

# VCM과 Beat Tracking을 이용한 음악의 명암 분류 기법 개발

## Development of Music Classification of Light and Shade using VCM and Beat Tracking

박승민 · 박준형 · 이영환 · 고광은 · 심귀보\*

Seung-Min Park, Jun-Heong Park, Young-Hwan Lee, Kwang-Eun Ko and Kwee-Bo Sim\*

중앙대학교 전자전기공학부

### 요 약

최근 음악을 장르로 분류하는 다양한 연구가 진행되고 있다. 하지만 이러한 분류는 전문가들마다 분류하는 기준이 서로 상이하여 정확한 결과를 도출하기가 쉽지 않다. 또한 새로운 장르 출현 시, 새롭게 정의해야 하는 번거로움이 발생한다. 따라서 음악을 장르로 구분하기 보다는 감정단어들로 분류, 검색하여야 한다. 본 논문에서는 사람이 느끼는 감정 중, 밝음과 어두움을 기준으로 음악을 분류하려고 한다. 음악이 내포하고 있는 특성들에 VCM(Variance Considered Machines)을 적용하여 음악의 명암 분류 시스템을 제안한다. 본 논문에서 이용한 음악적 특성은 3가지이다. 설문조사를 통해 명암이 정의된 기준 음악을 음의 높고 낮음의 분포, 음색의 가늘고 굵음과 비트의 빠르기를 이용하여 VCM에 먼저 학습을 시킨 후, 학습된 VCM을 통하여 분류 되지 않은 음악을 정의하여 설문조사를 통한 결과와 비교 분석 하였다. 음 추출은 Matlab을 이용하여 샘플링된 음악을 일정한 간격으로 나누어 FFT를 통해 주파수 분석을 한 후 평균값을 그 구간의 대표음이라 가정하고 추출된 음들의 높낮이를 수치화 하여 전체 분포를 파악하였다. 음색 부분에서는 음 추출에서 사용된 주파수 영역에서 전체 주파수 누적분포의 차이를 이용하여 수치화 하였다. 이 세 가지 특성을 VCM에 적용하여 실험 결과와 설문 조사 결과 비교하여 보니 약 95.4%의 확률로 음악의 명암이 분리된 것을 확인 하였다.

**키워드** : Intelligent Musical Fountain, Support Vector Machine, Variance Considered Machines, Light and Shade, Emotion, Music Analysis

### Abstract

Recently, a music genre classification has been studied. However, experts use different criteria to classify each of these classifications is difficult to derive accurate results. In addition, when the emergence of a new genre of music genre is a newly re-defined. Music as a genre rather than to separate search should be classified as emotional words. In this paper, the feelings of people on the basis of brightness and darkness tries to categorize music. The proposed classification system by applying VCM(Variance Considered Machines) is the contrast of the music. In this paper, we are using three kinds of musical characteristics. Based on surveys made throughout the learning, based on musical attributes(beat, timbre, note) was used to study in the VCM. VCM is classified by the trained compared with the results of the survey were analyzed. Note extraction using the MATLAB, sampled at regular intervals to share music via the FFT frequency analysis by the sector average is defined as representing the element extracted note by quantifying the height of the entire distribution was identified. Cumulative frequency distribution in the entire frequency rage, using the difference in Timbre and were quantified. VCM applied to these three characteristics with the experimental results by comparing the survey results to see the contrast of the music with a probability of 95.4% confirmed that the two separate.

**Key Words** : Intelligent Musical Fountain, Support Vector Machine, Variance Considered Machines, Light and Shade, Emotion, Music Analysis

## 1. 서 론

접수일자 : 2010년 10월 21일

완료일자 : 2010년 11월 25일

\* 교신 저자

※ 본 논문은 본 학회 2010년도 추계학술대회에서 선정된 우수논문입니다.

감사의 글 : 본 논문은 중소기업청·서울시의 2010년도 산학연공동기술개발사업 연구비 지원을 받아 수행된 연구입니다. 연구비 지원에 감사드립니다.

현재 지능형 음악분수에 대한 연구는 아직 시작단계에서 연구가 이루어지고 있다. 지능형 음악 분수 시스템에서는 분수 주변을 센서 탐지를 통해 상황을 인지하고 그에 따른 알맞은 음악을 실시간으로 검색하여 연결하여야 하기 때문에 정확하고 빠른 음악 정보 검색 시스템의 구축이 필수적이다.

