韓國컴퓨터情報學會 論文誌
第15卷 第11號, 2010. 11.

2010-15-11-2-5

온라인 음악 관리 서비스를 위한 사용자 음원 인식 시스템 개발

성 보 경*, 고 일 주**

Development of User Music Recognition System For Online Music Management Service

Bokyung Sung*, Ilju Ko**

요 약

최근 디지털 콘텐츠 서비스 분야에서 사용자 맞춤형 서비스를 위해 사용자 자원 인식의 필요성이 대두되고 있다. 특히 온라인 기반 음악 서비스의 경우 사용자 취향 분석, 음원 추천 및 음악 관련 정보 제공을 위해 사용자 음원 인식 기술이 요구되고 있다. 현재 태그정보를 기초로 사용자 음원 인식 후 음악 관련 정보를 제공하는 서비스가 제공되고 있지만, 태그정보의 변조 및 삭제 등의 취약점으로 인식 오류가 급증하고 있다. 이러한 문제의 보완 방안으로 음악 자체를 이용하는 내용기반 사용자 음원 인식 기법에 대한 연구가 이루어지고 있다. 본 논문에서는 음악의 파형에서 추출된 특징 정보를 기초로 온라인상에서 사용자 음원을 인식하는 방법에 대해 논하고자 한다. 사용자 음원의 내용기반 인식을 위해 구조에 적합한 음원의 전처리 후 특징 추출을 하였다. 추출된 특징은 음악 서버에 특징 형태로 저장된 음원과의 매칭 과정을 통한 인식을 진행하여 태그데이터에 독립적으로 사용자 음원을 인식할 수 있게 되었다. 제안된 사용자 음원 인식 방법의 검증을 위해 600개의 음악을 무작위 선정하고, 각각을 5가지 음질로 변화하였다. 이렇게 생성된 3000개의 실험음원을 30만곡을 포함하는 음악 서버를 기준으로 인식실험을 진행하였다. 평균 인식율은 85%를 나타내었다. 제안하는 내용기반 음원 인식을 통하여 태그기반 음원 인식의 취약점에 대한 극복을 하였으며, 음원 인식의 성능은 실제 온라인 음악 서비스에 적용할 가능성을 보여주었다.

Abstract

Recently, recognizing user resource for personalized service has been needed in digital content service fields. Especially, to analyze user taste, recommend music and service music related information need recognition of user music file in case of online music service. Music related information service is offered through recognizing user music based on tag information. Recognition error has grown by weak points like changing and removing of tag information. Techniques of

• 제1저자 : 성보경 교신저자 : 고일주
• 투고일 : 2010. 06. 12, 심사일 : 2010. 08. 06, 게재확정일 : 2010. 09. 09.
* 숭실대학교 미디어학과 박사과정 ** 숭실대학교 IT대학 글로벌미디어학부 부교수
※ 이 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임
(KRF-2008-331-D00401)

content based user music recognition with music signal itself are researched for solving upper problems. In this paper, we propose user music recognition on the internet by extracted feature from music signal. Features are extracted after suitable preprocessing for structure of content based user music recognition. Recognizing on music server consist of feature form are progressed with extracted feature. Through this, user music can be recognized independently of tag data. 600 music was collected and converted to each 5 music qualities for proving of proposed recognition. Converted 3000 experiment music on this method is used for recognition experiment on music server including 300,000 music. Average of recognition ratio was 85%. Weak points of tag based music recognition were overcome through proposed content based music recognition. Recognition performance of proposed method show a possibility that can be adapt to online music service in practice.

▸ Keyword : 내용기반 음원인식(Content Based Music Recognition), 사용자 음원(User Music), 온라인 음악서비스(Online Music Service)

Ⅰ. 서 론

온라인 음악서비스의 출현은 '콘텐츠의 디지털화'를 통해 '음반의 판매'를 '디지털 파일 형태의 음악을 네트워크를 통해 소비자에게 판매'하는 형태로 변화 시켰다. 이는 OSMU(One Source Multi Use)를 지향하는 디지털 콘텐츠 서비스의 초기변화로서, 네트워크 기술이 융합된 다양한 디바이스를 통한 고도화된 서비스가 요구되는 시점이다. 최근 디지털 음악 서비스는 자동화 형태의 '개인화된 서비스'라는 방향으로 변화하고 있다. 이러한 예로 사용자에게 음원에 관련된 정보를 제공하거나 음원의 수집부터 재생까지 일관적으로 관리하는 서비스를 들 수 있다. 이에 필요한 핵심 기술로 사용자의 음원을 인식하는 기술이 요구 되고 있다.

