

저작권 보호를 위한 HMM기반의 음악 식별 시스템 HMM-based Music Identification System for Copyright Protection

김희동¹⁾ · 김도현²⁾ · 김지환³⁾

Kim, Hee-Dong · Kim, Do-Hyun · Kim, Ji-Hwan

ABSTRACT

In this paper, in order to protect music copyrights, we propose a music identification system which is scalable to the number of pieces of registered music and robust to signal-level variations of registered music. For its implementation, we define the new concepts of 'music word' and 'music phoneme' as recognition units to construct 'music acoustic models'. Then, with these concepts, we apply the HMM-based framework used in continuous speech recognition to identify the music. Each music file is transformed to a sequence of 39-dimensional vectors. This sequence of vectors is represented as ordered states with Gaussian mixtures. These ordered states are trained using Baum-Welch re-estimation method. Music files with a suspicious copyright are also transformed to a sequence of vectors. Then, the most probable music file is identified using Viterbi algorithm through the music identification network. We implemented a music identification system for 1,000 MP3 music files and tested this system with variations in terms of MP3 bit rate and music speed rate. Our proposed music identification system demonstrates robust performance to signal variations. In addition, scalability of this system is independent of the number of registered music files, since our system is based on HMM method.

Keywords: music identification, HMM, copyright protection, music word, music phoneme

1. 서론

웹 서버의 대형화 및 자료 전송속도 향상으로 인하여 이용자 간의 대용량 멀티미디어 파일의 공유가 증가하고 있다. 특히 멀티미디어 파일 중 음악자료는 비교적 작은 용량과, 휴대용 음악 플레이어의 높은 보급으로 웹상에서 공유가 빈번하게 일어나고 있다. 공유되고 있는 많은 수의 음악파일들은 저작권의 보호를 받아야 하는 자료이나 저작권의 침해가 빈번히 일어나고 있다. 이 문제는 음반산업의 침체, 국민의 저작권 의식의 약화 등을 초래하고 있다.

웹상에서 공유가 되는 파일은 개수가 너무나 많기 때문에 파일별 저작권 침해 여부를 판단은 수동으로는 불가능 하다. 따라

서 공유된 음악파일들에 대한 저작권 위배여부를 효과적으로 식별해 주는 자동화된 시스템이 필요하다. 상용화된 시스템에서 일반적으로 볼 수 있는 필터링 시스템은 파일명, ID3TAG 등 텍스트기반으로 하는 것이다. 이 방법은 같은 노래지만 단순히 텍스트정보를 변형하더라도 필터링이 제대로 되지 않게 되어 텍스트변형에 취약하다.

음악의 오디오 정보를 이용하여 자동으로 저작권 침해 여부를 판단하기 위해서, 시스템에 등록된 저작권 보호 대상 음악들에 대해서 다음의 두가지 기능의 구현을 필요로 한다. 1) 등록된 음악들에 대해 업로드 하고자 하는 음악의 식별 기능 2) 업로드 하고자 하는 음악이 등록된 음악들 중 하나가 아니어도 등록된 음악들 중 확률값이 가장 큰 음악이 식별 결과로 생성 되기 때문에, 식별 결과를 분석하여 등록된 음악들 중 하나인지의 여부를 판별하는 기능. 본 연구에서는 업로드 하고자 하는 음악이 등록된 음악들 중 하나라고 가정하고, 등록된 곡들 중 어느 곡인지를 식별하는 효율적인 방법 구현에 초점을 맞춘다.

기존의 음악 식별 시스템에서는 음악에서 추출한 패턴간의 유사성을 비교하는 방법을 사용했다[1], [2]. 이러한 방법은 원본 음악 파일로부터 추출한 패턴과 비교 대상 음악 파일로부터 추출한

1) 서강대학교 eflin@sogang.ac.kr
2) 서강대학교 kimshako@sogang.ac.kr
3) 서강대학교 kimjihwan@sogang.ac.kr, 교신저자

패턴의 유사도를 측정하기 때문에 사람에게 동일한 곡으로 인지 되는 음악 파일이더라도, 샘플링 레이트(sampling rate), 재생 속도 등에서 차이가 있는 파일에 대해서는 저작권 위반 여부를 판별하기가 힘든 단점을 가지고 있다.

