

# 한국 전통음악 (국악)에 대한 자동 장르 분류 시스템 구현

## An Implementation of Automatic Genre Classification System for Korean Traditional Music

이 강 규\*, 윤 원 중\*, 박 규 식\*  
(Kang-Kyu Lee\*, Won-Jung Yoon\*, Kyu-Sik Park\*)

\*단국대학교 정보·컴퓨터공학부

(접수일자: 2004년 12월 1일; 채택일자: 2004년 12월 28일)

본 논문은 한국의 전통 음악, 즉 국악 장르를 자동으로 분류하는 시스템을 제안한다. 제안된 시스템은 입력 음악의 내용기반 분석을 통하여 궁중음악, 풍류방음악, 민속성악, 민속기악, 불교음악, 무속음악 등 6가지 장르중 하나로 자동 분류하여 해당 음악의 장르 결과를 보여준다. 국악 장르 분류에 사용된 내용기반 알고리즘은 크게 음악의 특징 벡터 추출 그리고 장르 분류를 위한 패턴인식 과정 2가지로 구성된다. 음악의 특징 벡터 추출은 디지털 신호 처리기술을 이용하여 해당 음악의 spectral centroid, rolloff, flux 등 STFT (Short Time Fourier Transform) 기반의 특징 계수들과 MFCC (Mel frequency cepstral coefficient), LPC (Linear predictive coding) 등의 계수들을 구한 후 SFS (Sequential Forward Selection) 최적 특징 벡터 열을 선별하여 사용하였으며 패턴 분류 알고리즘으로는 k-NN (k - Nearest Neighbor), Gaussian, GMM (Gaussian Mixture Model), SVM (Support Vector Machine) 분류기를 사용하였다. 특히 본 연구에서는 입력 질의의 패턴 (혹은 구간) 변화에 따른 시스템의 불확실성을 개선하기 위하여 MFC (Multi Feature Clustering) 방법을 이용하여 DB를 구축하였다. 모의실험 결과 k-NN 과 SVM 분류기 모두 97% 이상의 장르 분류 성공률을 보였으나, SVM 이 k-NN에 비해 약 3배 이상의 빠른 분류 성능을 가지고 있음을 확인하였다.

**핵심용어:** 국악 장르 분류, MFC, 패턴 분류, 특징 벡터 추출, SVM

**투고분야:** 음향 신호처리 분야 (1,1)

This paper proposes an automatic genre classification system for Korean traditional music. The proposed system accepts and classifies queried input music as one of the six musical genres such as Royal Shrine Music, Classical Chamber Music, Folk Song, Folk Music, Buddhist Music, Shamanist Music based on music contents. In general, content-based music genre classification consists of two stages - music feature vector extraction and pattern classification. For feature extraction, the system extracts 58 dimensional feature vectors including spectral centroid, spectral rolloff and spectral flux based on STFT and also the coefficient domain features such as LPC, MFCC, and then these features are further optimized using SFS method. For pattern or genre classification, k-NN, Gaussian, GMM and SVM algorithms are considered. In addition, the proposed system adopts MFC method to settle down the uncertainty problem of the system performance due to the different query patterns (or portions). From the experimental results, we verify the successful genre classification performance over 97% for both the k-NN and SVM classifier, however SVM classifier provides almost three times faster classification performance than the k-NN.

**Keywords:** Korean traditional music, MFC, Pattern classification, Feature extraction, SVM

**ASK subject classification:** Acoustic Signal Processing (1,1)

## I. 서론

지식기반 산업사회가 도래하면서 디지털 기술이 급속히 발전하고 있으며, 우리 생활 주변의 모든 정보들도 디지털 화 되어 가고 있다. 폭발적으로 늘어나는 방대한 데이터의 양으로 인해 앞으로는 정보 자체보다 정보의 관리 및 정보에 관한 정보가 더욱 가치 있는 정보가 될 것이다. 즉, 저장된 수많은 데이터들도 그 내용이 아무리 중요하다 할지라도 적시에 검색되어 활용될 수 없다면 가치 있는 정보라 할 수 없다. 이에 따라 방대한 양의 다양한 디지털 콘텐츠들을 효과적으로 분류 및 관리 할 수 있는 내용기반 분류 시스템에 대한 연구가 최근 몇 년간 꾸준히 관심을 모으고 있다. 그러나 대부분의 관련 연구가 디지털 영상 콘텐츠에 치우쳐져 있고, 오디오의 경우에도 음성, 특수 음향, 서양 음악 등에만 국한되어 있어 한국 전통 음악에 대한 연구가 필요한 실정이다. 한국 전통 음악인 국악은 우리 민족의 전통과 문화 양식에 가장 밀접한 관련이 있으며 옛 선조들의 문화 정신과 창의 정신이 녹아있는 중요한 문화유산 콘텐츠로 국악에 대한 체계적인 연구와 관리가 절대적으로 필요한 시점이다.

내용기반 음악 장르 분류 및 검색은 지난 몇 년간 꾸준히 관심을 모으고 있는 연구 분야로써 음악 디지털 라이브러리 (MDL), 방송연예 산업, 가상 현실 그리고 웹 어플리케이션 등 활용 분야가 무궁무진하다고 할 수 있다.