사용자 음원 인식을 위한 기술은 크게 두 갈래로 구분 될 수 있다. 현재 가장 많이 활용되고 있는 '태그기반 사용자 음원 인식'방법이 있다. 이 인식방법은 키워드 기반 검색 기술의 차용으로 이루어 졌다. 음악 콘텐츠는 텍스트 데이터와는 다르게 내용자체를 명확하게 수치적으로 표현할 수 있는 방법이 없다. 그래서 그동안 디지털 음악의 정보는 일괄적인 전처리 과정을 통해 생산 되고, 태그형태로 저장되었다. 콘텐츠는 태그 정보를 이용하여 검색 혹은 관리 되었다. 이러한 인식방법은 지속적인 활용에 한계를 나타내는 문제를 나타내고 있다. 첫째, 사용자들은 정형화된 형태의 태그정보 보다 실제 음악의 내용에 더 관심을 가지고 있지만 태그에 기록되어져 있는 정보 이외의 것에 대한 검색이 어렵다. 둘째, 일반 개인 사용자들이 자체적인 기준으로 음원을 생산하기 시작하면서 모든 디지털 음악의 태그정보가 올바르게 저장되지 않을 가능성이 증가하고 있다.

위와 같은 단점의 극복을 위해 내용기반 음원인식에 대한 연구가 진행되고 있으며, 여러 환경에 적용되고 있다. 내용기반 음원인식은 음원의 음악 신호 영역에서 미리 정의 되지 않은 음악정보를 추출하고 가공하여 인식하는 방법이다. 이 방법은 과거 한정된 음원 제공자에 의해 유통되던 구조에서는 좋은 성능을 보였다. 현재는 다양한 음악 재생 기기들로 인해 사용되는 음질규격이 다양해 졌다. 이로 인해 유통되는 많은 음원이 서로 상이한 규격을 지니고 있는 형태이다. 내용기반 음악 검색 분야의 주된 이슈중 하나는 다양한 음질 규격을 고려해야 한다는 것이다. 음악 디바이스 및 사용목적의 다양화로 유통되는 많은 음악들이 일관된 규격을 지니지 않게 되었다. 이와 같은 한계의 극복을 위해 디지털 음악 콘텐츠의 음악 파형 자체를 통한 내용기반 음원 인식 기법이 필요하다.

본 논문은 다양한 음질 규격의 음악 콘텐츠를 위한 온라인 음원 관리 서비스의 핵심 기술로 요구되는 온라인 사용자 음원인식 시스템을 제안을 한다. 제안하는 시스템은 음원 음질의 변화에 강인함을 나타내기 위한 음원의 전처리 및 특징을 추출하는 부분과, 중복되는 인식작업 제거를 포함하는 인식구조로 구성된다.

인식을 위해 입력되는 음원들 중 동일한 음악일지라도 서로 다른 음질의 파일일 가능성이 높다. 이런 경우 스트링 매칭에 보편적으로 활용되는 해시 기술만으로는 음악의 인식이 어렵다. 음질 변화에 관계없이 일관성을 나타내는 특징추출을 위해 적합한 전처리 과정을 거쳐야 한다. 음원인식에 필요한 만큼의 음악파형을 취한다. 전체 파형을 사용할 경우 발생되

는 불필요한 연산량을 피할 수 있다. 다양한 음질에 대한 인식이 목표 이므로 확률적으로 가장적은 오류를 낼 수 있는 기준 음질을 설정하고 취해진 음악을 기준 음질로 변환한다. 압축 오디오를 비압축 형태의 PCM 으로 변환하고 파형을 정규화 한다. 인코딩 환경에 따라 가변적으로 삽입되는 음악 앞부분의 노이즈 및 무음영역을 제거한다. 음질변화에 관계없이 일관성을 나타내는 특징추출을 위해 Delta-Grouped MFCCs (DG-MFCCs) 를 사용한다. 이 특징은 MFCCs를 일정한 단위 길이로 그룹화를 통해 생성된 축약된 특징열을 1차 미분한 것이다. 이러한 두 과정은 각각 차원감소를 통한 단순화 및 변화경향을 강조하는 효과를 준다. 이 방식을 통해 MFCCs의 단점인 음질 변화에 민감하다는 것과 데이터의 개수가 많다는 점을 극복하였다.