구현하고자 하는 음악 검색 시스템에서는 동일한 곡이더라도 사용자들의 취향 또는 요구에 따라 압축률, 시작과 끝 지점의 무음구간 길이, 곡의 진폭 등의 차이를 가지는 다양한 파일들이 존재한다. 따라서 본 논문에서는 원본과 동일한 음악 파일뿐 아니라, 샘플링 레이트 차이 등과 같이 변형된 음악 파일에 대해서도 효과적인 음악 식별 방법을 제안한다. 제안한 방법에서는 '음악 음향 모델'을 구현하기 위해서 '음악 단어', '음악 음소'의 개념을 새로이 정의한다. 이 개념을 바탕으로 연속음성 인식에서 사용되는 HMM(Hidden Markov Model) 기반의 방식을 음악식별에 적용한다. 따라서 제안한 방법은 수치상으로는 다른 패턴이지만, 인지적으로 동일한 음악 파일로부터 추출된 패턴의 정확한 식별이 가능하다. 또한 HMM에 기반을 하고 있기 때문에 시스템에 등록된 음악 파일 개수에 대해 확장이 용이하게 된다.

본 논문의 구성은 다음과 같다. 2장에서는 음악에서 추출한 패턴간의 유사성을 비교하는 기존의 방법을 서술한다. 3장에 음악 단어와 음악 음소에 대해 정의 하고, 본 논문에서 제안한 음악 음향 모델의 생성 방법과 HMM기반의 음악 식별 시스템을 설명한다. 4장에서는 실험 방법과 결과를 기술한다. 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

음악으로부터 정보 추출하여 추출한 정보를 이용하는 시스템과 이들 시스템이 사용한 특징 파라미터(feature parameter)들의 개요가 소개되었다[3]. 사용한 특징 파라미터들은 크게 각 윈도우 별로 추출한 주파수 대역별 에너지, 피치(pitch) 및 MFCC(Mel-Frequency Cepstral Coefficient) 등이 있다. 소개된 시스템들 중 음악 검색 방법으로는 Music Genome Project, Audio Fingerprinting, Query by Humming이 소개되었다.

Music Genome Project는 음악의 특징을 구분지어 주는 고유의 정보를 음악 유전자(gene)로 정의하고, 각 음악에 대한 유전자정보를 전문가들이 판별하여 음악 데이터베이스를 구축하는 프로젝트이다[4]. 이 기술을 응용하고 있는 대표적인 서비스는 Pandora Radio[5]이다. Pandora Radio에서는 하나의 곡을 나타내기 위해 팝, 락에 대해서 약150개, 랩에 대해서 약350개, 재즈에 대해서 약400개, 기타 장르에 대해서는 약500개의 유전자를 정의했다. 유전자에 해당하는 특징의 예는 가수의 성별, 사용된 악기, 곡의 느낌 등이 있다. 각 유전자의 값은 1에서 5사이에 0.5 단위의 수로 음악 전문가에 의해서 결정된다. 즉, 모든 곡을 음악 전문가가 듣고, 장르를 파악한 다음, 장르에 해당하는 유전

자의 값을 전문가가 기입한다. Pandora Radio에서는 이러한 음악 유전자 정보를 활용하여 사용자가 음악의 특성들을 입력하였을 경우, 해당하는 음악을 검색하여준다. 그러나 Music Genome Project에서는 특징 추출에 있어서 자동 방식이 아닌 수동 방식을 사용했기 때문에, 음악 유전자 정보의 데이터베이스로 구축하는 비용이 매우 높은 단점을 가지게 된다.

