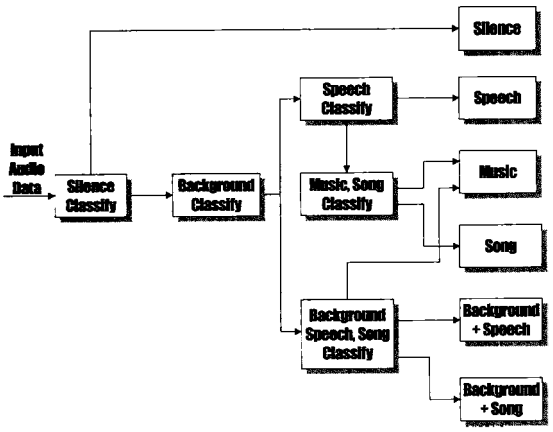


그림 11. 오디오 데이터의 분류 절차
Fig. 11. Classify procedure of auditory data.

오디오 데이터에 대한 내용기반 분석결과를 바탕으로 다음과 같은 실험에 의하여 정한 임계값에 기반한 분류 절차를 구성한다.

V. 실험 결과 및 고찰

5.1. 분류항목에 따른 파라메터 풀의 분석결과

분류항목에 따른 파라메터 풀의 분석결과는 표 1과 같다.

5.2. 분류결과

묵음인 경우에는 100%의 분류율을 보였으며 나머지 카테고리에 대하여는 표 2, 3과 같은 분류결과를 보였다.

표 2. 오디오 데이터에 대한 분류 결과
Table 2. The result of classification for the audio data.

	음성	음악	노래	배경 음성	배경 노래	분류율 (%)
음성	35	0	3	1	1	87.5
음악	2	20	8	5	5	50.0
노래	8	0	21	3	8	52.5
배경 음성	15	0	7	12	6	30.0
배경 노래	2	0	4	2	32	80.0

표 3. 오디오 데이터에 대한 분류 결과
Table 3. The result of classification for the audio data.

	음성 배경+음성	노래 배경+노래	음악	분류율 (%)
음성 배경+음성	63	17	0	78.6
노래 배경+노래	15	65	0	81.3
음악	7	13	20	50.0

VI. 결 론

본 논문에서는 실시간 오디오 색인 · 검색 시스템을 구현하기 위하여 필요한 기초 연구로 오디오 신호에 유용한 특징 파라메터들을 분석한 후, 이를 바탕으로 몇가지 새로운 파라메터를 제안하였다. 그리고 이를 바탕으로 파라메터 풀을 구성한 후, 분류절차를 구성하여 분류 실험을 행하였다. 본 논문에서는 특히 기존의 단순히 음악/음성의 범주를 확장하여 음악, 음성, 노래, 그리고 오디오 데이터의 대부분이 배경음을 포함하므로 이를 범주에 포함시켜 배경음을 가진 음성과 노래, 그리고 묵음을 포함하여 분류 범주를 확장하여 설정하였다. 본 논문에서는 오디오 데이터의 분류에 이용가능한 특징 파라메터를 중심으로 연구하였으며, 분류는 정량적인 임계값을 바탕으로 분류절차를 구성하였다. 결론적으로, 오디오 데이터와 같이 다양하고 동적인 패턴에 대한 분류를 위해서는 단일 파라메터를 이용하는 것보다는 다양한 특징파라메터를 유기적으로 조합하여 파라메터 풀을 구성하여 이용하는 것이 적절한 것으로 사료된다. 또한 분류 단계에서 이러한 파라메터 풀 내의 각 파라메터들로 차원을 형성할 경우, 기존의 성능이 뛰어난 분류 알고리즘을 이용한 분류도 가능할 것이다.

참 고 문 헌

1. M. J. Carey, "A Comparison of Feature for Speech, Music Discrimination," *Proc. ICASSP*, vol. 1, pp. 149-152, 1999.
2. J. Saunders, "Real Time Discrimination of Broadcast Speech /Music," *Proc. ICASSP*, vol. 2, pp. 141-144, 1996.
3. 이경록, 서봉수, 김진영, "오디오 인덱싱을 위한 음성/음악 분류 특징 비교," 한국음향학회지, 제20권 제2호, pp. 10-15, 2001.
4. Y. Medan, E. Yair and D. Chazan, "Super Resolution Pitch Determination of Speech Signals," *IEEE Trans. On Signal Processing*, vol. 39, no. 1, pp. 40-48, 1991.
5. 한학용, 고시영, 허강인, "우리말 연속음성의 음절분할법," 한국음향학회지, 제20권 제4호, pp. 70-75, 2001.
6. T. Zgang and C.-C. J. Kuo, "Heuristic Approach for Generic Audio Data Segmentation and Annotation," *Proc. ACM Multimedia 99*, Nov. 5, pp. 67-76, 1999.

저자 약력

● 한 학 용 (Hag Yong Han)

1994년 2월: 동아대학교 전자공학과 (공학사)
1994년~1997년: 경남에너지(주) 근무
1998년 2월: 동아대학교 대학원 전자공학과 (공학석사)
2001년 2월: 동아대학교 대학원 박사과정 수료
2001년 3월~현재: (주)이지하모니 무설 기술연구소, 동명정보대학교 정보공학부 겸임교수
※ 주관심분야: 패턴인식, DSP응용, 음성신호처리

● 김 수 훈 (Soo Hoon Kim)

1991년 2월: 동아대학교 전자공학과 (공학사)
1993년 2월: 동아대학교 대학원 전자공학과 (공학석사)
1999년 2월: 동아대학교 대학원 전자공학과 (공학박사)
2001년 3월~현재: 부천대학 정보통신계열 전임강사
※ 주관심분야: DSP, 음성인식, 신경망, 인공지능

● 허 강 인 (Kang In Hur)
한국음향학회지 제20권 제8호 참조