음원 인식 과정은 크게 기존에 인식과정을 거쳤는지 확인하는 과정과 검색을 통한 인식이다. 인식과정에 선행적으로 입력음악의 해시코드를 추출하고 기존 인식콘텐츠들의 해시코드 테이블과 비교하여 이전에 인식한 것인지 확인한다. 이러한 과정의 추가를 통해 중복 인식 과정을 제거하여 시스템의 효율성을 높였다.

II. 기존 연구

음악의 소비가 단순한 청취에서 온라인을 통한 다목적 배포가 되면서 사용자가 보유한 음원을 인식하는 기술의 중요성이 점차 부각되고 있다. 현재 음원인식은 크게 2가지 방향으로 방법론이 연구 및 적용되고 있다. 태그기반 음원인식, 내용기반 음원인식이 그것이다.

태그기반 음원 인식은 음악파일에 삽입된 태그 정보를 기초로 한 인식 방법이다. 음악파일은 그 음악에 관련된 텍스트 정보와 음악파형 정보로 구성된다. 이것의 보편적인 예로서, MP3 파일은 ID3태그 규격으로 관련된 텍스트 정보를 저장하고 있다.[1] 태그정보는 음악의 객관적인 정보로 제목, 가수, 앨범명, 발매연도, 장르, 원곡가수, 저작권, 매체종류 등이 있다. 이외에도 가사, 이미지, 볼륨, 잔향 설정 등과 같이 확장된 정보도 저장할 수 있도록 규격화되어있다. 이러한 규격화된 정보를 통해 음원의 인식 및 검색이 가능하다. 일부의 정보가 누락되었다 하더라도 텍스트 정보 항목들의 조합을 통해 인식이 가능하다. 그러나 태그정보가 완전히 삭제되거나 실제 파형정보가 태그정보와 다른 음원일수도 있다. 태그기반 인식모듈을 통한 인식 테스트에서 태그정보가 완전히 삭제된 샘플의 경우는 인식판정이 되지 않고, 음원과 다른 태그정보가 삽입된 경우는 False Positive(긍정오류) 판정으로 인식오류가 발생됨을 나타낸다. 저작권 보호조치를 위한 초기의 불법 디지털 음원 검출 알고리즘의 원리가 음원의 태그정보를 활용하는 방식이었다. 이후 완전히 다른 태그정보 혹은 태그정보 전체를 삭제하고 음원을 유통하는 경우가 크게 증가하였다. 이러한 경우에는 태그기반으로는 음원인식이 불가능 하다는 단점이 지적되고 있다.

내용기반 음원 인식은 내용적 변형 가능성이 거의 없는 음악파형 자체만을 이용한 인식 방법이다. 태그정보와 다르게 파형정보는 정보검색에 적합한 형태로 가공하는 방법이 다양하게 존재한다. 분석을 위한 연산시간 및 정확도 등이 온라인 음원 유통 환경의 변화에 민감하게 변화하는 이유로 내용기반 음원 인식은 현재도 활발히 연구되고 있다.

내용기반 음원 인식에서의 핵심은 파형으로부터 인식성능이 좋은 특징을 추출하는 것이다. 현재 연구되는 대표적인 특징 추출 알고리즘은 분석 기반의 관점에서 크게 3가지로 분류된다. 3가지 분류는 시간 기반 특징, 주파수 기반 특징, 혼합 특징 이다.[2] 시간 도메인의 특징의 대표적인 예로 Zero Crossing Rate(ZCR)[3], Root Mean Square(RMS)[4] 가 있다. 주파수 도메인 기반의 특징추출 방법으로는 Linear Predictive Coding(LPC)[5-6], Mel-Frequency Cepstrum Coefficient(MFCC)[7-11], Spectral Centroid[12], Spectral Flux[13], Spectral Roll-off[14] 이 있다. 최근에는 각각의 장단점 보안을 위해 위의 방식들을 혼합한 형태의 특징추출 알고리즘에 대한 연구도 진행되고 있다. 최근에는 각각의 장단점 보안을 위해 위의 방식들을 혼합한 형태의 특징추출 알고리즘에 대한 연구도 진행되고 있다.