기존의 내용 기반 오디오 장르 분류 시스템은 크게 해당 음악의 특징 벡터 추출, 해당 오디오의 장르 분류를 위한 패턴인식 과정 2가지로 구성된다. Wold는 [1]에서 "Muscle Fish"라는 내용기반 분석 시스템을 개발하여 오디오를 크기 (loudness), 밝기 (brightness), 피치 (Pitch), 음색 (timbre) 등의 지각 특징들을 통해 표현하였으며, nearest neighbor (NN) 패턴 규칙을 이용하여 해당 오디오의 장르를 분류하였다. 그러나 이 시스템은 악기 소리, 동물 소리 그리고 환경 소리 등의 음향 효과 등에 국한되어 일반적인 음악 정보의 분류에 적용하기에는 무리가 있다. 음악 DB만을 대상으로 한 최초의 시도로 Tzanetakis 와 Perry [2]는 음악 신호로부터 음색 특징, 리듬 정보 그리고 피치 정보에 해당하는 특징 벡터를 추출한 후 10개 서양 음악 장르에 대하여 약 70%의 분류 성공률을 보였고, 리듬과 피치 정보만을 사용하였을 경우에는 23%~28%의 아주 저조한 결과를 나타내었다. 논문 [3]에서 Li 는 Daubechies 웨이블릿 계수와 논문 [2]에서 사용된 특징들을 선별 조합하여 여러 가지 비

교 실험을 수행하였고, 음악 장르 분류에는 리듬이나 피치 정보 보다는 음색 특징들이 훨씬 효과적이라는 결론을 내렸다. Jiang은 [4]에서 spectral contrast라는 새로운 특징 벡터를 제안하였고, Baroque, Romantic, Pop, Jazz 그리고 Rock등 5가지 서양 음악에 대하여 약 82%의 분류 성공률을 얻었다. 또한 Guo와 Li는 [5]에서 이전 트리 인식과 support vector machines (SVM)의 사용을 제안하였고, distance-from boundary (DFB)라는 새로운 측정 방법을 사용하여 음악 장르 간의 패턴 유사도를 측정하였다. 이외에도 음성, 특수음향, 서양 음악 등에 대한 특징 벡터 추출 및 장르 분류 관한 다양한 연구들이 논문 [6-8]에 설명되어있다. 이상에서 살펴본 바와 같이 대부분의 기존 연구들은 음성, 특수음향, 서양음악 등에 대한 내용기반 시스템의 분류 성능 향상을 위한 특징 벡터 추출 및 패턴 인식 기법에 관한 연구들이 주를 이루고 있다.

이상의 연구 결과들을 토대로 본 연구에서는 국악 자동 장르 분류 시스템을 구축함에 있어 3가지 사항을 고려하였다.

첫째, 입력 질의 (Query) 음악을 국립 국악원에서 공식적인 국악 장르로 분류하고 있는 궁중음악, 풍류방음악, 민속성악, 민속기악, 불교음악, 무속음악 등 6개 장르 중 하나의 장르로 자동 분류한다.

둘째, 질의 음악 처리 시 입력 질의 패턴 (혹은 구간)의 변화에 따라 발생하는 시스템의 성능 불확실성을 제거하기 위하여 전체 음악 신호를 특징자출 수 있는 강인한 특징벡터 추출 방법인 Multi-Feature Clustering (MFC) 기법을 사용하여 국악 DB를 구축한다.

셋째, 패턴 분류기에 따른 시스템의 장르 분류 성공률과 연산 속도를 비교 분석 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 특징 벡터 추출 방법에 대하여 설명하였으며 3장에서는 국악 장르 분류에 사용된 패턴 인식 알고리즘에 대하여 설명하였다. 4장에서는 모의실험을 통한 비교 분석을 수행하였고, 마지막으로는 결론으로 끝을 맺는다.

## II. 특징 벡터 추출 및 최적화

### 2.1. 특징 벡터 추출

각 특징 벡터의 분석 및 추출은 다음과 같은 조건에서

수행하였다. 각 국악 파일로부터 추출된 오디오 신호는 22050Hz, 16bits, mono로 샘플링 되었다. 오디오 신호는 시간에 따른 신호의 주파수 성분이 일정한 특성 (time-invariant, stationary)을 나타내도록 23ms 크기를 사용하였다. 각각의 프레임은 Hamming window를 50%씩 중복하여 이동하였으며 각 23ms 프레임으로부터 특징을 추출하여 평균과 분산 값을 조합해서 총 58차의 특징 벡터를 구하게 된다.

본 논문에서는 SFS (Sequential Forward Selection) 기법[9]을 사용하여 총 58차 특징 벡터로부터 최적의 벡터 열을 선택하여 질의로 사용하였으며, 또한 MFC (Multi Feature Clustering) 기법을 사용하여 음악 파일의 전 구간에 걸친 특징을 반영하여 DB를 구축하였다.

다음은 본 논문에서 사용된 특징 벡터들을 간략하게 소개하였다.