음악 정보 검색(MIR; Music Information Retrieval) 분

야는 위에서 열거한 서비스 뿐 만 아니라 다른 다양한 분야에 적용이 가능하다. 이 때문에 또 다른 수익 창출이 기대되어 앞으로 갈수록 시장성이 기대되는 연구 분야이다. 하지만 예전과 다르게 지금 이 순간에도 수많은 새로운 음악 콘텐츠들이 생성되고 퍼져 나가고 있는 현재 상황에서 적절한 음악을 찾는 것은 점점 어려워지고 있다. 또한 음악 정보 검색은 현재 MIREX처럼 규모가 큰 콘테스트가 진행이 되고 있지만, 실질적으로 상품화하여 실생활에 적용 가능한 알고리즘 연구는 아직 미비한 실정이다[1].

현재 음악 분류, 검색 연구에서 많이 사용되는 기준은 장르이다[2]. 하지만 장르로 음악을 분류한다는 것은 여러 문제점을 지니고 있다. 장르라는 것이 음악에서 추출 가능한 특징을 가지고 분류 하는 것이 아니라 시대나 지역 등, 음악 자체와 무관한 기준들이 있다. 또한 음악 전문가들조차도 하나의 음악을 전혀 다른 장르로 구분하는 경우도 심심치 않게 접 할 수 있으며, 새로운 장르 출현 시 새로운 분류 기준 구축 및 과거 음악 재분류를 해야 하는 번거로움이 있다. 또한 같은 장르라고 해도 각각의 음악들은 서로 다른 감정을 내포하고 있기 때문에 상황에 따른 적절한 검색 결과를 나타내기가 쉽지 않다. 따라서 음악을 상황에 적절하고 빠르게 검색 가능하도록 하기 위해 음악을 장르에 따라 분류, 검색하는 것이 아니라 음악이 내재하고 있는 감정 단어들의 집합으로 분류, 검색 하여야 한다. 본 논문에서는 인간이 느낄 수 있는 여러 감정 단어 중, 밝고 어두움에 관하여 분류를 하려고 한다.

## 2. 음악에서의 명암 분류 정의

본 논문에서는 음악의 밝고 어두움을 아래와 같이 정의하였다. 아래에 나열한 정의를 음악의 특성 중, 음과 박자, 그리고 음색 등 3개의 특성을 이용하여 분류 하고자 한다.

사람은 같은 특성을 지닌 음악이라 할지라도 다른 의미의 가사를 적용하면 서로 반대되는 감정을 느낀다. 하지만 아직 음성인식 분야는 연구가 시작 되지 얼마 되지 않았고 발표된 연구 성과들도 실제 본 실험에 적용하기에 부족했다. 따라서 노래 가사의 의미를 제외하고 음악 자체의 특성으로 표현이 가능한 감정형용사들로 밝고 어두움을 정의하였다.

표 1. 음악의 명암 분류 정의  
Table 1. Definition of music of light and shade

명암	밝음	어두움
감정 형용사	경쾌하다 빠르다 발랄하다 산뜻하다 가늘다	암울하다 느리다 슬프다 음침하다 굵다
주요 특징	박자가 빠르다	박자가 느리다
	높은 음이 많다.	낮은 음이 많다.
	음색이 가늘다	음색이 두껍다

## 3. 음악적 명암 특징 추출 기법 제안

### 3.1 음 추출

음 분석 및 추출은 다음과 같은 조건에서 수행되었다. 음악은 WAV파일을 MATLAB 기본 내장 라이브러리를 이용하여 44100Hz, 16bits로 샘플링 하였다. 다음으로 메트로놈표에 의거한 알레그로 빠르기를 적용하여 1분에 120번 J(4분음표)를 연주하는 것을 기준으로 하여 음을 추출한 간격을 정의 하였다. 이보다 느리게 간격을 정의 하면 빠른 박자를 지닌 음악은 간격 내에서 여러 음이 중첩이 되기 때문에 음 추출이 정확하게 이루어지지 않을 가능성이 매우 높다. 음악이 44100Hz로 샘플링 되었기 때문에 1초에 44100개의 행렬 데이터가 존재한다. 위에서 빠르기를 알레그로를 기준으로 정하였으므로 행렬 데이터의 1/2인 22050행렬 데이터 간격 마다 FFT(Fast Fourier transform)을 적용하여 시간영역의 음 데이터를 주파수 영역으로 표현 하였다.