최근까지 대부분의 인식 모듈은 태그 기반으로 사용자의 음악 콘텐츠의 태그 정보를 활용한다. 고도화된 서비스와 자동화를 위해서는 태그 정보가 아닌 음원 자체 정보를 가지고 인식 하여야 한다. 잘못 정제된 태그 정보로 인해 발생되는 인식의 오류를 최소화 시킬 수 있기 때문이다. 다변화 하는 온라인 환경 내에서 인식정확도 측면에서도 내용기반 인식 방법이 보다 효과적이다. 오류를 발생시킬 만한 요소들에 더 강인한 형태이기 때문이다.[16-17]

III. 사용자 음원 인식을 위한 제안 방법

본 제안은 사용자가 보유한 콘텐츠에 대한 자동적 관리 및 보유한 모바일 디바이스를 통한 원활한 소비가 가능하게 하는 것을 목적으로 한다. 이와 같은 목적은 사용자 음원의 실제 음원파형만으로 음원을 인식하고 그 음원에 적합한 메타 데이터를 제공하여 정확하고 일관적인 규격의 관련 정보의 제공을

위미한다. 또한 효율적인 음원관리를 위해 사용자의 중복된 음원에 대해서는 중복적인 업로드나 저장 공간을 차지하지 않도록 한다.

제안하는 방법은 전체적 구조의 제안, 음악의 전처리 및 특징 추출 그리고 음원의 인식 부분으로 구분하여 나타낸다.

1. 전체적 구조

그림 1.은 제안된 사용자 음원 인식을 위한 제안방법의 구조를 나타낸다. 사용자 음원 인식 시스템의 구조는 크게 3부분의 세부 영역으로 구성된다. 세부영역은 사용자 음원입력, 음원의 전처리 및 특징 추출, 음원의 인식 및 관리서비스이다.

사용자의 음원입력은 해시값 산출 과정과 파싱 과정을 거친다. 해시값은 인식대상의 콘텐츠가 기존에 인식 과정을 거친 콘텐츠 였는지 판단하는 과정에서 요구된다. 이 중복 검수 판단 과정을 통해 중복된 인식과정을 생략할 수 있다. 음원의 파싱 과정은 음원에서 태그영역을 제거한 후 음악 파형만을 취한다. 음악 파형만을 통한 인식은 기존 태그정보의 다양한 변형 및 잘못된 정보로 인하여 발생되는 인식 오류 가능성을 낮춰준다. 또한 일정 영역만의 음원을 인식의 특징 추출에 사용함으로써 빠른 처리 과정을 가능하게 한다.

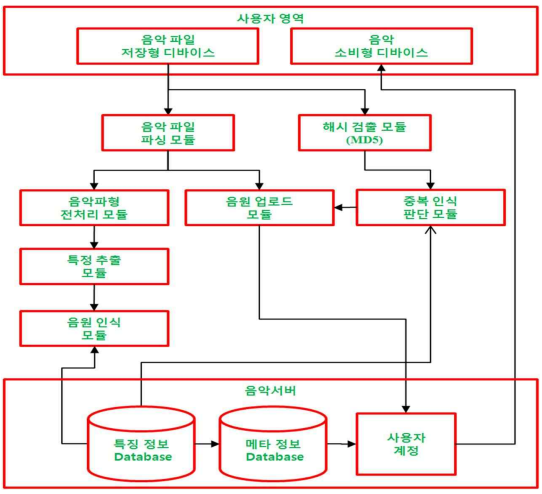

그림 1. 사용자 음원 인식의 전체 구조
Fig 1. System Architecture for user music recognition

전처리 과정은 음악 인식에 적합한 특징인식 성능을 높이는 것을 가장 큰 목적으로 진행한다. 전처리 과정은 특징 추출에 적합한 형태로의 정규화를 진행한다. 정규화를 통해 음악 인식에 영향을 미치는 음질변환, 음악 파일규격 등의 요소들의 영향을 최소화 할 수 있다.

특징 추출은 실시간 처리에 적합할 수 있을 만큼의 데이터 리덕션 및 인식률 유지라는 두 개의 목적을 기초로 구성된다. 음원의 특징을 단순화된 특징값 으로 표현하는 것은 인식과정의 효율성을 만족 시킬 수 있는 형태이다.