### 2.1.1. Spectral Centroid

Centroid[10]는 STFT (Short Time Fourier Transform)의 magnitude 스펙트럼의 중심을 뜻한다.

$$C_t = \frac{\sum_{n=1}^N M_t[n] * n}{\sum_{n=1}^N M_t[n]} \quad (1)$$

여기서  $M_t[n]$ 은 프레임  $t$ 와 주파수 Bin  $n$ 에서의 스펙트럼 magnitude에 해당한다. Centroid는 스펙트럼의 형태를 측정하는 방법 중의 하나이다.

### 2.1.2. Spectral Rolloff

Rolloff는 스펙트럼 magnitude 분포의 80%가 집중되어 있는 주파수  $R_t$ 이하를 나타낸다.

$$\sum_{n=1}^{R_t} M_t[n] = 0.8 * \sum_{n=1}^N m_t[n] \quad (2)$$

Rolloff는 스펙트럼의 형태와 낮은 주파수 영역에 신호의 에너지가 얼마나 집중되어 있는지를 보여준다.

### 2.1.3. Spectral Flux

Flux[10]는 연속된 스펙트럼 분포에서 정규화 된 magnitude들 간의 차이를 제공해서 구할 수 있다.

$$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2 \quad (3)$$

여기서  $N_t[n]$ ,  $N_{t-1}[n]$ 은 각각 현재 프레임  $t$ 와 이전 프레임  $t-1$ 에서의 FT의 정규화 된 magnitude이다. Flux는 스펙트럼 변화의 양을 계산할 수 있다.

### 2.1.4. ZCR (Zero Crossing Rate)

영 교차율 (Zero Crossing Rate)[11]은 오디오 신호 파형의 위상이 중심축을 통과하는 회수를 나타낸다. 영 교차율은 신호의 주파수 내용을 측정하는 가장 간단한 특징으로 음성인식에서 유·무성음의 판별에 사용된다.

### 2.1.5. LPC (Linear Predictive Coding)

선형 예측 계수(Linear Predictive Coding)[12]는 인간의 발성 모델에 입각해서 음성 신호를 부호화하는 방법으로 오디오 파형의 샘플 값에서 필터 계수를 구하여 성대에서 입, 코까지의 성도 특성을 8~12차의 전극형 (All-pole) 필터에 근사 시키는 방법이다. 본 논문에서는 10차 계수를 사용하였다.

### 2.1.6. MFCC (Mel Frequency Cepstrum Coefficient)

MFCC[13]는 인간의 청각 특성을 모델링 하는 방법으로 오디오 신호의 magnitude 스펙트럼을 log scale 한 후 FFT bin을 그룹화하여 인간의 청각 특성에 맞는 Mel-Frequency 스케일로 변환한 것이다. 본 논문에서는 13차 계수를 사용하였다.

## 2.2. SFS (Sequential Forward Selection)를

### 이용한 최적 특징 벡터 열 생성

각 23ms 프레임에서 추출하게 되는 각 특징 계수는 spectral centroid, rolloff, flux, ZCR 각 1개, 그리고 12차 LPC, 13차 MFCC로서 총 29개 계수이다. 따라서 1개의 음악 파일에 대해 각 프레임 별 특징 계수의 평균과 분산을 조합하게 되면 총 58차 특징 벡터 계수가 존재하게 된다. 본 논문에서는 SFS 기법을 사용하여 전체 특징 벡터 중 가장 성공률이 높게 나오는 차수까지의 최적 특징 벡터 열만을 선정하여 국악 장르 분류에 사용하게 된다. SFS 기법은 전체 특징 벡터에 존재하는 상관성을 제거하여, 특징 벡터의 차수를 줄일 수 있는 방법으로 장르 분류 성공률 향상과 시스템 연산 효율성을 위하여 반드시 필요한 과정이다.

SFS는 먼저 각 특징 계수들을 개별적으로 사용하여 장르 분류를 한 후, 가장 좋은 성공률을 나타내는 특징 부터 순차적으로 하나씩 특징들을 추가하면서 성공률을 계산하여 새로운 58차의 특징 벡터 열을 만든다. 맨 처음 성공률 측정시 58번의 성공률 측정이 필요하며, 특징 벡터 열이 1개씩 추가 될 때마다 성공률 측정 횟수는 1 회 씩 줄어든다. 58개의 특징들이 조합 될 수 있는 모든 경우의 수를 계산하지 않으므로 실질적으로 SFS 방식은 유사 최적 (Sub Optimal) 특징 벡터 열 선정 방식 이라고 할 수 있다. SFS를 사용한 최적 특징 벡터의 선정과 분류기의 성능은 4장에서 자세히 다루기로 한다.