주파수 영역에서 가장 많이 누적된 주파수를 그 간격의 대표 주파수로 가정하고 그 주파수를 음으로 표현하였다[3]. 주파수를 음으로 표현하는 방법은 Bruce Goldsteind이 주파수와 음의 관계에 대해서 설명한 방법으로 구현 하였다[4]. 1 옥타브(Octave)를 주파수의 비가 2 : 1 인 것으로 정의 하고 각 옥타브를 7등분 하여 나란히 늘어놓은 음의 간격을 음정이라 하였다. 그래서 가장 낮은 음부터 각 음의 주파수와 비교하여 선정된 주파수가 클 경우 값이 1씩 증가하는 방법으로 음을 수치화하였다. 수치화된 음들의 평균값을 그 음악의 음 높이라 정의 하였다. 그림 1. 과 그림 2. 에서 보이는 것과 같이 어두운 음악과 밝은 음악과의 음의 분포 차이가 확연히 다른 것을 확인 할 수 있다.

표 2. 그림 1과 그림 2에 사용된 음악  
Table 2. Music used in fig 1. and fig 2.

밝 음	Uptown Girl - Billy Joel (그림 1)
어두움	Watch The World Burn - Hans Zimmer (그림 2)

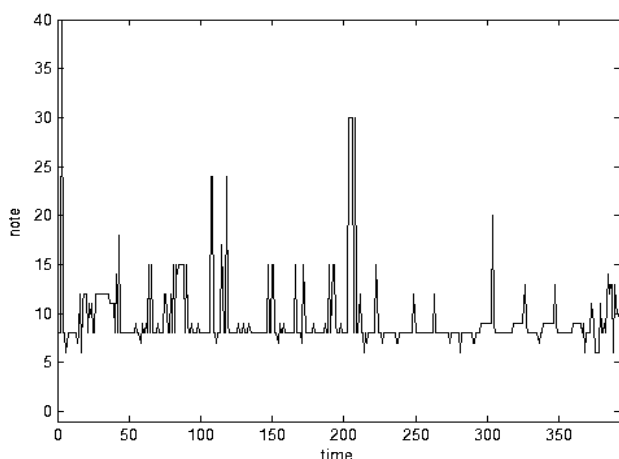


그림 1. 어두운 음악의 음 분포 중 일부, 평균 값 : 9.34  
Fig. 1. Distributon of node of shade music, Average : 9.34

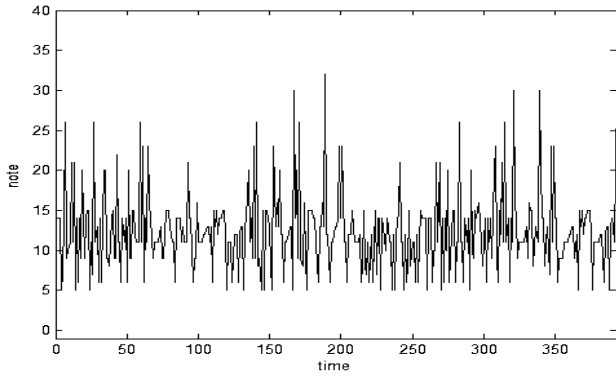


그림 2. 밝은 음악의 음 분포 중 일부, 평균 값 : 12.40  
 Fig. 2. Distributon of node of light music, Average : 12.40

**3.2 음색 추출**

음색추출은 음 추출과 비슷한 방법으로 진행하였다. 음 추출을 위해 일정한 간격으로 나누어진 후, FFT변환을 통해 시간 영역을 주파수 영역으로 표현된 상태에서 기준 주파수의 양쪽의 분포 차이를 수치화 하였다. 기준 주파수는 일반적인 남성의 음성 분포 주파수인 500Hz와 일반적인 여성의 음성 분포 주파수인 1000Hz사이의 중간인 750Hz를 기준으로 잡고 실험을 진행하였다. 이 기준 주파수를 중심으로 주파수영역을 양 쪽으로 분할하고 각 영역의 분포를 체크한 뒤 분포의 차이를 수치화 하였다. 만약 음색이 가늘면 기준 주파수보다 높은 영역의 분포가 낮은 영역의 분포보다 많기 때문에 차이는 높은 수치를 나타낼 것이며 음색이 굵으면 낮은 영역의 분포가 상대적으로 많기 때문에 낮은 수치를 나타낼 것이다.

