

# 검색의 일관성원리와 피드백을 이용한 감성기반 음악 검색 시스템

신 송 이<sup>†</sup> · 박 은 종<sup>\*\*</sup> · 엄 경 배<sup>\*\*\*</sup> · 이 준 환<sup>\*\*\*\*</sup>

## 요 약

본 논문에서는 검색의 일관성원리와 다중질의 방법을 이용하여 감성을 기반으로 음악을 검색하는 알고리즘을 제안한다. 사용되어지는 특징들은 MPEG-7에서 제공하는 오디오 서술자들이며 이 정보들을 사용하는 것은 국제표준화가 용이하다는 장점을 가지고 있기 때문이다. 또한 감성에 따라 MPEG-7 오디오 서술자들의 가중치를 부여하는 방법을 제안하여 계산량을 줄이는 방법을 고안하였으며 일관성원리와 다중질의 방법을 사용하는 적합성 피드백을 사용하여 검색의 적합률을 향상시켰다.

키워드 : 감성 기반 음악 검색 시스템, MPEG-7 오디오 서술자, 적합성 피드백

## Emotion-Based Music Retrieval Using Consistency Principle and Multi-Query Feedback

Song-Yi Shin<sup>†</sup> · En-Jong Park<sup>\*\*</sup> · Kyoung-Bae Eum<sup>\*\*\*</sup> · Joon-Whoan Lee<sup>\*\*\*\*</sup>

## ABSTRACT

In this paper, we propose the construction of multi-queries and consistency principle for the user's emotion-based music retrieval system. The features used in the system are MPEG-7 audio descriptors, which are international standards recommended for content-based audio retrievals. In addition we propose the method to determine the weight that represent the importance of each descriptor for each emotion in order to reduce the computation. Also, the proposed retrieval algorithm that uses the relevance feedback based on consistency principal and multi-queries improves the success ratio of musics corresponding to user's emotion.

Keywords : Emotion Based Music Retrieval System; MPEG-7 Audio Descriptors; Relevance Feedback

## 1. 서 론

과거의 검색은 '얼마나 많은 자료를 찾아 보여주는 가'가 중심이었다. 때문에 많은 자료를 저장하는 방법에 많은 심혈을 기울였었다. 하지만 급속도로 하드웨어 기술이 발전하여 많은 대용량의 데이터를 저장할 수 있게 된 현대시대에서는 '많은 자료를 찾아주는 것'보다 대용량의 데이터에서

사용자가 원하는 데이터를 '빠른 시간 내에 정확하게 검색'하는가, 즉 검색엔진들은 사용자들을 만족시키기 위하여 대용량의 자료를 저장하는 방법이 아닌 '어떻게 하면 사용자가 원하는 자료를 정확하게 찾아 줄 수 있는지'에 관심을 가지고 연구하고 있다[1-3]. 특히 과거 음악이나 동영상과 같은 멀티미디어 데이터는 특정한 질의어만을 이용하여 검색할 경우 단편적이고 보편적인 검색 결과만을 보여주기 때문에 사용자가 원하는 음악이나 동영상을 얻기 위하여 많은 피드백 과정이 소요되는 문제점 등을 가지고 있었다[4]. 이런 전통적인 키워드 기반의 검색방법은 사용자가 원하는 음악이나 동영상을 검색하기 위해 질의어를 적절하게 구성해서 입력해야 하는 어려움이 있고, 설명 정확한 질의어를 구성하여 입력하였다 할지라도 관리자가 해당 음악이나 동영상에 대한 키워드(Keyword)를 적절히 구성하지 않았다면 원치 않는 검색 결과를 가져오게 된다.

※ 본 연구는 지식경제부 우정사업본부의 우정기술연구개발사업의 일환으로 수행하였음. [2006-X-001-02, 실시간 우편물류 운영기술 개발]  
This work was supported by the Postal Technology R&D program of Korea Post. [2006-X-001-02, Development of Real-time Postal Logistics System].

† 준 회 원 : 전북대학교 컴퓨터공학과 공학석사

\*\* 정 회 원 : 한국전자통신 연구원

\*\*\* 정 회 원 : 군산대학교 컴퓨터정보공학과 교수

\*\*\*\* 정 회 원 : 전북대학교 전자정보공학부 교수

논문접수: 2009년 11월 16일

수정일: 1차 2010년 1월 20일, 2차 2010년 2월 9일

심사완료: 2010년 2월 10일

이러한 문제점들을 해결하기 위하여 최근 질의 자체가 단순한 텍스트 질의 형태가 아닌 하나의 음악으로 구성하고 구성된 질의 음악과 유사한 음악을 검색하는 내용기반 음악 검색 시스템을 활발히 연구하여왔다[5, 15].

기본적인 내용기반 음악 검색 방법은 질의 음악의 속성들을 사전에 미리 결정하고 결정된 속성들을 추출한다. 또한 DB에 있는 모든 후보 음악들에 대해서도 동일하게 속성들을 추출하고 사용자의 질의 음악 입력 시 질의 음악의 속성과 DB에 있는 후보 음악들의 속성들의 유사성 측도를 이용하여 계산하게 된다. 계산된 척도 값 중에서 질의 음악과 가장 유사한  $n$ 개의 후보 음악을 사용자에게 보여줌으로써 사용자가 원하는 음악을 찾아낼 수 있었다. 그러나 사용자의 개성이 중시되는 가치관의 변화에 따라 같은 질의 음악에 대해 나만의 음악을 찾고자 하는 요구가 늘어나면서 내용기반 음악 검색방법보다 보다 진보적인 검색방법인 감성기반 음악 검색방법이 연구되어 왔다.