### 2.3. Multi-Feature Clustering(MFC)를 이용한 국악 DB구축

동일한 질의 음악 파일 내에서도 입력 질의 패턴 (혹은 구간)에 따라 장르 분류 결과는 심각한 영향을 받을 수 있으며 이는 곧 시스템의 성능 불확실성 문제를 야기 할 수 있다. 이러한 문제를 해결하기 위해 본 논문에서는 k-means clustering 알고리즘에 기반한 multi-feature clustering (MFC) [9] 이라는 특징 벡터 추출 방법을 사용한다. MFC 알고리즘의 기본 동작은 다음과 같다. 먼저 DB로 구축될 각 음악 신호 전체 길이에 대해 15초 단위의 윈도우를 쉼표가면서 각 윈도우에 SFS를 적용하여 최적의 특징 벡터 열을 추출한다. 다음으로 이 결과로 얻어진 특징 벡터 열 들을 k-means clustering 기법을 이용하여 최종 몇 개의 특징 벡터 열로 분류한다. 기존 방법과 차이점은 DB에 구축될 하나의 음악 신호에 대해 1개의 특징 벡터만을 사용하는 대신 음악 신호 전체를 대표할 수 있는 몇 개의 특징 벡터 열을 사용한다는 점으로, 시스템 질의 구간이나 질의 길이에 따른 장르 분류의 불확실성 문제를 해결 하여 시스

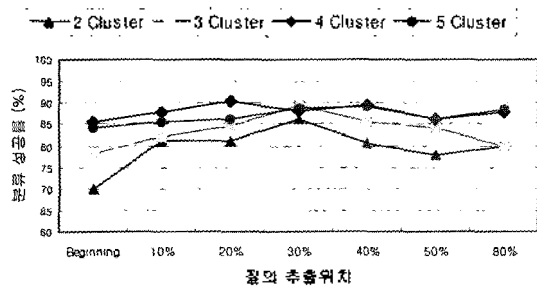


그림 1. 벡터 열 Clustering 개수와 질의 입력 위치에 따른 장르 분류 성공률  
 Fig. 1. Classification accuracy with respect to different number of clusters at varying query position.

템의 신뢰도를 높일 수 있다. 그림 1 은 음악파일 별로 clustering 된 벡터열 개수와 음악의 일정 구간에서 추출한 질의를 입력 하여 측정한 장르 분류 성공률을 나타 낸 것이다. 여기서 질의는 각 구간별 시작위치에서 15초 간 추출하여 사용하였다. 그림에서 보듯이 음악 파일 전 구간에서 추출된 특징벡터를 4개의 특징벡터 열로 clustering 하였을 경우 음악의 전 구간에서 가장 높은 성공률을 보여주고 있다. 그림 2 는 각 음악 파일을 4개의 특징 벡터 열로 clustering 한 MFC 알고리즘의 구체적인 절차를 설명하고 있다.

### III. 장르 분류 알고리즘

장르 분류 알고리즘은 질의 음악을 궁중음악, 풍류방 음악, 민속성악, 민속기악, 불교음악, 무속음악 등 6개의 국악 장르 중 하나로 분류하는데 사용되며 크게 인공 신경망을 이용한 방법과 통계적 방법 등이 있다. 일반적으로 인공 신경망을 이용한 방법은 통계적 방법의 분류 알고리즘 보다 좋은 성능을 보여주지만 복잡도가 높은 반면 통계적 패턴 인식 기법은 연산 속도가 빠르다는 장점이 있다. 본 논문에서는 일반적으로 많이 사용되고 있는 패턴 학습 및 인식 기법으로서 k-NN, Gaussian, GMM, SVM의 4가지 분류기[14] 를 사용하였다.

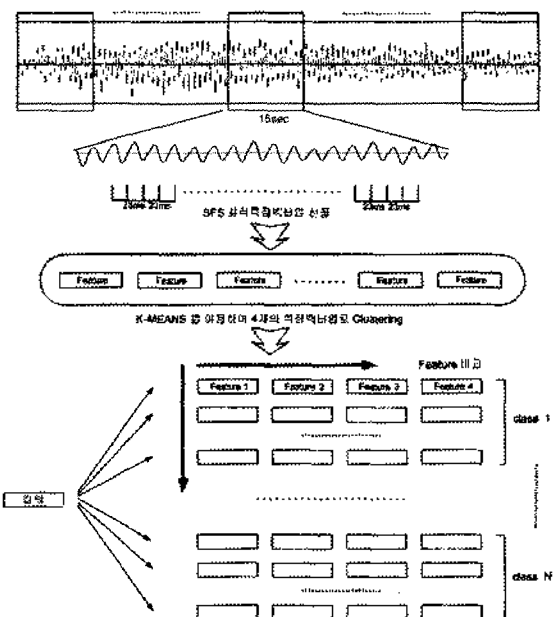


그림 2. 학습 DB 구축을 위한 Multi-feature clustering (MFC) 절차  
 Fig. 2. Multi-feature clustering for training and class classification procedure.

### 3.1. k-NN 분류기

k-NN 분류기는 임의의 특징 벡터가 어떤 장르와 더 유사한가를 결정하여 분류하는 방법이다. 특징 벡터간의 유사도를 계산하기 위해 비교적 단순하고 직관적인 접근 방법인 Euclidean 거리 함수를 이용한다. 입력된 데이터에서 추출한 특징 벡터와 미리 분류되어 있는 장르별 표준 특징 벡터와의 거리 값을 계산하여 유사도를 측정한다. K는 최근접 특징 벡터들의 개수를 의미하는 것으로서, 각 장르별로 DB에 미리 준비된 표준 특징 벡터들과의 거리를 계산하여 가장 가까운 k 개의 특징 벡터가 속해있는 장르로 질의 데이터를 분류한다.