음원의 인식은 특징 값으로 변환된 음원간의 Euclidean Distance를 통한 유사도 측정을 통해 진행되며 음악인식 이후 관련된 메타정보를 자동으로 부여하여 사용자 계정을 구성한다.

2. 음악의 전처리 및 특징 추출

음악의 전처리는 5단계 세부 처리 과정의 연속으로 구성된다. 각 세부 처리 과정은 음악파일 분할, 기준음질로의 변환, 디코딩, 정규화, 무음영역제거 이다.

음악파일 분할은 음악파일로부터 원하는 부분만을 취하기 위한 단계이다. 인식을 위해 사용자의 음원의 전체 영역을 사용하지 않는다. 콘텐츠 인식을 위해 전체를 사용할 경우 고도의 인식률을 나타내겠지만, 절대적으로 높은 연산시간이 요구되므로 시스템 구성관점에서는 비효율적이다. 연산의 효율 및 안정적인 온라인 서비스 구조를 위해 최소영역의 음원만을 사용하는 것이 바람직하다. 이를 위해 최소한의 유효영역만 분할 후 사용한다.

분할된 음악파일은 기준음질로 변환하는 과정을 거친다. 사용자의 음원은 샘플링비율, 비트율, 오디오 채널수 등의 다양한 조합으로 인코딩되며, 이것은 음원인식의 오류를 발생시키는 요소로 작용한다. 동일 음악에 대한 인식 오류를 최소화하기 위해 기준음질을 정하고 이 기준으로 변환하여야 한다. 기준음질의 경우는 일반적으로 유통 음원의 음질 규격들의 평균치로 지정할 수 있지만, 평균치는 평균점을 중심으로 음질의 차이 발생시 마다 동일하게 오류요소가 누적될 수 있다. 또한 평균점 이하의 음질은 저음질 에서 고음질로 변환해야 된다. 확률적으로 오류를 최소로 하기위해, 기준음질은 유통되는 음원 규격의 최소 규격으로 하여야 한다. 최소 규격의 경우 최종 음질 변환후 정보손실을 가져오지 않기 때문이다. 본 논문에서는 온라인에서 유통되는 일반적인 최소 음질 음원 규격인 mp3 64kbps 을 기준음질로 정하였다.

복호화 과정은 압축형태의 음악으로부터 비압축 파형을 나타내는 PCM 으로 변환하는 과정이다. 본 논문에서는 44100Hz 샘플링, 16bit 양자화, Mono 채널을 기준 PCM 규격으로 정하였다.

기준음질로부터 변환된 PCM 데이터는 원본 음원의 볼륨 레벨에 따라 값이 차이가 커진다. 그러나 동일 음원의 경우는 샘플들간의 변화비율은 동일하기 때문에 정규화를 통해 파형

의 절대값 차이를 최소화 한다.

무음 영역 제거는 입력음악의 시작부분에 가변적인 길이로 삽입된 무음 신호 구간을 제거하기 위한 단계이다. 무음 신호는 음악의 디지털화 과정에서 삽입된다. 또한 디지털 컨버팅 과정 때마다 음원의 부분별 길이의 변화가 발생되고 변화가 누적된다. 무음신호 구간 길이의 차이는 음원인식율을 저하시키는 주된 요소중 하나이다. 무음제거를 통해 실제적으로 음악이 시작하는 지점을 찾아 최초 무음 구간에 의해 발생할 수 있는 시간축 기준의 데이터 쉬프팅 현상을 줄일 수 있다. 무음제거 구간은 무음구간에서 유음구간으로 전환되는 형태를 크게 4가지로 유형화하여 이것을 기준으로 전환지점을 결정하는 룰베이스 방식을 적용하였다.[18]

전처리 과정을 거친 음악으로 부터의 특징추출은 MFCC 추출, 단위길이 블록화, Delta 연산의 과정으로 이루어진다. 이 과정을 통해 최종 Delta-Grouped MFCCs (DG-MFCCs)의 특징이 추출된다.