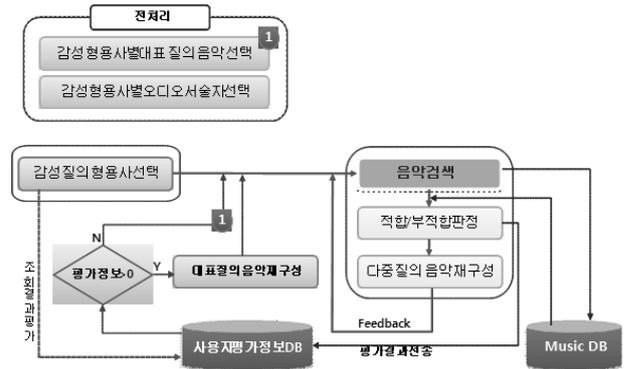
그러나 기존 연구되어진 감성기반 음악 검색 방법은 MFCC(Mel-frequency cepstral coefficients)나 LPC(Lin ear prediction coding)와 같은 속성정보들을 이용하기 때문에 표준화 하는데 어려움이 있으며, 사용자가 한번 부적합하다고 판단한 검색결과를 계속적으로 보여줌으로써 사용자가 만족을 하기위하여 다른 질의어를 입력하여 검색을 하여야 하는 번거로움이 있었다[6].

본 논문에는 이러한 단점들을 보완하고자 MPEG-7에서 제안한 오디오 서술자들(Audio Descriptors)만을 사용하여 기존의 내용기반, 감성기반 음악 검색 방법들의 확장 및 통합이 용이하도록 한다. 또한 검색하고자 하는 감성 형용사마다 그 형용사를 가장 잘 표현해주는 서술자를 사전에 미리 결정하고 서술자마다 검색에 이용되는 가중치를 다르게 하여 최적의 검색결과를 얻는 방법을 제안한다. 마지막으로 일관성 원리와 다중질의 방법을 이용한 피드백을 사용하여 사용자에게 ‘부적합’판정을 받은 음악들은 검색되지 않고, ‘적합’판정을 받은 음악들은 계속 검색되어지며, ‘적합’ 판정을 받은 음악들의 서술자를 사용하여 재검색하는 방법을 제안한다.

본 논문의 2장에서는 제안된 감성기반 음악검색 시스템의 구조, 감성모델, MPEG-7 오디오 서술자, 서술자들의 가중치 결정방법, 일관성 원리와 다중질의 방법을 통한 검색 방법을 서술하며, 3장에서는 제안된 음악검색의 방법의 실험결과와 검토를 기술하며, 4장에서 결론을 정리하였다.

## 2. 감성기반 음악 검색 시스템

제안하는 전체적인 감성기반 음악검색 시스템은 (그림 1)과 같이 감성형용사 선택, 음악검색, 적합/부적합 판정, 피드백 과정을 거친다. 즉 사용자는 원하는 감성의 음악검색을 형용사를 선택함으로써 시작하며, 선택된 형용사에 부합하는 미리 평가된 대표 음악들을 제시하며, 사용자는 대표음악 중에 ‘적합하다’라고 판정한 음악을 선택하여 유사한 감성의



(그림 1) 감성기반 음악 검색 시스템

음악검색을 시작하게 된다. 이에 시스템은 미리 결정된 서술자와 가중치를 이용하여 가장 유사한 음악들을 제시하며 검색을 완성해 나간다.

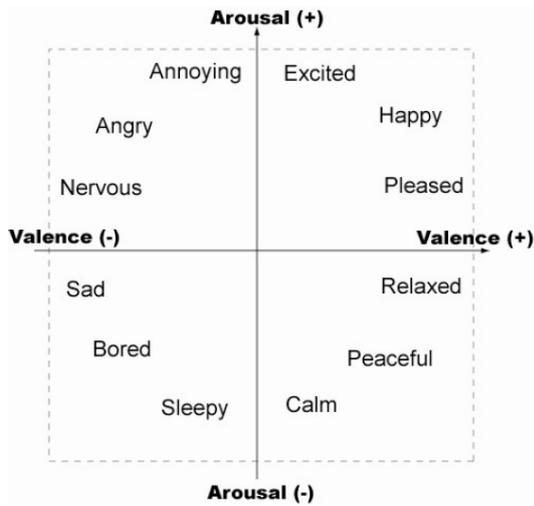
(그림 1)에서 옅 라인의 전처리부분의 감성형용사별 대표 질의 음악선택은 설문조사를 통하여 선정하였고, 감성 형용사선택은 Thayer의 감성모델을 기반으로 하였다[8]. 또한 감성형용사별 오디오 서술자 및 가중치 선택은 Mercer Kernel-Based Clustering in Feature space(MKBC) 알고리즘을 이용한 포함정도(Inclusion Degree)를 사용하여 선택하였다. 마지막 적합성 피드백(Relevance Feedback)은 일관성 원리와 다중질의 방법을 이용하였다.