Audio Fingerprinting을 이용한 음악 검색 방법이 제안되었다 [1], [2]. 이 방법에서는 윈도우의 길이는 0.37sec이고, 윈도우를 11.6ms씩 이동하여 각 윈도우 별로 sub-fingerprint를 생성한다. 생성된 sub-fingerprint를 256개씩 그룹핑하여 하나의 fingerprint를 생성한다. 따라서 하나의 fingerprint의 길이는 3초가 된다. 이 fingerprint들에 대해서 데이터베이스를 구축한다. 각 sub-fingerprint는 32비트로 구성이 되는데 300~2000Hz 사이의 대역에서 중복되지 않는 33개의 주파수 대역의 에너지들 간의 차이로 0 또는 1값이 부여된다. 테스트 시 입력된 sub-fingerprint와 데이터베이스 내에 해당 되는 sub-fingerprint가 비트 수준까지 거의 동일한 경우, 찾고자 하는 fingerprint를 효율적으로 찾아준다. 그러나 곡의 속도 변환과 같이 fingerprint의 시작과 끝 지점이 제대로 정합되지 않은 경우에는 sub-fingerprint간의 불일치가 크게 발생하여 원하는 fingerprint를 찾기 힘들어 진다. 본 논문에서 구현하고자 하는 음악 검색 시스템에서는 동일한 곡이더라도 사용자들의 취향 또는 요구에 따라 압축률, 시작과 끝 지점의 무음구간 길이, 곡의 진폭 등의 차이를 가지는 다양한 파일들이 존재 한다. 이러한 환경에서는 sub-fingerprint의 비트 수준의 값이 차이가 많이 발생하기 때문에 Audio Fingerprinting을 이용하기에 한계가 있다. 또한 구현하고자 하는 시스템은 곡 단위로 인식 되는데, 비트 수준에서 차이가 많이 발생하는 경우 Audio Fingerprinting을 곡 단위 인식으로 확장하기 힘든 단점이 있다.

Humming을 이용하여 음악을 검색하는 연구가 진행되었다 [6], [7]. Humming을 입력으로 받아 A/D(Analog to Digital)변환 후, autocorrelation등을 이용해 피치를 측정한다. 피치의 변화를 스크립트로 표현하여 데이터베이스에 저장되어 있는 음악의 피치 스크립트들과 비교하여 가장 비슷한 피치 스크립트를 가지는 노래를 검색 결과로 제공한다. 그러나 본 논문에서 구현하고자 하는 시스템에서는 각 곡에 배경음악, 악기, 보컬 등이 섞여 있는 음악으로부터 피치를 정확하게 측정하는 것이 힘들게 때문에 humming을 이용한 검색을 적용하기 힘들다.

3. HMM기반의 음악 식별 시스템

기존의 방식은 압축률 차이, 곡의 속도 변화와 같이 비트 수준에서의 차이가 있는 경우 인식이 잘되지 않고 또한 곡 단위의 음악 식별 시스템으로의 확장에 어려움이 있다.

본 장에서는 HMM을 이용한 연속음성인식 기법에 기반하여 음악 식별 시스템 구현한다. HMM은 음향모델 구현에서 GMM

(Gaussian Mixture Model)등을 활용하는 경우 특징 벡터의 차이에 대해 강인하고, 시스템의 확장이 용이한 장점을 가지고 있어 연속 음성 인식에 널리 사용되고 있다.

제안하는 방법에서의 인식 기본 단위로 음악 단어(music word)와 음악 음소(music phoneme)를 정의 하고, 이에 대해 생성된 음악 음향 모델을 이용하여 음악 식별 시스템을 구현한다. 3.1절에서는 특징 추출에 대해서 설명한다. 3.2절에서는 음악 단어와 음악 음소를 정의한다. 3.3절에서는 음악 음향 모델 생성 방법에 대해서 기술 한다. 3.4절에서는 생성한 모델로부터 테스트자료를 탐색(search)하는 방법에 대해서 논한다.

3.1 특징 추출

하나의 음악을 20ms길이의 윈도우를 10ms씩 이동하면서 각 윈도우 별로 12차의 MFCC와 1차의 에너지 값을 추출 한다. 그 후 위의 13차 벡터에 대해서 이전 윈도우에서의 해당 13차 벡터와의 차이인 13차 delta 벡터를 구하고, 같은 방법으로 delta 벡터의 차이인 13차 acceleration 벡터를 구해서 총 39차의 벡터를 생성한다.

3.2 음악 단어와 음악 음소

본 연구에서는 인지적인 측면에서 사람이 들었을 때 동일한 특성을 가지는 음악 단위를 음악 음소로 정의 한다. 예를 들어 '도~레~미~'로 구성된 음악에서 '도', '레', '미' 각 음은 각기 다른 특성을 가지므로, 해당 음악을 세 개의 음악 음소의 연결로서 모델링 가능하다. 이 정의에 따라, 동일 음악 음소에 정렬되는 특징 벡터들은 벡터 공간상에서 근접화가 이루어져야 한다. 음악 음소는 음향 모델 학습의 기본 단위로, 각 음악 음소 별로 음향 모델을 보유하게 된다.