MFCCs 는 인간 청각기관 모델을 기초로 설계된 특징추출 알고리즘의 하나다. 초창기 음성인식 시스템이나 화자인식 시스템 등에서 인식을 위한 특징파라메터로 사용되었다. 이후 잡음에 약한 단점에도 불구하고 높은 인식 성능으로 인해 음악장르인식과 같이 음악정보 검색 분야로 그 활용 분야가 확대 되어져 왔다. 일반적으로 통용되는 표준 MFCCs의 정의는 다음의 과정과 같다. 기준 길이의 Window 크기 와 Hop 크기 로 프레임화 된 시계열의 신호를 주파수계로 전환한다. 주파수 데이터에 상대적으로 인간의 인지능력이 높게 발현되는 1Hz 이하 소리에 대한 분석 밀도를 높여주는 Mel Filter Bank 처리를 한다. 인간의 소리 인지 형태로 신호를 스케일링 하기 위해 로그를 적용한다. 출력노드들 간의 상관관계를 최소화하기 위해 DCT 연산을 수행한다. 본 논문에서는 Window 크기와 Hop 크기를 일반적으로 사용하는 각각 20ms, 10ms 로 하였다. 1~12차 까지 12개의 특징을 사용하기 위해 Mel Filter Bank의 출력 노드를 12개로 설계하였다.[7-11]

하나의 콘텐츠로부터 식1의 연산 과정을 통해 총 80개의 DG-MFCCs특징을 추출 한다. 식1의 연산 과정은 다음과 같은 일련의 과정들의 수식적 표현이다. MFCCs 특징을 블록 단위로 구분하고 각각의 블록내부는 sum 연산한다. 단일 그룹의 크기는 50개 프레임 으로 정하였다. 그룹화된 값들은 Delta 연산을 하여 연속데이터의 1차 미분된 형태의 결과를 도출한다. 그룹화는 직접적으로 연속된 특징데이터의 개수를 줄이며 데이터의 단순화 하는 역할을 하고, Delta 는 단순화된 형태의 패턴을 강조 하는 역할을 한다.

$$DG\_MFCCs = \sum_{n=1}^{50}\sum_{k=1}^{12} MFCC(k)(n)$$

..... (식1)

3. 음원인식

음원 인식은 2단계의 처리과정을 통해 구성된다. 각각은 중복인식 판단, 음악인식 판단이다.

중복 인식과정 판단은 Message-Digest algorithm 5 (MD5) 알고리즘을 통해 추출된 해시값을 이용하여 판단한다. MD5는 IETF RFC 1321에 명시되어 있는 128비트 암호화 해시 함수이다. 규약에 따르면 입력된 어떤 두 개의 메시지가 동일한 메시지 축약을 결과로 내거나, 또는 어떤 메시지 축약을 통해 엉뚱한 메시지가 만들어지는 것은 "계산적으로 불가능"하다고 한다. 이러한 원리를 기초로 음악파일이 원본 그대로인지를 확인하는 무결성 검사에 사용하는 것이다. MD5는 임의의 길이의 메시지(variable-length message)를 입력받아, 128비트짜리 고정 길이의 출력값을 고속 연산을 통해 도출해 낸다. 즉 사용자의 음원이 서버에 있는 음원과 모든 조건이 완전 일치 한다면 동일한 해시값이 생성이 되고 이것의 비교 과정만을 통해 중복되는 음원의 인식 과정과 업로드 과정을 생략할 수 있게 해준다.

파형만을 이용한 인식 판단은 위에서 제시한 DG-MFCCs 특징들간의 Euclidean Distance 비교를 통해서 한다. 인식 판단기준은 입력된 특징열과 유사한 서버내의 1순위 음원과 2순위 음원의 특징 간의 Delta값과, 2순위 음원과 3순위 음원의 특징 간의 Delta값의 비교를 통해 판단한다. 1순위 음원과 2순위 음원의 Delta값과 2순위 음원과 3순위 음원의 Delta값의 차가 상대적으로 크면 클수록 확실한 인식으로 고려된다. 초기 실험 데이터에 의해 조합으로 그 차이 판단에 적합한 임계값을 찾아낸다.

IV. 실험 및 결과

실험에서 사용된 음원은 크게 서버에 저장된 음원, 인식 실험을 위한 음원으로 나뉜다. 음악 서버의 경우는 누적 수집된 30만곡의 음악을 보유한 서버를 사용하였다. 인식 실험을 위한 음악은 2008년1월부터 2010년 1월까지 랜덤하게 선택된 각각 여섯 개월에 새로 발매된 음원 100개, 총 600개의 음악을 사용하였다. 실험 결과는 시기별로 성능의 변화 추이를 분석하기 위해 월 단위로 평균치를 산출 하였다.