### 2.1 Thayer의 감성모델

감성을 형용사로 표현하는 것과 감성을 어떻게 분류할 것인가에 해당하는 문제는 것은 매우 어렵고 모호하다[7]. Thayer의 감성모델은 이런 어렵고 모호한 문제를 해결한 하나의 모델이며 감성을 연구하는데 주로 채택되어 사용되어 왔다. 이에 본 논문에서도 임의의 음악에 대하여 사람이 느끼는 감정을 추출하는 방법으로 Thayer의 감성모델을 사용하였다.

Thayer 모델은 음악에서 느껴지는 감성을 Arousal과 Valence를 두 축을 사용하여 표현한다. Arousal은 감성의 강도를 의미하며 이 값이 커질수록 흥분한(Excited) 기분을, 작을수록 고요한(Calm) 기분을 의미한다. Valence는 감성의 긍정적/부정적 정도를 나타내며 값이 클수록 긍정적 감성을 나타내며, 값이 작을수록 부정적 감성을 나타낸다. 또한 Arousal과 Valence는 각각 하나의 축으로써 감성을 2차원 상의 좌표로 나타내며 총 12개의 감성 형용사를 정의 한다. (그림 2)는 Thayer의 감성모델을 그림으로 표현한 것이다.

Thayer의 감성모델방법을 이용하는 연구는 [9, 10] 등에서 찾아볼 수 있다. 다양한 장르의 음악에 대하여 사람들의 감성을 설문조사를 통하여 각 음악에 대한 Arousal 값과 Valence 값을 추출한 후, 이를 각 음악에서 계산된 오디오 특징들과 연관시켜 음악에 대한 감성 모델을 수립하는 방법으로 연구되어졌다. 본 연구에서는 (그림 2)에 보이는 12가지 형용사를 모두 사용하며 이들은 서로 대립을 이루는 6개의 쌍을 구성한다.



(그림 2) Thayer의 감성모델

2.2 MPEG-7 오디오 서술자

2.2.1 저수준 오디오 서술자(Low-Level Audio Descriptors)

저수준 오디오 서술자는 “표준번호 ISO/IEC FDIS 15938-4”에 정의된 오디오세그먼트(AudioSegment : 오디오를 시간적으로 분할한 일부분)에서 사용되도록 고안된 간단하고 낮은 복잡성을 가진 서술자의 집합으로 구성되며 (그림 3)에 보이는 서술자들로 구성된다[11].

(그림 3)의 서술자들을 간단히 설명하면 아래와 같다.

- *AudioSpectrumBasis* : 스펙트럼 차원의 감소 및 요약을 위해 사용되는 스펙트로그램의 통계적 기본함수.
- *AudioSpectrumCentroid* : 로그 주파수 전력 스펙트럼의 무게중심을 서술.
- *AudioSpectrumEnvelope* : 오디오 신호에 대한 전력 스펙트럼에 의한 서술.

Audio Framework	
Silence D	
Timbral Temporal LogAttackTime D Temporal Centroid D	Timbral Spectral HarmonicSpectralCentroid D HarmonicSpectralDeviation D HarmonicSpectralSpread D HarmonicSpectralVariation D SpectralCentroid D
Basic Spectral AudioSpectrumEnvelope D AudioSpectrumCentroid D AudioSpectrumSpread D AudioSpectrumFlatness D	Spectral Basis AudioSpectrumBasis D AudioSpectrumProjection D
Basic AudioWaveform D AudioPower D	Signal Parameters AudioHarmonicity D AudioFundamentalFrequency D

(그림 3) 저수준 오디오 서술자

- *AudioSpectrumFlatness* : 오디오 신호의 단기 전력 스펙트럼의 편평성을 서술.
- *AudioHarmonicicy* : 오디오 신호의 조화성을 서술.
- *AudioSignature* : 오디오의 강력한 자동식별을 위해 필요한, 유일한 콘텐츠 식별자를 제공하기 위해 고안된 오디오 신호의 축약된 표현.
- *AudioWaveform* : 오디오 신호의 파형 서술.
- *AudioFundamentalFrequency* : 오디오 신호의 기본 주파수를 서술.
- *DcOffset* : *AudioSegment*의 각 채널의 최댓값에 대해 상대적인 평균값을 기술.

2.2.2 고수준 오디오 도구(High-Level Audio Tool)

고수준 오디오 도구는 MPEG-7에서 고수준으로 분류되는 서술자와 서술 체계를 포함하는데 구조나 응용에 대한 것들이다. 이들은 저수준 오디오 서술자를 응용해서 사용하며, 크게 오디오 서명 서술자와 음색 서술자 두 가지가 있다.