음악 단어는 식별의 대상이 되는 음악들에서 연속해서 나오는 음악 음소의 열로, 음악 식별 시스템의 인식 네트워크 상에서 인식의 기본 단위로 정의한다. 가요등의 전주에서는 유사한 비트가 반복이 되는 경우가 많다. '쿵~짝~쿵~짝~쿵~짝~'과 같이 서로 다른 두개의 음악 음소가 계속 반복되는 경우 '쿵~'과 '짝~' 이들 두 음소로 하나의 음악 단어를 구성하고, 이 음악 단어가 반복 되는 것으로 인식 네트워크 구성이 가능하다. 이 경우, 인식 네트워크 구성이 단순해지고, 다른 곡에서 같은 음악 단어를 사용하는 경우 음향 모델의 공유가 가능해지게 되어 메모리 및 식별 속도 향상에 도움이 된다. 또한, 연속음성인식에서 사용되는 인식 프레임워크를 그대로 활용할 수 있게 되는 장점이 있다. 곡의 수가 늘어남에 따라 음향 모델의 수가 늘어나게 되어, 식별시 음성인식과 유사하게 삽입(insertion)이 많이 발생하게 된다. 음악 단어 기반으로 인식 네트워크를 만든 경우, 삽입 벌칙(insertion penalty)을 이용해서 의도하지 않은 음향 모델과의 정합의 수를 줄일 수 있다.

3.3 음악 음향 모델

일반적으로 가요 한곡은 4분 내외의 길이를 가지고 있다. 음악 한 곡이 4분으로 구성되어 있고, 각 윈도우별로 추출한 39차 벡터값이 t 초 단위로 군집화 된다고 가정했을 때, 해당곡을 $4 \times 60/t$ 개의 음악 음소를 가지는 left-to-right HMM 음악 음향 모델로 모델링 할 수 있다. 음악은 음악을 구성하는 각 음간의 시간의 전후 관계가 명확하므로 left-to-right 모델을 적용하는 것이 적합하다. <그림1>은 4분 길이의 음악을 240개의 음악 음소 음향 모델로 모델링 한 결과를 보여준다.

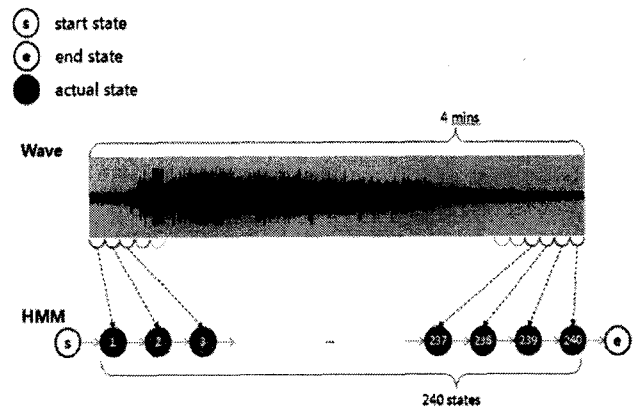


그림 1. 음악 음향 모델의 예제
Figure 1. Example of music acoustic model

p 개의 음악 음소를 가지는 음악 음향 모델의 경우에 대해 초기 모델 생성 방법은 다음과 같다. 해당 음악을 동일한 길이의 p 개의 조각으로 나눈다. 각 조각별로 39차 벡터들의 평균과 분산값을 구한다. 이들 평균과 분산값으로부터 초기 GMM을 생성한다. 이 초기 모델로부터 해당 음악에 대해 maximum likelihood 방법으로 각 음악 음향 모델을 학습한다. 즉, 초기 모델을 이용해서 해당 음악에 대해서 likelihood가 최대가 되도록 auto segmentation을 수행하고, auto segmentation 결과에 따라 음악을 p 개의 조각들로 나눈 후, 각 조각별로 39차 벡터들의 평균과 분산값을 구하고, 이들 평균과 분산값을 이용하여 모델을 업데이트 한다. 이 과정을 likelihood가 증가되는 동안 계속해서 반복한다.