### 3.2. Gaussian 분류기

Gaussian 분류기는 다차원의 가우시안 (Gaussian) 분포로서 각 장르를 모델화한다. 가우시안 분포는 각 장르별 샘플로부터 추출한 특징벡터의 평균과 공분산으로 이루어져 있다. 학습된 DB는 각 장르 별로 하나의 Gaussian 분포를 가진다.

Gaussian 분류기는 국악 DB내 오디오 장르별 평균 ( $\mu$ )과 공분산 ( $C$ )을 미리 계산하여 질의 오디오가 입력되었을 때 장르별 수식 (4)의 확률 밀도 함수를 계산하여 이 값이 가장 높은 장르를 해당 장르로 결정한다.

$$p(x) = \frac{1}{(2\pi)^{d/2} |C|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right] \quad (4)$$

### 3.3. GMM (Gaussian mixture model)

GMM은 각 장르별 확률 밀도 함수를 몇 개의 가우시

안 확률 밀도 함수의 선형 결합으로 정의한다. 가우시안 분류기가 각 장르별 확률 밀도 함수를 단일 가우시안으로 모델링하는 반면 GMM 분류기는 여러 Gaussian의 선형 결합으로 모델링함으로써 각 장르의 패턴 데이터들을 좀더 정확하게 묘사할 수 있는 장점이 있다. 그러나 패턴 학습 시 GMM 파라 메타 계산을 위한 EM (Expectation-Maximization) 알고리즘이 추가적으로 사용되어 높은 연산속도가 필요하며 알고리즘 구현 시 반복 루프로 인한 연산 정확도가 민감해지는 단점이 있다.

### 3.4. SVM(Support Vector Machine)

SVM (Support Vector Machine)은 입력된 학습용DB를 두 개의 장르로 구분하는 함수를 추정하는 것이다. 이 함수는 n차원의 벡터공간의 경우 (n-1)차원의 초평면 (hyperplane)의 형태로 나타난다. 단순히 주어진 특징벡터들을 두 개의 장르로 구분하는 함수는 여러개가 존재할 수 있으나, 두 장르간의 모든 점들 사이의 거리를 최대화하도록 제한을 두면 하나의 유일한 평면만이 존재하게 된다. 이러한 선형 평면 분류 경계를 OSH (optimal separating hyperplane) 라고 한다. 본 논문에서는 SVM을 이용하여 여러개의 장르를 분류하기 위한 방안으로써, 벡터공간을 d차원의 다항식으로 확장하여 비선형 분류 경계를 구하는 방식을 사용한다.

## IV. 컴퓨터 모의실험 및 분석

### 4.1. 모의실험 환경

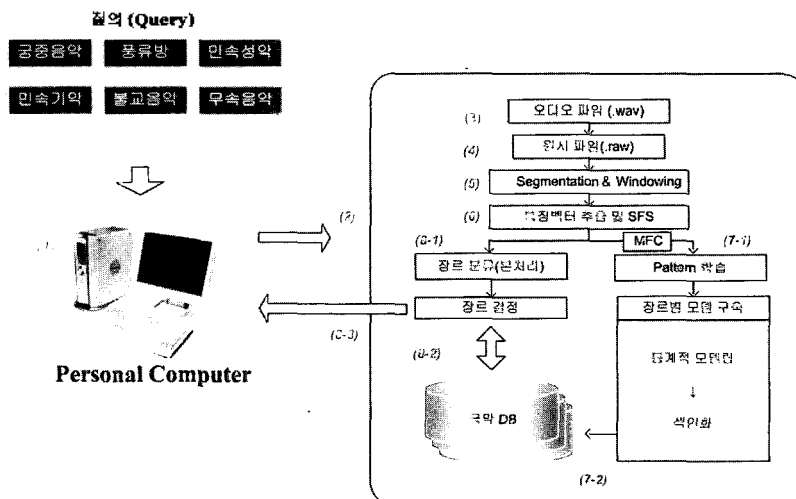


그림 3. 제안된 국악장르 분류 시스템  
Fig. 3. Overall structure of the proposed system.

제안된 시스템의 성능 분석을 위해 사용된 음악 DB는 인터넷 국악 전문 사이트, MP3 파일, 음악 CD 그리고 라디오로부터 추출하여 구축하였다. 6개의 국악장르는 국립국악원에서 공식적으로 국악 장르로 분류하고 있는 궁중음악, 풍류방음악, 민속성악, 민속기악, 불교음악, 무속음악을 사용하였으며 각 장르별로 30곡을 선정하여 총 180곡의 국악 DB를 구축하였다. 국악 DB는 2장의 MFC 기법을 이용하여 메타데이터 형식으로 구축되었으며 실험에 사용된 모든 음악 파일은 22050Hz, 16bits, mono의 wave 파일로 변환하여 사용하였다.