- 오디오 서명(Audio Signature): 오디오의 자동식별을 위해 필요한 유일한 콘텐츠 식별자를 제공하기 위해 고안됨. 오디오 신호의 축약된 표현으로 *AudioSpectrumFlatness Type*의 연속 값에 대한 통계적 데이터 요약을 사용함.
- 음색(Timbre): 악기 소리에 대한 청각 특징을 서술하기 위한 것임. 같은 음 높이와 음 크기를 갖는 2개의 소리를 다르게 듣는 청각 특징으로서 정의됨. 음색서술 체계의 목적은 이러한 청각 특징을 몇 개의 서술자의 집합으로 서술하기 위한 것임. 이 서술자는 소리에 대한 “공격(attack)”, “밝기(brightness)” 또는 “풍요(richness)”와 같은 의미와 관련된 서술로 조화 음과 타악기 음 소리 계열에 동시에 속하는 음색의 인식을 서술하기 위하여 만들어졌으며 *InstrumentTimbreType*이라 함. *InstrumentTimbreType*의 종류는 2가지로 조화음 소리계열에 속하는 소리의 음색인식을 서술하기 위하여 만들어진 서술자 집합과, 타악음 소리 계열에 속하는 소리의 음색 인식을 서술하기 위해 만들어진 서술자의 집합이 있음. 조화음을 위한 서술자는 *HarmonicInstrumentTimbre*로 *Harmonic Spectra lCentroid, Deviation, Spread, Variation, LogAttackTime*의 집합으로 서술되며, 타악음을 위한 음색 서술자는 *Per cussiveInstrumentTimbre*로 *LogAttackTime, SpectralCentroid, TemporalCentroid*의 조합으로 서술됨.

2.3 포함정도를 이용한 서술자별 가중치 설정

감성형용사에 부합하는 음악을 검색하는데 있어 각각의 감성마다 어떤 서술자들은 다른 서술자에 비해 더 중요할 수 있다. 즉 감성마다 서술자들의 가중치를 다르게 설정하여 검색하는 것이 계산량 관점이나 검색성능관점에서 보다 효율적일 수 있다. 본 연구에서는 검색성능을 개선하기 위하여 감성마다 사용되는 서술자를 다르게 하고, 또한 사용

되는 서술자마다 중요도에 따른 가중치를 다르게 사용한다.

그러면 어떤 방법으로 중요도를 결정하는 것이 바람직할 것인가? 일반적으로 어떤 서술자의 유사성을 기반으로 군집화한 결과가 사람이 분류한 결과와 일치한다면 그런 서술자는 중요한 의미를 가질 수 있을 것이며, 만약 사람이 군집화한 결과와 전혀 일치하지 않는다면 이는 중요성이 전혀 없는 서술자로 간주할 수 있을 것이다. 또한 어떤 서술자 관점에서 유사성을 기반으로 한 군집화시 군집의 개수가 작을 때 일치하는 정도가 큰 경우가 군집의 개수가 작을 때 일치하는 정도가 큰 경우보다 더 서술자의 중요성을 판단하는데 의미가 있다. 즉 사람이 분류한 결과와 서술자의 유사성을 기반으로 하여 군집화한 결과의 포함정도(Inclusion degree)는 서술자의 중요도를 판단하는 수단이 될 수 있다.

본 논문에서는 사람이 형용사별로 분류된 음악과 이들 음악의 MPEG-7 서술자의 유사성을 기반으로 한 군집화 결과의 포함정도를 비교분석하여 해당 형용사의 음악 선택 시 서술자의 가중치를 다르게 적용하였다. 본 연구에서의 가중치 결정과정에는 어떤 서술자 관점에서 음악 상호간의 유사성을 기반으로 하는 군집화 알고리즘은 모두 사용이 가능하다. 본 연구에서는 MKBC(Mercer kernel-based clustering) 알고리즘을 적용하였다[12]. MKBC 알고리즘은 원래 커널(kernel)을 기반으로 유사도를 산출하여 군집화하는 알고리즘이기 때문에 본 논문에서 제안한 MPEG-7 오디오 서술자 관점에서 음악 상호간의 유사도를 이용한 분류에 적합하다.

일단 여러 사람들이 평균적인 의미에서 어떤 형용사  $adj$ 에 해당하는 음악을 평가/분류했다고 하고 이들 부류를  $HC_i^{adj}, (i=1,2)$ 라 하자. 본 연구에서 감성은 서로 반대되는 의미의 형용사 쌍으로 표현되는 축으로 표현되며  $HC_1^{adj}$ 는  $HC_2^{adj}$ 와 반대되는 느낌의 음악들이다. 이들 부류의 음악을 어떤 서술자  $D^j$  관점에서의 상호 유사성을 이용하여, MKBC 알고리즘이  $c$ 개의 부류로 군집화 했다고 하자. 그러면 이  $c$ 개의 부류 중에  $k$ 번째 부류의  $HC_i^{adj}$ 에 포함정도는 식 (1)과 같다.

$$v_k^i = \frac{card(X_k^c \cap HC_i^{adj})}{card(X_k^c)}, i=1,2 \quad (1)$$

식 (1)에서  $card$ 는 집합의 원소 개수(cardinality)를 의미하고,  $v_k^i$ 는 해당 서술자  $D^j$  관점의  $k$ 번째 군집 결과가 사람이 분류한 결과에 일치하는 정도를 나타내는 척도이다. 그러나 알고리즘이 군집화한 부류  $X_k^c$ 가  $HC_1^{adj}$  또는  $HC_2^{adj}$ 에 각각 반씩 포함된다면 각각 반반씩 사람이 분류한 부류에 걸쳐 나타나기 때문에 가장 의미없는 부류라 볼 수 있다. 따라서 본 연구에서는  $X_k^c$ 가  $HC_1^{adj}$ 에 소속되는 정도는  $A(x)$ ,  $HC_2^{adj}$ 에 소속되는 정도는  $B(x)$ 라고 (그림 4) 또는 식 (2)와 같이 재 정의하여 활용하였다.

$$A(x)(\text{or } B(x)) = \begin{cases} 0 & \text{when } x < \frac{1}{2} \\ 2(x-1/2) & \text{when } \frac{1}{2} \leq x \leq 1 \end{cases} \quad (2)$$

(그림 4)에서  $x$ 는 식 (1)에 의해 계산되는  $v_k^i$ 를 의미하며, 소속정도는 식 (2)와 같이 조정된다.