3.4 탐색(Search)

식별 가능한 m 개의 음악에 대한 입력된 음악의 식별은 인식 네트워크상에서 이루어진다. <그림2>는 인식 네트워크 구성방법을 보여준다. m 개의 음악 음향 모델들의 첫 번째 상태(state) 들을 연결하는 더미 상태(dummy state)인 시작 상태(start state)를 생성한다. 이 때, 시작 상태에서 각 음악 음향 모델의 첫 번째 상태로 이동이 가능하도록 상태 전이(state transition)들을 설정한다. 같은 방식으로 마지막 상태들을 연결하여 주는 더미 상태인 최종 상태(end state)를 생성 한 후 음악 음향 모델의 마지막 상태들로부터 최종 상태로 이동이 가능하도록 상태 전이들을 설정한다.

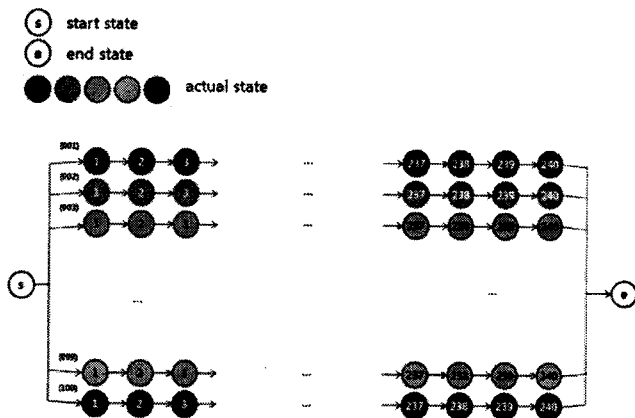


그림 2. 음악 식별 시스템에서의 인식 네트워크

Figure 2. Recognition network of music identification system

인식 네트워크를 이용하여 입력된 음악에 대해 음악 식별되는 탐색 과정은 다음과 같다. 입력된 음악의 각 윈도우 별로 39차의 벡터를 생성한다. 이들 벡터 열에 대해서, Viterbi 탐색 방법을 이용하여 인식 네트워크로부터 가장 최적의 상태열을 찾아 이에 해당하는 음악을 식별 결과로써 생성한다.

4. 실험 및 결과

본 연구에서는 음악 식별 시스템을 만들기 위해 1,000개의 MP3파일들을 수집하였다. 이들 1,000곡은 최근 10년 이내에 출시된 한국 인기가요들로 구성되어 있으며, 128Kbps의 압축률로 녹음된 고품질 MP3 파일들이다. <표1>은 이들 1,000곡에 대한 요약 정보이다.

표 1. 학습 자료의 구성
Table 1. Training data set

항목	내용
곡 수	1,000
파일 형태	MP3
비트레이트	128Kbps
선정 기준	최근 10년 이내에 출시된 한국 인기가요

음성인식에서 널리 사용되는 HTK[8]를 이용하여 음악 식별 시스템을 구현하였다. Stereo로 녹음 되어있는 MP3파일을 16K, mono(left) PCM파일로 변환 하고, 20ms 길이의 윈도우에 대해서 10ms씩 이동하는 각 윈도우에 대해서 39차 특징 벡터를 생성하였다.

수집된 1,000곡의 평균 길이가 228초(약 4분)인 점을 감안하고, 1초 또는 0.5초 단위로 39차 특징 벡터 값의 군집화 되는 영역이 달라진다고 가정하여 각 음악을 240개 또는 480개의 음악 음소로 구성하였다. 또한 각 음악 음소의 정렬되는 구간 별로 39차 특징 벡터 값이 하나 또는 두 개로 군집화 된다고 가

정하고 mixture의 개수를 1개 또는 2개로 달리하여 음악 음향 모델을 구성하였다. <표2>는 이들 음악 음향 모델의 종류를 정리한 내용이다.

표 2. 음악 음향 모델의 종류
Table 2. Classification of music acoustic models

모델 이름	곡당 음악 음소 개수	음소 모델별 Mixture 개수
240p_1m	240개	1
240p_2m	240개	2
480p_1m	480개	1
480p_2m	480개	2

각 음악 음향 모델별의 학습 방법은 다음과 같다. 음악들을 동일한 길이의 곡당 음악 음소 개수(240개 또는 480개)의 조각으로 나눈다. 각 조각별로 39차 벡터들의 평균과 분산 값으로부터 초기 GMM을 생성한다. 이 초기 모델로부터 해당 음악에 대해 maximum likelihood 방법으로 각 음악 음향 모델을 학습한다.