그림 3.은 VCM 학습에 사용된 음악들의 음색 수치를 그래프로 표시한 것이다[12]. ‘o’ 가 밝은 음악이며 ‘\*’ 가 어두운 음악이다. 그래프에서 각 음악들의 음색 분포가 확연히 분류가 가능 할 수 있도록 나뉘어 표시되어 있다.

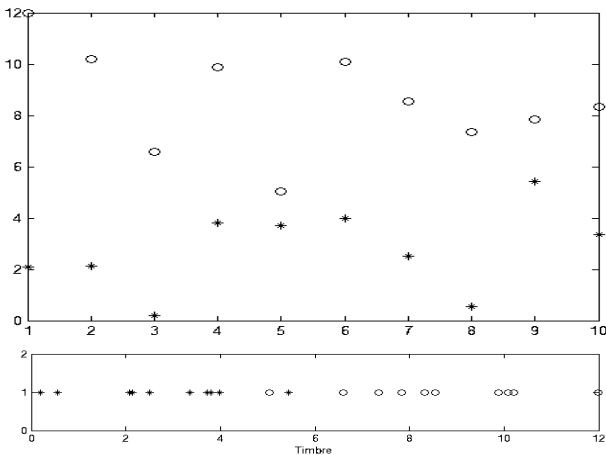


그림 3. 밝은 음악(o)과 어두운 음악(\*)의 음색 분포  
 Fig. 3. Timbre distribution of light music(o) and shade music(\*)

**3.3 박자추출**

D.P.W Ellis 는 소리 신호에서 FFT 변환을 통하여 주파

수 영역으로 변환 한 뒤, MEL 상수를 곱하여 청각적인 모델을 추출하고 미분값을 계산하여 Onset Energy를 구하고 Autocorrelation을 취하여 가장 빈번한 온셋 간의 시간을 구한다. 이때 가중치를 적절한 곳에 두고 다이내믹 프로그래밍을 사용하여 최적화된 비트를 구한하는 비트 트래킹 기술을 제안하였다[5].

본 논문에서는 D.P.W Ellis 가 제안한 비트 트래킹 기술을 개선한 알고리즘을 이용하였다. 기존의 비트 트래킹 기술은 전체 주파수 영역을 대상으로 하여 비트를 추출했다. 하지만 이 방법은 클라이막스나 다른 중요 부분에서 비트가 아닌데도 비트보다 음량이 커질 경우 비트가 묻히게 되어 비트 추출이 어려워지게 된다. 개선된 알고리즘은 대부분의 음악의 비트가 저 주파수 부분에 집중되어 있다는 가정 하에 LPF(Low Pass Filter)를 이용하여 저 주파수 부분에서 따로 비트를 추출하는 방법이다[6]. 그림 4.은 이 알고리즘을 이용하여 음악의 임의의 일부 구간에서 추출된 비트의 모습입니다.