이러한 과정에서 조정된 소속정도는  $D^j$  관점에서  $c$ 개의 부류로 군집화 했을 경우 사람이 분류한  $HC_1^{adj}$  또는  $HC_2^{adj}$ 와 일치하는 정도  $M_c$ 를 제공한다.

$$M_c = \rho_c \frac{\sum_{k=1}^c (v_k^1 + v_k^2)}{c}, \text{ where} \quad (3)$$

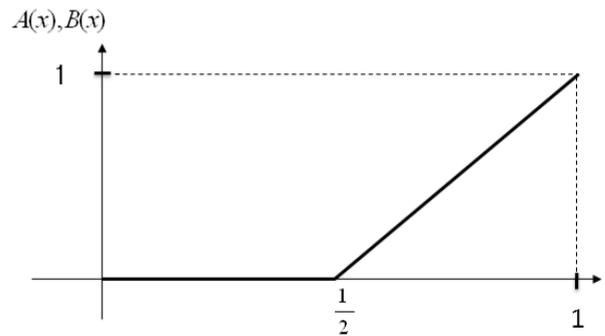
$$\rho_c = \begin{cases} -0.1c+1.2 & \text{when } 2 \leq c \leq 10 \\ 0 & \text{when } c \geq 10 \end{cases}$$

식 (3)에서  $v_1^k = A(v_k^1)$ 이며  $v_2^k = B(v_k^2)$ 를 의미한다. 식 (3)의 분자는 언제나 분모  $c$ 보다 크지 않다. 또한 MKBC 알고리즘에 의해 군집화되는 부류의 개수  $c$ 를 늘리면 작은 규모의 부류들이 생성되어 식(1)의 모든  $v_k^i$  값들 1에 가깝게 만들기 때문에 변별력을 가지기 어렵다. 이러한 이유 때문에 식(3)에서  $c$ 개의 부류로 일치하는 정도  $M_c$ 의  $\rho_c$ 는  $c$ 가 증가하면 감소하도록 구성하였다.  $D^j$  관점에서의 음악의 유사성에 기반한  $c$ 개( $c=2,3,4,5$ )의 클러스터링 결과의 포함정도를 계산하였으면 이들을 합하여 식(4)와 같이  $D^j$ 의 중요도를 구하고 이를 정규화하여 서술자  $D^j$ 의 가중치  $w^j$ 를 계산하였다.

$$ID^j = \sum_{c=2}^5 M_c \quad (4)$$

$$w^j = \frac{ID^j}{\sum_{\text{all } j} ID^j}$$

본 논문의 가중치 결정방법은 사람이 분류한 결과( $HC_1^{adj}$  또는  $HC_2^{adj}$ )와 어떤 서술자 관점에서의 작은 수의 부류로



(그림 4)  $v_k^i$ 의 소속 함수

유사성을 기반으로 한 군집화결과의 일치도가 크면 큰 가중치를 가지며 일치도가 떨어지거나, 많은 수의 부류로 분류한 결과와 일치도가 크면 작은 가중치 값을 가지게 하는 결과를 제공한다[13].

즉 군집의 파편화(fragmentation)가 일어나지 않는 범위에서 일치도를 판별하여 서술자의 가중치 내지는 중요도를 부여하는 기존 유사성 기반 러프집합이론의 문제점을 개선한 방법이다.

<표 1>은 이와 같은 방법으로 Thayer의 감성모델에 정의되어있는 “Angry-Peaceful”, “Annoying-Calm”, “Excited-Sleepy”, “Happy-Board”, “Nervous-Relaxed”, “Pleased-Sad” 등의 6가지 형용사 쌍 각각에 대하여 감성 형용사의 17개의 MPEG-7 오디오 서술자(AudioSpectrumBasis, AudioSpectrumCentroid, AudioSpectrumEnvelope, AudioSpectrumFlatness, AudioHarmonicity, DcOffest, AudioSignature, AudioWaveform, AudioSignature, AudioFundamentalFrequency, InstrumentTimbreType, HarmonicInstrumentTimbreType)관점으로 분류할 경우의 포함정도를 바탕으로 한 중요도 ID<sup>2</sup>들을 구한 것이다.

<표 2>는 <표 1>의 결과를 토대로 형용사 쌍별로 포함정도가 높은 서술자를 1개부터 5개까지 사용하여 10개의 곡을 검색 시 사용자에게 ‘적합’판정을 받은 음악의 개수이다. 서술자를 5개까지만 사용한 이유는 너무 많은 서술자를 포함하여 검색 시 계산의 속도가 느리게 하지 않으면서 실험 시 경우의 수를 줄이기 위함이었다.