학습된 1,000개의 음악 음향 모델로부터 3.4장에서 기술한 방식에 따라 인식 네트워크를 구성한다. 이때 각 음악 음소를 하나의 음악 단어에 대응 시키고, 삽입 벌칙은 0으로, 빔 사이즈(beam size)는 500으로 설정 하였다. 이 인식 네트워크를 바탕으로, 1,000곡에 대해서 Viterbi 알고리즘에 기반하여 테스트를 수행하였다. 동일한 곡이라도 사용자들의 취향 또는 요구에 따라 비트레이트 등의 차이를 가지는 다양한 파일들이 존재한다. 따라서 원본과 동일한 음악 파일뿐 아니라, 음악 파형 수준에서 변형된 음악에 대해서도 변형에 강인한 정도를 측정한다. 이 변형 정도를 정량화하기 위해서 원본 음악 파일의 음악 속도를 +2%, +5%, -2%, -5%로 변형하고 비트레이트를 32Kbps, 64Kbps로 변환하여 테스트 하였다.4)

<표3>은 각 음악 음향 모델 별 실험 조건 대한 인식 결과를 보여주고 있다. 먼저 원곡에 대한 인식률은 네 가지 음악 음향 모델 모두에서 100% 수준의 인식률을 보여주었다. 음악 속도 변형에 대해서는 240p_1m(음악음소 개수 240, 상태 당 mixture 개수 1개)모델에 대해 2%의 속도 변형에 평균 98.7%, 5%의 속도 변형에 대해 평균 98.4%의 인식률을 보여주었다. 음악 음소의 개수를 두 배로 늘리거나, 상태 당 mixture의 개수를 두 배로 증가시킴으로써 100%에 가까운 인식 결과를 얻었다. 비트레이트가 64Kbps인 경우 240p_1m 모델에서 98.0%의 인식률을 보였다. 이 경우에도 음악 음소의 개수를 2배로 증가 하거나, 상태 당 mixture의 개수를 두 배로 증가시킴으로써 100%에 가까운 인식 결과를 얻었다. 비트레이트가 32Kbps인 경우, 240p_1m 모델의 경우 약 54%, 240p_2m 이나 480p_1m의 경우에는 약 69% 내외, 480p_2m의 경우에는 약 81%의 식별률을 보였다. 그

4) 원음에 가까운 음질을 보장하기 위해서 일반적으로 128Kbps 이상의 비트레이트로 저장한다.

러나 32Kbps의 음질이 전화 음질 수준임을 고려했을 때, 파일 공유 시스템에서는 거의 사용 되지 않는 비트레이트이다. 따라서 1,000곡 수준의 파일 공유 시스템에서는 음악 1초당 1개의 음악 음소를 정의 하고, 상태 당 2개의 mixture로 모델링 하거나, 또는 음악 0.5초당 1개의 음악 음소를 정의하는 경우 100% 수준의 시스템의 구현이 가능하다.

표 3. 각 음악 음향 모델별 실험 조건에 대한 실험결과 (MP3-64, MP3-32는 64Kbps, 32Kbps로 녹음된 MP3파일을 사용.원곡은 128Kbps로 녹음되었음)

Table 3. Experimental results of music acoustic models for different experimental conditions (original music: 128Kbps bit rate, MP3-64: 64Kbps bit rate, MP3-32: 32Kbps bit rate)

실험 조건	인식률 (%)			
	240p_1m	240p_2m	480p_1m	480p_2m
원 곡	99.9	99.9	100	100
음악 속도 +2%	98.7	99.7	99.7	99.7
음악 속도 -2%	98.7	99.9	100	100
음악 속도 +5%	98.3	99.7	99.7	99.7
음악 속도 -5%	98.5	99.8	99.8	99.9
MP3-64	98.0	99.7	99.6	99.7
MP3-32	53.6	67.8	69.5	80.9