다음의 그림 3은 제안된 국악 장르 분류 시스템으로 입력 질의, 특징 벡터 추출, 패턴 학습 및 인식 과정을 거쳐 최종 장르 분류 결과를 제공한다.

#### 4.2. 실험 결과 및 분석

본 논문에서는 효율적인 국악장르 분류 시스템의 구축과 시스템 성능 확인을 위하여 다음의 2가지 실험을 수행하였다.

- 실험 1 : SFS 최적 특징벡터 열 생성
- 실험 2 : 분류기의 성능 평가 분석

##### 4.2.1. 실험 1: SFS 최적 특징벡터 열 생성

다음의 그림 4는 음악 파일의 약 40% 지점에서 15초 길이의 질의를 입력한 후 SFS 방법을 이용하여 각 k-NN, Gaussian, GMM, SVM 분류기 별로 구한 성공률을 보여준다.

실험 결과 k(1)-NN 이 가장 높은 분류 성공률을 보여 주었으며 SVM이 두 번째로 높은 성공률을 보여 주고있다. Gaussian 과 GMM 의 경우는 일정 수준까지 특징벡터를 더해 갈수록 성공률이 향상되었으나 필요 이상으로 많은 특징벡터를 사용한 경우 오히려 성공률이 떨어짐을 알 수 있다. 이는 K-NN 과 SVM 분류기와 달리 Gaussian 과 GMM이 통계적 모델링에 기반한 패턴 분

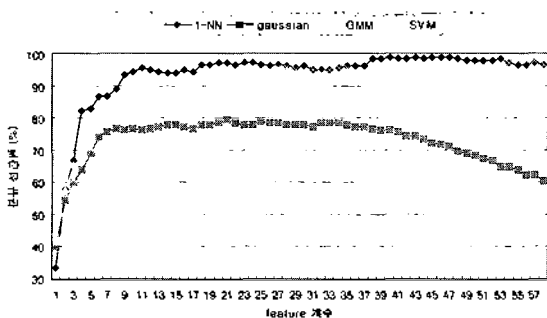


그림 4. 분류기별 SFS를 이용한 최적 특징벡터 선정 과정  
Fig. 4. classification accuracy rate using SFS method.

표 1. 최적 특징 벡터 열을 사용한 장르 분류 성공률 결과  
Table 1. Classification success rate which uses an optimum feature vector.

분류기	성공률(%)	특징벡터 차수
k-NN	98.889	40
gaussian	79.444	22
GMM	88.889	25
SVM	97.778	56

류기로서, 사용되는 특징벡터 종류에 따라 성공률에 많은 영향을 받는다는 것을 알 수 있다.

k(1)-NN 의 경우 40개의 최적화된 특징벡터를 사용한 경우 98.889% 의 성공률을 보여 주었으며, SVM은 56개의 특징벡터를 이용하여 97.778%의 성공률을 보였다. 다음의 표 1, 은 각 분류기별 최적 특징 벡터 열에서의 성공률을 나타낸다.

표 2, 표 3, 은 각각 k-NN, SVM 을 이용하여 분류한 결과를 나타낸다. 궁중음악의 경우 다른 장르의 음악에 비하여 음악의 박자, 사용되는 악기 등에서 뚜렷한 차이가 있으며 이로 인하여 거의 모든 분류기가 정확하게 분류를 해 내었다. 그러나, 풍류방 음악의 경우 시조가 많이 포함되어 있고, 민속 성악의 경우 판소리가 주류를 이루고 있으며, 이렇게 두 장르 모두 사람의 음성을 주로 사용하는 음악이 주류를 이루고 있는 관계로 풍류방 음악과 민속 성악 간에 오분류 되는 경우가 발생하였다. 불교 음악의 경우는 염불, 반야심경, 회심곡 등에서 민속 성악 및 기악과 비슷한 음악적 특징을 가지고 있으며, 무속 음악의 경우 굿, 시나위, 살풀이 등 가장 다양한 장르의 음악적 특징을 가지고 있다. 분류기별 특징을 살펴 보면, k-NN의 경우 풍류방과, 민속 기악등 주로 음악적 특징이 강한 장르의 경우 오분류가 나타날 확률이 높았으며, SVM의 경우 민속성악, 불교음악, 무속음악 등에서 오분류된 패턴을 보았을때 음악에 음성적 특징이 강할 경우 장르에 대한 성공률이 약간 떨어지는 것으로 보인다.

국악 장르 분류 성공률에 가장 큰 영향을 미치는 최적 특징 벡터 열은 사용된 패턴 분류기에 따라 약간씩 달라지는 경향이 있지만 대체적으로 LPC 계수와 MFCC 계수가 분류 성공률에 가장 큰 영향을 주었으며, 다음으로는 FLUX, Rolloff, Centroid, ZCR 등의 순서대로 분류 성공률에 영향을 주었다. LPC 계수 경우 인간의 음성 발생을 모델링 한 것으로서 국악에 포함되어있는 음성 신호가 장르 분류에 큰 영향을 미친다고 볼 수 있다. 한편 MFCC 계수 경우는 사람이 음악을 들을 때 민감하게 반

표 2. k (1)-NN을 이용하여 분류한 결과  
Table 2. confusion matrix (k-NN).