<표 2>는 각 형용사별로 서술자를 1~5까지 사용하여 검색한 검색결과이다. <표 2>에서 볼 수 있듯이 “Angry,

<표 2> 각 형용사별 1~5개의 서술자를 사용한 검색결과

감성형용사 \ 서술자 개수	1개	2개	3개	4개	5개
Angry	4	4	4	4	4
Peaceful	4	4	4	4	4
Annoying	1	1	1	1	2
Calm	2	5	6	6	8
Excited	9	2	5	5	5
Sleepy	1	1	2	2	2
Happy	2	2	2	2	2
Board	3	3	3	3	3
Nervous	3	3	3	3	3
Relaxed	2	2	2	3	3
Pleased	3	4	4	1	2
Sad	5	6	6	6	6

Peaceful , Happy, Board, Nervous”는 서술자 사용 개수에 관계없이 검색된 결과가 같은 것을 알 수 있다. 따라서 사용하는 서술자의 수가 적을수록 수렴속도가 빠르기 때문에 가장 적은 수의 서술자를 사용하였다. 단 “Angry, Peaceful”의 경우 “AudioFundamentalFrequency”와 “HarmonicSpectralCentroid”의 중요도가 “1”로 같기 때문에 두 개의 특징모두 사용하였다. 나머지 감성 형용사들은 형용사별로 가장 검색이 잘된(Calm의 경우는 5개) 개수를 검색에 사용하는 서술자의 개수로 정하였다.

2.4 일관성원리 및 다중질의 방법을 이용한 피드백

사용자는 자신이 ‘부적합’하다고 판단한 정보가 계속 검색

<표 1> 6쌍 형용사 별 사용하는 오디오 서술자 및 포함정도

형용사 \ 오디오 서술자	Angry - Peaceful	Annoying - Calm	Excited - Sleepy	Happy - Board	Nervous - Relaxed	Pleased - Sad
AudioSpectrumBasis	0.568627	0.884314	0.663137	0.663137	0.663137	0.663137
AudioSpectrumCentroid	0.795378	0.583754	0.581373	0.639216	0.381373	0.495098
AudioSpectrumEnvelope	0.958824	0.617647	0.958824	0.756863	0.747339	0.288235
AudioSpectrumFlatness	0.950980	0.875686	0.95098	0.389281	0.87479	0.564379
AudioHarmonicity	0.477124	0.172549	0.728105	0.586928	0.641176	0.539869
DcOffest	0.762745	0.477124	0.460784	0.579902	0.871569	0.342577
AudioSignature	0.862745	0.806863	0.869281	0.307843	0.815966	0.466667
AudioWaveform	0.637255	0.639216	0.768137	0.713235	0.561401	0.371989
AudioSignature	0.862745	0.806863	0.869281	0.307843	0.815966	0.466667
AudioFundamentalFrequency	1.000000	0.159664	0.911765	0.425490	0.658824	0.425490
InstrumentTimbreType	0.360784	0.484314	0.229412	0.574020	0.463725	0.312745
HarmonicSpectralCentroid	1.000000	0.450000	0.823529	0.789216	0.855882	0.859804
HarmonicSpectralDeviation	0.741457	0.296471	0.788889	0.152941	0.728431	0.335714
HarmonicSpectralSpread	0.752941	0.481046	0.483399	0.668693	1.000000	0.345098
HarmonicSpectralVariation	0.809804	0.203361	0.625490	0.401961	0.509804	0.323529
LogAttackTime	0.443137	0.366667	0.172549	0.619118	0.487255	0.271569
SpectralCentroid	0.929412	0.442157	0.779412	0.813025	0.428852	0.856863
TemporalCentroid	0.647745	0.391036	0.389281	0.508263	0.823039	0.182745

되어지거나 ‘적합’하다고 판단한 정보가 재검색시에 사라진다면 검색시스템의 성능이 좋지 않다고 판단할 것이다. 일관성 원리는 이와 같은 상황을 고려하여 사용자가 적합 또는 부적합하다고 판단한 음악은 적합성 피드백을 통한 재검색시에도 적합 또는 부적합하다고 검색되도록 하는 원리이다. 또한 다중질의 방법을 이용하는 피드백은 검색된 결과에서 사용자가 적합하다고 판단한 모든 음악들의 서술자들을 이용하여 검색하는 방법이다[14].

이처럼 재검색되는 과정마다 ‘적합’하다고 판단한 음악들은 계속 검색되어지고 ‘부적합’하다고 판단된 음악들은 검색되어지지 않는 일관성 원리와, ‘적합’하다고 판단되어진 음악들의 서술자를 이용하여 피드백과정을 거쳐 다시 검색을 하는 방법은 사용자가 검색을 할 때마다 지속적으로 적용되어 최상의 검색결과를 보이도록 한다.

$$Sim_R^i = \frac{1}{n} \sum_{k=1}^n \sum_{c=1}^{DN} w^j \cdot S_c^{i,k}$$

$$Sim_{IR}^i = \frac{1}{m} \sum_{k=1}^m \sum_{c=1}^{DN} w^j \cdot S_c^{i,k} \quad (5)$$

where

- $n$  = 적합하다고 판단 된 음악의 수
- $m$  = 부적합하다고 판단 된 음악의 수
- $DN$  = 사용된 오디오 서술자의 수(Descriptor Number)
- $S_c^{i,k}$  = 서술자  $c$ 관점에서  $i, k$  음악의 유사도
- $w^j$  = 오디오 서술자  $D^j$ 의 가중치