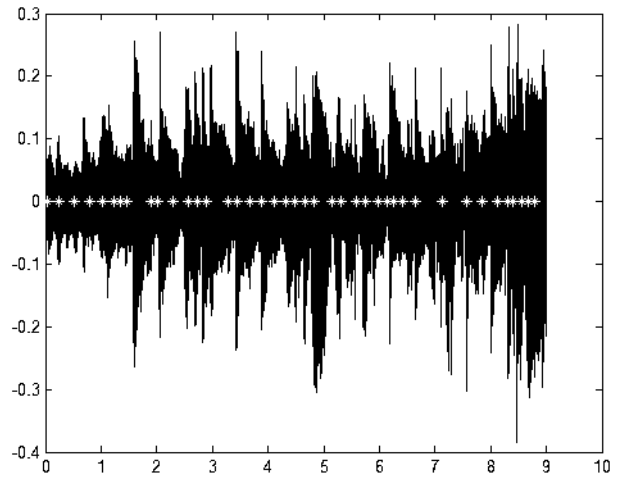


그림 4. 일부 구간의 비트(\*)위치  
 Fig. 4. Bit(\*) position in some sectors

개선된 비트 트래킹을 적용 후, 음악에서의 빠르기를 구하기 위해 추출된 비트의 개수를 측정한다. 다른 음악과 비교가 가능하도록 하기 위해 측정 후, 전체 음악의 길이로 나누어 평균값을 구하여 빠르기를 수치화 한다. 음 추출에서 사용된 음악과 같은 음악을 사용하여 비트를 추출 후 빠르기를 수치화하면 밝은 음악은 5.2168 이고 어두운 음악은 4.0362로 충분히 서로 반대되는 개념으로 분류가 가능한 차이를 보이고 있다.

**4. 음악적 명암 분류 알고리즘**

분류 알고리즘의 종류에는 여러 가지가 있다[9]. 기계 학습의 방법 중에 가장 간단한 방법 중 하나로 분류되고 있는 관찰치 특성을 기준으로 훈련 샘플 중에서 가장 가까운 관찰치들을 분류하는 K 근접 이웃(k-nearest neighbors algorithm (k-NN)), 과거에 있었던 사례들의 결과를 바탕으로 새로운 사례의 결과를 예측하는 기법인 사례기반추론(Case Based Reasoning - CBR), 광대한 탐색 공간 안에

서, 주어진 함수의 전역 최적 해에 대한 좋은 근사를 주는 전역 최적화 문제에 대한 일반적인 확률적 메타 알고리즘인 담금질 기법(Simulated Annealing), 자연세계의 진화과정에 기초한 계산 모델로서 John Holland에 의해서 1975년에 개발된 전역적인 최적화 알고리즘인 유전 알고리즘(Genetic Algorithm), 통계 분류와 회귀 분석을 쓰는, 지도 학습 방법인 Support Vector Machine등이 있다.

이 중에서 본 논문에서는 VCM을 선택하였다[12]. VCM이란 그림 5와 같이 예러의 발생확률을 줄이기 위해 SVM의 초평면을 데이터의 분산과 사전확률에 따라 수정한 알고리즘으로 SVM 초평면의 기울기는 그대로 유지하면서 y 절편을 수정하여 초평면을 데이터가 클래스에 속할 확률이 같아지는 지점으로 이동시켜 분류하는 방법이다.

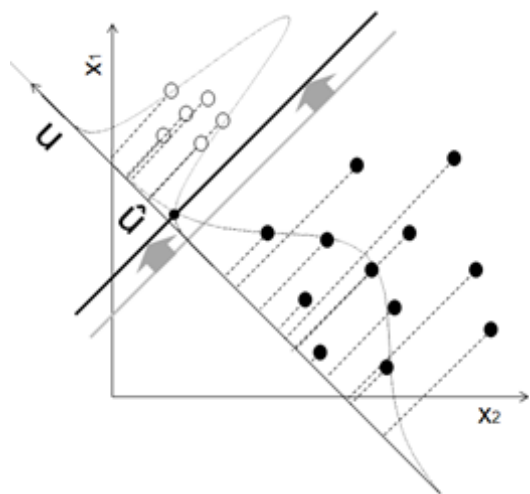


그림 5. 데이터의 분산과 존재확률을 고려하여 초평면을 이동시켜주어 최대 마진을 갖으면서도 예러가 발생할 확률이 줄어든 VCM에 의한 데이터 분류[12]

Fig. 5. Data classification by the VCM with not only maximum margin but also low error probabilities because we shift the optimal hyperplane according to the variances and prior probabilities[12]

### 5. 실험 결과 및 분석

실험에서는 클래식, 댄스, 동요 등 다양한 장르에서 각각의 강한 특징을 가지는 음악 Bach - Cello Suite No 1 in G Major, Erik Satie - Trois Gymnopdies Nol Lent et Douloureux 외 303개를 선정하였다. 총305개의 곡을 설문 조사를 통하여 밝은 음악과 어두운 음악으로 분류하였다. 설문 조사는 총 20명의 사람을 대상으로 진행 하였으며 남자 10명 여자 10명으로 구성하였다. 나이로는 10대 6명 20대 5명 30대 2명으로 성비는 모두 같도록 구성하였다. 설문 조사의 결과를 각 음악별로 밝으면 '1', 어두우면 '0' 으로 수치화한 다음 평균값을 구하고 반올림을 하는 방식으로 밝음과 어두움으로 분류 하였다. 객관적이고 전문적인 설문 진행이 아니므로 '표본추출'이나 '표본오차'와 같은 수치는 없음을 밝힌다.