		질의 장르					
		궁중음악	풍류방	민속성악	민속기악	불교음악	무속음악
DB	궁중음악	100	0	0	0	0	0
	풍류방	0	96.667	0	0	0	0
	민속성악	0	0	100	0	0	0
	민속기악	0	0	0	96.667	0	0
	불교음악	0	0	0	3.333	100	0
	무속음악	0	3.333	0	0	0	100
분류성공률		98.889 %					

표 3. SVM을 이용하여 분류한 결과  
Table 3. confusion matrix (SVM).

		질의 장르					
		궁중음악	풍류방	민속성악	민속기악	불교음악	무속음악
DB	궁중음악	100	0	0	0	0	3.333
	풍류방	0	100	3.333	0	0	0
	민속성악	0	0	96.667	0	3.333	0
	민속기악	0	0	0	100	0	0
	불교음악	0	0	0	0	96.667	3.333
	무속음악	0	0	0	0	0	93.333
분류성공률		97.778 %					

용하는 주파수 대역의 청각 특징들을 계수로 표현한 것으로 이 역시 장르 분류 성공률에 큰 영향을 미치고 있다. 반면 음악의 주파수적 특성 변화만을 고려한 STFT 기반의 특징벡터들은 장르 분류 성공률에 그다지 큰 기여를 하지 않는 것으로 볼 수 있으며 [3]에서도 이와 유사한 결론을 내리고 있음을 알 수 있다.

4.2.2. 실험 2: 분류기의 성능 평가 분석

실용적인 자동 장르 분류 시스템을 구축하기 위해서는 시스템의 안정성과 연산 속도가 매우 중요한 요소이다.

다음의 그림 5는 음악을 6개의 구간으로 분리한 후 각각의 구간의 시작점에서부터 15초 길이를 추출하여 분류한 결과로써 질의 구간 변화에 따른 장르 분류 성공률을 나타내고 있다.

k(1)-NN 의 경우 전 구간에서 90% 이상의 고른 성공률을 보이고 있고, SVM 분류기 역시 음악의 첫 시작 지점에 해당하는 질의 경우를 제외하고 전 구간에서 90%

이상의 성공률을 보여 주었다. 반면 Gaussian 과 GMM 분류기는 질의 구간 변화에 따른 장르 분류 성공률의 변화가 큰 것을 볼 수 있다.

제한된 시스템이 실용적으로 사용되기 위해서는 분류 성공률 이외에도 분류 결과를 빠르게 제공하는 것이 중요하다. 다음의 그림 6은 4개의 장르분류기 별로 최대 성공률을 보인 벡터 열을 이용한 분류 시간을 측정하는 것이다. 분류 시간 측정은 MATLAB의 clock 함수를 이용하였다. 객관적인 연산시간 측정을 위하여 각 분류기 별로 프로그래밍 방식에 따른 연산시간 변동이 없도록, 분류 수식이 연산되는 구간만을 분류 시간을 측정에 사용하였다. k-NN의 경우 질의와 국악DB 간의 거리값을 연산하는 시간을 측정하였고, 나머지 분류기의 경우는 장르별 모델링 (Gaussian, GMM), 결정 경계면 계산 (SVM)에 걸리는 시간을 제외하고 장르 분류에 소요되는 시간만을 측정하였다.

모든 측정치는 매회 측정시 커다란 변동이 없는 관계

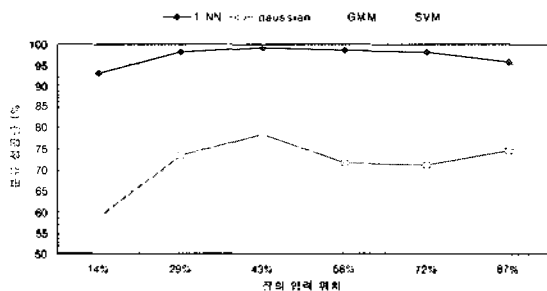


그림 5. 질의 구간별 분류성공률  
Fig. 5. Classification results with different query portions.

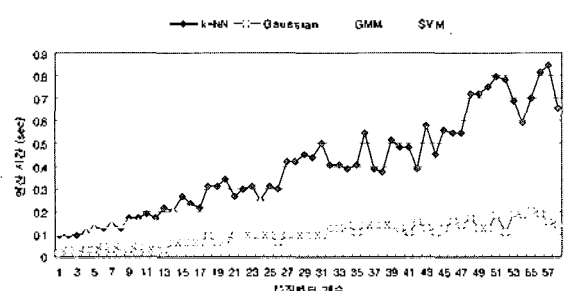


그림 6. 분류기별 연산 속도 측정 결과  
Fig. 4. Classify operation speed measurement result.

로 3회 측정 후 평균값을 취했다. 분류 시간 측정에 사용된 컴퓨터 시스템 사양은 Intel Xeon 2.4Ghz Dual Cpu, 2GB Ram 이다.