식 (5)는 이러한 다중질의 피드백을 수행하는 수식이다.  $Sim_R^i, Sim_{IR}^i$ 은 후보 음악  $i$ 가 ‘적합’과 ‘부적합’으로 판단된 군집들과의 유사성을 각각 의미하며, DB에 있는 모든 후보 음악들에 대하여 각각의  $Sim_R^i - Sim_{IR}^i$ 를 구하고  $Sim_R^i > Sim_{IR}^i$ 를 만족하는 모든 음악을 찾아내어  $Sim_R^i - Sim_{IR}^i$ 을 내림차순으로 정렬하여 값이 큰  $N$ 개의 후보 음악을 검색하였다.

### 3. 실험결과

실험에 사용한 노래는 360곡으로 MPEG-7에서 제공하는 wav형식의 음악파일을 사용하였다. 또한 실험결과를 검증하기 위해 미리 사람이 360곡을 각 형용사 느낌이 나는 곡으로 분류하였고, 이 분류로 시스템이 검색한 결과와 비교분석하였다. 음악은 하나의 감성만 느껴지는 것이 아니므로 분류 시 형용사의 중복을 허용하였다. 형용사 별로 사람이 분류한 각 감성의 곡의 수는 “Angry”는 89곡, “Annoying”은 16곡, “Board”는 26곡, “Calm”은 96곡, “Excited”는 126곡, “Happy”는 41곡 “Nervous”는 25곡, “Peaceful”는 32곡, “Pleased”는 37곡, “Relaxed”는 35곡, “Sad”는 99곡, “Sleepy”는 16곡이었다. 실험에 사용된 노래의 부분은 곡의 감성을 잘 표현하는 각각의 절정(Climax)부분을 10초가량을 사용하였으며, MPEG-7 오디오 서술자들은 MPEG-7에서 제공하

〈표 3〉 각 형용사별 1~5개의 서술자를 사용한 검색결과

감성형용사 \ 검색횟수	1차 검색	2차 검색
Angry	4	10
Peaceful	4	10
Annoying	2	10
Calm	8	10
Excited	9	10
Sleepy	2	10
Happy	2	10
Board	3	10
Nervous	3	10
Relaxed	3	10
Pleased	4	10
Sad	6	10

는 오픈소스를 사용하여 추출하였고, 나머지 MKBC 알고리즘, 포함정도를 이용한 가중치 계산, 검색시스템은 Visual c++를 사용하여 구현하였다.

본 연구의 1차 검색결과와 2.3절의 <표 2>를 통하여 알 수 있으며, 제안한 일관성 원리와 다중질의 방법을 적용한 적합성 피드백 후의 검색결과와 <표 3>에 나타났다. 1차 검색 시 10곡을 만족시키지 못하였던 검색결과가 일관성원리를 이용한 피드백과 다중질의 방법을 이용한 2차 검색에서는 모두 만족함을 볼 수 있었으며, 검색결과와 최대 10곡으로 제한하였다. 이는 1차 검색에서 일관성 원리에 의해 만족스런 검색결과가 계속되는 검색에서 검색결과에 반드시 포함되게 하며, 만족스럽지 못한 검색결과와 포함되지 않게 하기 때문이다. 또한 1차 검색에서 만족스런 검색 결과들과 불만족스러운 검색결과 들은 2차 검색에서 다중질의를 구성하여 검색되는 음악을 만족스러운 결과들과는 가깝고 만족스럽지 못한 검색결과들은 멀게 만들었기 때문이다.

<표 3>의 내용은 검색의 정확도(precision)에 해당한다. 즉 1차 및 2차 검색에서 최대 10곡을 검색할 경우 원하는 시스템이 감성의 음악 몇 곡을 검색했는가에 해당한다. 재현율(recall)의 경우는 형용사로 표현된 감성마다 해당 음악의 수가 일정하지 않고 최대 99곡(sad)까지의 검색을 필요로 하기 때문에 생략하였다.

### 4. 결론

본 연구는 Thayer의 감성모델을 이용하여 감성형용사를 선택하고, MPEG-7에서 제공하는 오디오 서술자만을 검색에 이용하였으며, MKBC 알고리즘을 이용한 포함정도(Inclusion Degree)를 사용하여 검색하는 감성마다 사용하는 서술자와 서술자의 가중치를 달리하고, 최종적으로 일관성 원리와 다중질의 방법을 이용하는 피드백을 사용하여 검색능력의 향상방법을 제안하였다.

제안한 시스템의 특징은 크게 4가지로 정리할 수 있다.