먼저 나누어진 음악들 중에서 강한 특징을 보이는 음악 각각 10곡씩 20곡을 가지고 VCM을 학습시켰다. 이 만들어진 기준으로 나머지 285개의 곡을 적용 시켜 보았다. 285개의

곡을 적용 시킬 때 전체 구간을 대상으로 연산을 할 경우 연산량이 많아져 MATLAB으로 연산이 쉽지 않아 각 음악의 1분 1초 부터 2분 58초 사이의 부분을 대상으로 실험을 진행하였다. 그 결과, VCM에 적용한 285곡 중 272개의 곡이 설문 조사와 같은 결과를 나타내었지만, 13개의 곡은 예상과 다르게 분류되었다. 13개의 곡중 11개의 곡은 경계면에 가까이 있어 성공률은 약 95.4%로 기타 장르 분석 알고리즘보다 높은 성능을 보여주었다[10]. 또한 VCM과 SVM과의 성능 비교를 위해 SVM으로도 실험을 진행하였다. 그 결과 VCM이 SVM보다 1.43% 높은 성공률을 보였다.

그림 8.에서 비트(Beat)와 음(Node)의 그래프를 보았을 때 둘의 분포는 서로 상관관계가 있어 대조적으로 표현되었다. 하지만 그림 7.과 그림 9.를 보면 음색으로 기준으로 수평이나 수직으로 초평면이 생성되어있다. 음색(Timbre)이 음의 명암의 VCM 학습에서 가장 비중치를 많이 가지는 것으로 파악된다.

표 3. 분류 결과

Table 3. Result of Classification

	SVM 분류 결과	올바르게 분류된 개수
VCM	95.43%	272
SVM	94.03%	268

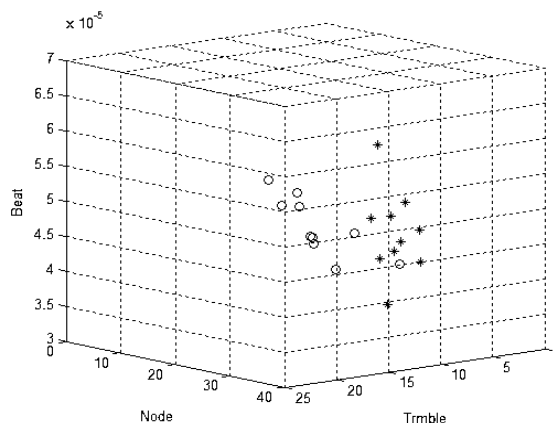


그림 6. VCM 학습  
Fig. 6. VCM learning

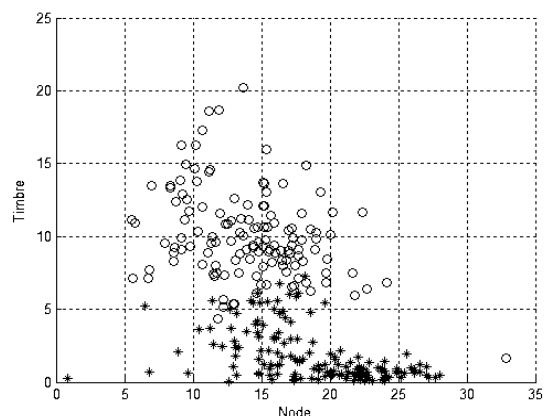


그림 7. 학습된 VCM 적용 (음색-음 그래프)  
Fig. 7. Apply learned VCM (Timbre-Node graph)

## 6. 결론 및 향후 연구

본 논문에서는 음색, 박자와 음을 이용하여 음악을 명암으로 분류하였다. 제안한 방법에 의해 음악을 분류 할 수 있었으며, 실험을 통하여 세 특징만 적용하였음에도 불구하고 장르를 기준으로 한 분류 알고리즘보다 20%정도의 성능의 개선이 있음을 알 수 있었다. 따라서 음악의 특성을 토대로 음악을 명암으로 구분하는 것이 음악 정보 검색에의 적용 가능성을 확인 할 수 있었다.