5. 결론 및 향후 연구

본 논문에서는 파일 공유 시스템에서의 음악 저작권 보호를 위해서 확장성이 뛰어나고, 수치상으로는 다른 패턴이지만 인지적으로 동일한 음악 파일을 정확하게 인식하는 음악 식별 시스템을 제안하고 검증하였다. 음악 음향 모델을 구현하기 위해서 음악 단어, 음악 음소의 개념을 새로이 정의하고, 이 개념을 바탕으로 HMM기반의 음악 식별 시스템을 구현하였다. 1,000곡에 대한 음악 식별 실험 결과 음악 1초당 두 개의 mixture를 가지는 한 개의 음악 음소로 모델링 하거나, 또는 음악 0.5초당 한 개의 음악 음소를 정의 하여 100% 수준의 인식률을 보여주는 시스템 구현이 가능함을 보였다.

서버 상에 구현된 1,000곡에 대한 식별 시스템에서는 약 4분 길이의 입력 음악의 식별에 대해 약 5초의 시간이 소요되었다. 그러나 파일 공유 시스템에서는 수 만곡 수준에서 식별이 가능하여야 하고, 많은 수의 동시 접속자 수를 고려하였을 때, 인식 속도의 개선이 추후 필요하다. 또한 본 연구에서는 각 음악에 대해 1초 또는 0.5초 단위로 특징 벡터의 군집화 특성이 달라진다고 가정하여 이 단위로 음악 음소의 개수를 정의하였다. 추후에는 인식률과 인식 속도 향상을 위하여 각 곡의 특성에 맞는 음악 음소의 정의를 찾고자 한다.

감사의 글

시스템 구현과 자료 수집에 많은 도움을 준 서강대학교 장병준, 최병준, 김광호, 임민규, 명진복, 반총희님께 감사의 뜻을 전합니다.

참고 문헌

- [1] J. Haitsma, T. Kalker, (2002). "A highly robust audio fingerprinting system", *Proc. International Conference on Music Information Retrieval (ISMIR)*, pp.144-148.
- [2] M. Miller, M. Rodriguez, I. Cox, (2002). "Audio fingerprinting: nearest neighbor search in high dimensional binary spaces", *Proc. IEEE International Workshop on Multimedia Signal Processing*, pp.182-185.
- [3] R. Typke, F. Wiering, R. Veltkamp, (2005). "A survey of music information retrieval systems", *Proc. International Conference on Music Information Retrieval*, pp.153-160.
- [4] W. Glaser, T. Westergren, J. Stearns, J. Kraft, "Consumer item matching method and system", U.S. Patent No. 7003515
- [5] T. Westergren, (2007). "The music genome project", www.pandora.com/mgp.shtml.
- [6] A. Ghias, J. Logan, D. Chamberlin, B. Smith, (1995). "Query by humming - musical information retrieval in an audio database", *Proc. ACM Multimedia*, pp.231-236.
- [7] E. Unal, S. Narayanan, E. Chew, (2004). "A statistical approach to retrieval under user-dependent uncertainty in query-by-humming systems", *Proc. ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp.113-118.
- [8] S. Young, G. Evermann, M. Gales et al., (2006). *The HTK book(for HTK version 3.4)*, Cambridge University.

• 김희동 (Kim, Hee-Dong)

서강대학교 컴퓨터공학과
서울시 마포구 신수동 1번지
Tel: 02-715-2715 Fax: 02-704-8273
Email: efline@sogang.ac.kr
관심분야: 음성인식, 멀티미디어 검색
현재 KAIST 전자전산학부 전산학전공 석사과정 재학중

• 김도현 (Kim, Do-Hyun)

서강대학교 컴퓨터공학과
서울시 마포구 신수동 1번지
Tel: 02-715-2715 Fax: 02-704-8273
Email: kimshako@sogang.ac.kr
관심분야: 음성인식, 멀티미디어 검색
현재 LG전자 재직중

• 김지환 (Kim, Ji-Hwan) 교신저자

서강대학교 컴퓨터공학과
서울시 마포구 신수동 1번지
Tel: 02-715-2715 Fax: 02-704-8273
Email: kimjihwan@sogang.ac.kr
관심분야: 멀티미디어검색, 대화이해, 임베디드용 음성인식
2007~현재 서강대학교 컴퓨터공학과 조교수