첫째, 내용기반 검색과 감성기반 검색의 두 가지 검색 방법을 모두 적용하였다. 기본적으로 내용기반 검색의 방법을 취하고 감성기반 검색 인터페이스를 추가, 즉 감성별 대표 음원을 제공하며 이를 통하여 내용기반 검색을 함으로써 감성기반 검색을 완성하였다는 것이다. 둘째, 내용기반 검색의 국제표준인 MPEG-7 오디오 서술자를 사용하여 표준화가 용이하다는 것이다. 여기서 MPEG-7 오디오 서술자는 음원 파일의 형태와는 관련이 없으며, 어떤 형태의 파일이 재현되는 과정에서 전기적인 음원신호로부터 추출해내는 서술자를 의미한다. 셋째, 유사성 기반의 러프집합 이론을 이용한 방법을 개선한 감성별 사람이 나눈 군집과 서술자의 유사성을 이용하여 분류한 군집들을 비교하여 사람이 분류한 집합과 비슷할 경우에는 소속정도를 크게 하고 결과가 다를 경우 소속정도를 작게 하여 각 감성을 검색하는데 서술자의 중요성을 다르게 적용하여 사용한 것이다. 넷째, 적합성 피드백(relevance feedback)이 실제 검색방법과 다른 점을 착안, 일관성 원리와 다중질의 방법을 피드백에 적용하여 사용자의 검색결과를 기반으로 재검색을 하는 것이다.

### 참 고 문 헌

[1] E. Agichtein, E. Brill, S. Dumais, and R.Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences." In Proceedings of the ACM Conference on Research and Development on Information Retrieval (SIGIR), 2006.

[2] C. J. C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, G. Hullender, "Learning to Rank using Gradient Descent, in Proceedings of the International Conference on Machine Learning," 2005.

[3] K Jarvelin and J. Kekalainen. "IR evaluation methods for retrieving highly relevant documents." In Proceedings of the ACM Conference on Research and Development on Information Retrieval (SIGIR), 2000.

[4] Gerard Salton, Christopher Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," Information Processing & Management, Vol.24, No.5, pp.513-523, 1988.

[5] Yibin Zhang, Jie Zhou, "A Study On Content-Based Music Classification," IEEE Proc. 7th International Symposium on Signal Processing and Its Applications, Vol.2, pp.113-116, July, 2003.

[6] B. Logan., "Music recommendation from song sets," Proc. of ISMZR. pp.211-212. 2005.

[7] T. Li and M. Ogihara, "Content-based music similarity search and emotion detection," ICASSP, pp.17-21, 2004.

[8] R. E. Thayer, "The Biopsychology of Mood and Arousal," New York, Oxford University Press, 1989.

[9] M. D. Korhonen, D. A. Clausi, M. E. Jernigan, "Modeling Emotional Content of Music Using System Identification," IEEE Transactions on Systems, Man, and Cybernetics - Part

B : Cybernetics, Vol.36, No.3, 2006.

[10] L. Lu, D. Liu, H-J, Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals," IEEE Transactions on Audio, Speech, and Language Processing, Vol.14, No.1, 2006.

[11] Information Technology Multimedia Content Description Interface Part 4: Audio, ISO/IEC FDIS 15938-4.

[12] Mark Cirolami, "Mercer kernel-Based Clustering in Feature Space," IEEE TRANSACTIONS ON NEURAL NETWORKS, Vol.13, No.3, pp.780-784, May, 2002.

[13] Y.J. Jiang, J. Chen and X.Y. Ruan, Fuzzy similarity-based rough set method for case-based reasoning and its application in tool selection, International Journal of Machine Tools & Manufacture 46, pp.107-113, 2006.

[14] 박은중, "검색의 일관성 원리와 다중 질의 영상을 이용한 감성기반 영상 검색," 전북대학교, 박사학위 졸업논문, 2008. 8.

[15] 이준환 외 5명, "칼라영상의 감성평가와 이를 이용한 내용기반 영상검색", 정보처리논문지B, 제10권 제2호, pp.179-188, 2003.



### 신 승 이

e-mail : syshin@chonbuk.ac.kr

2008년 전북대학교 전자정보공학부(학사)

2008년~현재 전북대학교 컴퓨터공학과  
(공학석사)

관심분야: 감성기반 음악검색, 패턴인식



### 박 은 중

e-mail : for511@etri.re.kr

2002년 군산대학교 컴퓨터공학과(학사)

2005년 전북대학교 전자공학과(공학석사)

2008년 전북대학교 영상공학과(공학박사)

2008년~2009년 전라북도전략산업혁신을 위한 전자정보고급인력양성사업단(Post-  
Doc.)

2009년~현재 한국전자통신 연구원

관심분야: 감성기반 영상검색, 멀티미디어



### 엄 경 배

e-mail : kbeum@kunsan.ac.kr

1984년 전북대학교 전자공학과(학사)

1986년 전북대학교 전자공학과(공학석사)

1990년 전북대학교 전자공학과(공학박사)

1995년 캐나다 토론토대학교 전기 및 컴퓨터공학과(Post Doc.)

1989년~현재 군산대학교 컴퓨터정보공학과 교수

관심분야: 영상처리 및 컴퓨터 비전



## 이 준 환

e-mail : chlee@chonbuk.ac.kr

1980년 한양대학교 전자공학과(학사)

1982년 한국과학기술원 전자공학과(공학석사)

1982년~1985년 전북대학교 전자공학과 조교

1985년~1987년 전북대학교 전자공학과 전임  
강사

1990년 미주리대학 전산학과(공학박사)

1990년~현재 전북대학교 전자정보공학부 교수

관심분야: 내용기반 영상검색, 멀티미디어