그림 6. 에서 보듯이 SFS 결과 가장 높은 성공률을 보였던 k-NN 분류기의 경우 타 분류기에 비하여 2배 이상의 연산 시간이 걸렸다. 최대 성공률이 k-NN 에 비해 1% 정도 떨어졌던 SVM 의 경우 Gaussain 분류기와 함께 k-NN에 비하여 3배 이상의 빠른 연산 속도를 보여 주었다. 이는 k(1)-NN의 DB내 모든 국악 메타데이터에 대해 유클리디안 거리 값을 구하여 일일이 비교하는 방식임에 비해, SVM 의 경우 한번 결정함수를 구해 놓으면 다음 분류부터는 결정 함수를 계산하는 과정이 필요 없기 때문에 연산 속도 면에서 유리하다 할 수 있다. 그림 5 와 6, 그리고 표 1로부터 장르 분류 성공률과 연산 속도, 최적 특징벡터열의 차수를 모두 감안하였을 경우 SVM 분류기가 가장 효율적임을 알 수 있다.

## V. 결론 및 향후 연구과제

본 논문에서는 질의 음악을 국립 국악원에서 공식적인 국악 장르로 분류하고 있는 궁중음악, 풍류방음악, 민속성악, 민속기악, 불교음악, 무속음악의 6개 장르 중 하나로 자동 분류하는 내용기반 국악 장르 분류 시스템을 제안하였다. 각 음악에 대해 총 58개의 특징벡터를 추출한 후 SFS 기법을 이용하여 최적의 특징벡터 열만을 선별하였으며, 장르 분류 알고리즘으로는 k-NN, Gaussian, GMM, SVM 분류기를 사용하였다. 특히 본 연구에서는 입력 질의의 패턴 (혹은 구간) 변화에 따른 시스템의 불확실성을 개선하기 위하여MFC (Multi Feature Clustring) 방법을 이용하여 DB를 구축하였다. 총 180곡의 국악 DB를 대상으로 한 컴퓨터 모의실험 및 성능분석 결과 SVM 분류기가 장르 분류 성공률 (97% 이상) 과 연산속도 면에서 가장 우수함을 확인하였다.

향후 연구 방향으로써 다양한 국악 장르에 대한 신뢰성 있는 내용기반 시스템 구현, 마이크로폰 질의 입력 음악에 대한 실시간 장르 분류 및 검색 시스템 구현 등을 연구중에 있다.

## 감사의 글

본 논문은 정보통신부의 출연금으로 수행한 IT Soc 사업단의 연구비지원으로 이루어 졌습니다. 본 논문의 내용을 발표할 때는 반드시 IT SoC 핵심 설계 인력 양성 사업의 결과임을 밝혀야 합니다.

## 참고 문헌

1. E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio," IEEE Multimedia, 3(2), 1996.
2. G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Trans. on Speech and Audio Processing, 10(5), pp. 293-302, July 2002.
3. T. Li, M. Ogihara and Q. Li, "A comparative study on content-based music genre classification," in Proc. of the 26th annual internal ACM SIGIR, pp. 282-289, ACM Press, July 2003.
4. D. N. Jiang, L. Lu and H. J. Zhang, "Music type classification by spectra contrast features," in Proc. ICME2002, 2002, pp. 113-116.
5. G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machine," IEEE Trans. on neural networks, 14(1), pp. 209-215, Jan. 2003.
6. J. Foote et al, "An overview of audio information retrieval," ACM-Springer Multimedia Systems, 7(1), pp. 2-11, Jan. 1999.
7. Y. Wang, Z. Liu and J. Huang, "Multimedia content analysis: using both audio and visual clues," IEEE Signal Proc. Mag., Nov. 20000
8. S. Blackburn, "Content based retrieval and navigation of music", 1999, Mini-thesis, University of Southampton.
9. Kyu-Sik Park, Won-Jung Yoon, Kang-Kue Lee, "A Robust Approach to Content-Based Musical Genre Classification and Retrieval Using Multi-Feature Clustering" ASIAN2004, Thailand, Dec. 2004
10. J. M. Gray, "An Exploration of Musical Timbre". PhD thesis, Dept. of Psychology, Stanford University, 1975.
11. M. J. Carey, E. S. Parris, and H. Lloyd-Thomas, "A comparison of features for speech, music discrimination", In Proc. ICASSP, pp. 1432-1436, March 1999.
12. J. Makhoul, "Linear prediction: A tutorial overview", Proceedings of the IEEE, Apr. 1975.
13. B.Logan, "Mel Frequency Cepstral Coefficients for Music Modeling". In proc. Int. Symposium on Music Information Retrieval (ISMIR), 2000.
14. R. Duda, P. Hart and D. Stork, Pattern Classification, 2nd Ed., Wiley-Interscience Publication, 2001



---

## 저자 이력

---

- 이 강 규 (Kang-Kyu Lee)

한국음향학회지 제 23권 7호 참조

- 윤 원 중 (Won-Jung Yoon)

한국음향학회지 제 23권 7호 참조

- 박 규 식 (Kyu-Sik Park)

한국음향학회지 제 23권 7호 참조