음악 분류 시 특징을 박자, 음 그리고 음색 세 가지를 가지고 적용을 하였다. 분류가 제대로 되지 않은 음악들을 조사한 결과, 분류 알고리즘이나 특성 부분에서 오류가 있었던 것이 아니라 음악 자체가 중간적인 특징을 내포하고 있어서 분류해내기가 난해했던 것으로, 설문조사 대상자들 서로 다른 의견을 제출했었던 음악들이었다.

또한 음악을 명암으로 구분하기 위해 분류 알고리즘인 VCM을 적용하였다. 추후 VCM을 학습시키기 위해 적절한 음악의 선택과 개수를 증가시키고, 이용한 세 가지 특징 뿐만 아니라 음량, 리듬, 화성, 템포 등 다른 추가적인 음악 특징을 추가하고, 음악을 임의적인 구간이 아닌 대표구간을 선택과 함께 설문조사 대상의 표본을 객관적이고 전문적으로 하여 학습데이터를 수집 후, 학습을 시키면 음악의 특징 추출과 명암 정의 알고리즘이 매우 개선되어 대중적으로 적용 가능한 최적 분리 경계면이 생성 되어 음악 분류 시스템으로 적용이 가능 할 것으로 보인다[11].

## 참고 문헌

- [1] The Music Information Retrieval Evaluation eXchange, "MIREX HOME", Available : <http://www.music-ir.org/mirex/wiki/> ,2010, [Accessed : September 3, 2010]
- [2] 윤원중, 이강규, 박규식, "Multi-feature clustering 을 이용한 강인한 내용 기반 음악 장르 분류 시스템에 관한 연구", *대한전자공학회, 전자공학회논문지* vol.42 no.3, pp. 115~120, 2005.
- [3] 백준기, *신호와 시스템*, INFINITYBOOKS, 2008
- [4] E. Bruce Goldstein, *Sensation and Perception*, Wadsworth Publishing Company, 2009.
- [5] D.P.W, "Beat Tracking by Dynamic Programming", *Journal of New Music Research* vol.36, no 3, pp.51-60, 2007.
- [6] 박승민, 박준형, 이영환, 고희은, 심귀보, "음악적 표현 최적화를 위한 Beat Tracking 기반 음악 특징 검출 기법 개발", *한국지능시스템학회, 2010 추계학술발표논문집*, pp. 63-66, 2010. 11.
- [7] Kecman, V. *Learning and Soft Computing*, Cambridge, MA: MIT Press, 2001.
- [8] Cristianini, N. and Shawe-Taylor, J, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, First Edition ,Cambridge: Cambridge University Press, 2000.
- [9] Ethem Alpaydin, *Introduction to machine learning*, MIT Press, 2004.
- [10] 이종인, 김병만, "대표구간의 음악 특징에 기반한

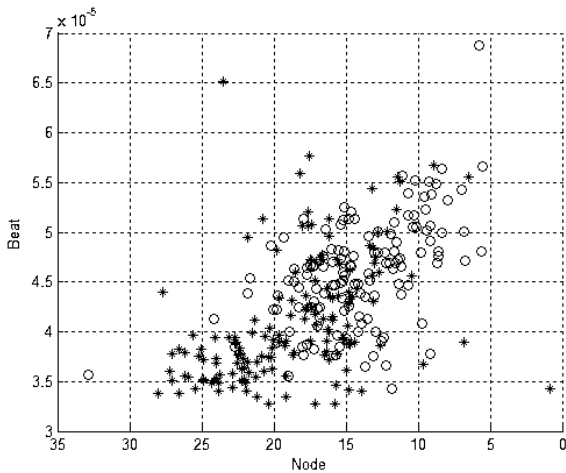


그림 8. 학습된 VCM 적용 (비트-음 그래프)  
Fig. 8. Apply learned VCM (Beat-Node graph)

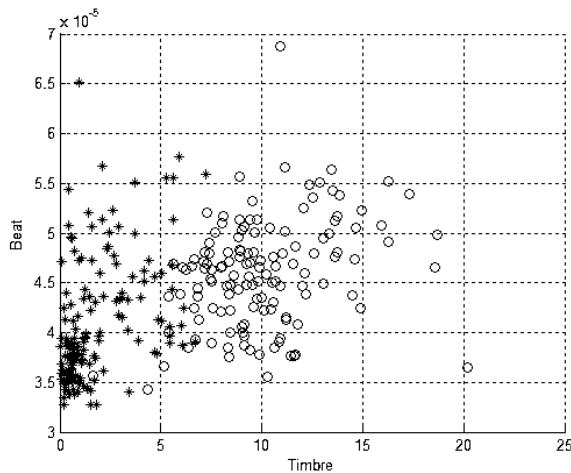


그림 9. 학습된 VCM 적용 (비트-음색 그래프)  
Fig. 9. Apply learned VCM (Beat-Timbre graph)

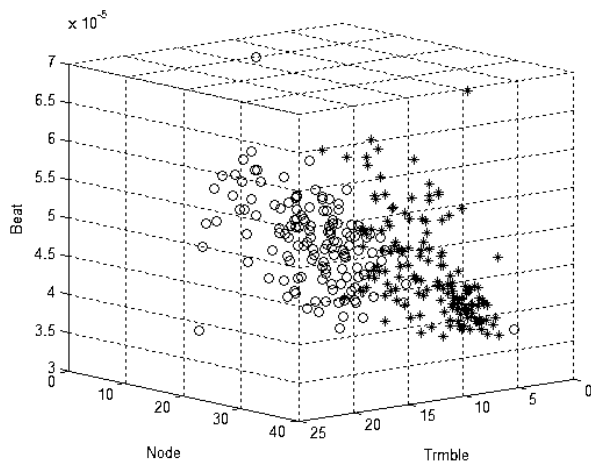
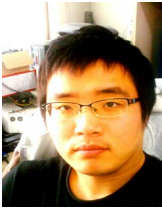


그림 10. 학습된 VCM 적용 (음-비트-음색 그래프)  
Fig. 10. Apply learned VCM (Node-Beat-Timbre graph)

음악 장르 분류”, *한국정보과학회 정보과학회논문지*  
: 소프트웨어 및 응용, vol.35 no.11 pp. 653~708,  
2008.

- [11] 김현주, "음악 장르 구분을 위한 대표 샘플 추출  
방법", *한국정보과학회 가을학술발표논문집*, vol 35,  
no 2(B), pp. 165~168, 2008.
- [12] H. G. Yeom, I. H. Jang, and K. B. Sim,  
"Variance Considered Machines: Modification of  
Optimal Hyperplanes in Support Vector  
Machines", *IEEE Int. Symp. Industrial  
Electronics*, Seoul, Korea, pp. 1144-1147, July  
2009.

## 저 자 소 개



**박승민(Seung-Min Park)**

2010년 : 중앙대학교 전자전기공학부 공학사  
2010년 ~ 현재 : 중앙대학교 대학원 전자  
전기공학부 석사과정

관심분야 : Brain-Computer Interface, Intention Recognition  
Soft Computing



**박준형(Junheong Park)**

2004년 ~ 현재 : 중앙대학교 전자전기공학부  
학사과정

관심분야 : Intention Recognition, Soft Computing



**이영환(Young-Hwan Lee)**

2009년 ~ 현재 : 중앙대학교 정보대학원  
석사과정

관심분야 : Intelligent Musical Fountain



**고광은(Kwang-Eun Ko)**

2007년 : 중앙대학교 전자전기공학부 공학사  
2007년 ~ 현재 : 중앙대학교 대학원 전자전  
기공학부 석박사통합과정

관심분야 : Multi-Agent Robotic Systems (MARS),  
Machine Learning Context Awareness, Emotion  
Recognition Systems



**심귀보 (Kwee-Bo Sim)**

1990년 : The University of Tokyo  
전자공학과 공학박사

[제20권 제4호 (2010년 8월호) 참조]

1991년 ~ 현재 : 중앙대학교 전자전기공학부 교수  
2006년 ~ 2007년 : 한국지능시스템학회 회장

E-mail : kbsim@cau.ac.kr

Homepage URL : <http://alife.cau.ac.kr>