사용자 음원인식 실험은 다음과 같다. 인식할 하나의 음원이 선택되면, 앞절에서 정의한 기준을 통해 음악 인식에 필요한 영역만 취한 후 전처리 한다. 전처리된 데이터로부터 연속된 특징열을 추출한다. 최종 산출된 특징열은 특징블록 형태로 저장되어 있는 서버의 모든 음악과 유사도 비교를 한다. 마지막으로 설정된 임계치를 고려하여 음악 서버내에서 가장 높은 유사도를 나타내는 특징 데이터의 음악 파일을 동일 음원으로 인식한다. 음원은 음악관련 디지털 기기들이 주로 사용하는 음질 변화에도 강인함을 보임을 증명하기 위해 총 5가지로 변환하여 총 3000번의 인식 실험을 하였다. 5가지 음질 규격은 192kbps, 160kbsp, 128kbps, 96kbps, 64kbps 로 모두 mp3 포맷으로 실험하였다.

그림 2. 는 무작위로 산출된 여섯 개월의 음원인식 실험의 인식율의 평균을 나타내고 있다. 이 도표는 각 1개월을 1개의 세트로 간주 하여 구분하였으며, 7번째 세트는 세트A에서 세트F까지 6개 실험세트의 평균을 나타낸다. 즉 6개월의 평균을 의미한다. 각 세트 내에서는 동일한 음원 100개를 5개의 음질별로 인식률을 나타내고 있으며 5개 음질 전체에 대한 평균도 나타내고 있다. 예를 들어 세트A의 64kbps 막대그래프는 64kbps 음질의 세트A의 실험음원 100곡의 인식율을 의미한다. Average 는 각 5개 음질 인식률의 평균을 의미한다. 세로축은 인식률을 의미하며 단위는 백분율이다. 각 세트에서 음질 변화에 따른 인식률 사이에 특별한 패턴이 나타나지 않았다. 세트간에 인식율의 차이가 나타나는 경우(세트 C)가 있지만, 전체 실험 결과에 대한 평균을 통해 장기간실험된 평균적인 결과는 음질에 상관없이 일정한 인식률로 수렴하는 것을 보여주고 있다. 인식율의 수렴은 약 85%로 나타나고 있다.

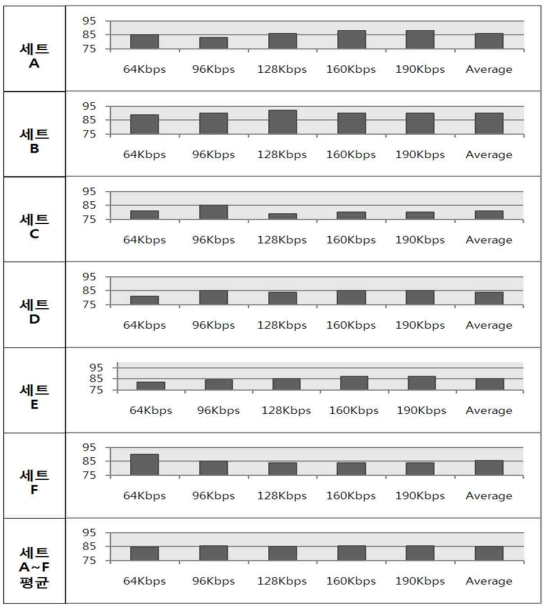

그림 2. 내용기반 사용자 음원인식 시스템의 인식율
Fig 2. Recognition rate of content based user music recognition system

그림 3. 은 음원 인식 속도의 평균을 나타내고 있다. 이 도표는 그림2 와 같은 형식이며 인식에 걸리는 전체 시간을 나타내고 있다. 세로축은 인식에 걸린 시간을 의미하며 단위는 '초'이다. 각 세트에서 음질 변화에 따른 인식속도 변화에서는 기준 음질에 가까운 음원일 수록 근소한 차이로 인식속도가 단축 됨을 나타내는 것 이외에 특별한 패턴이 나타나지 않았다. 전체 실험 결과에 대한 평균을 통해 장기간 실험된 결과들의 평균은 음질에 상관없이 일정한 인식 속도로 수렴하는 것을 보여주고 있다. 인식 속도의 수렴은 약 6.9초로 나타나고 있다.

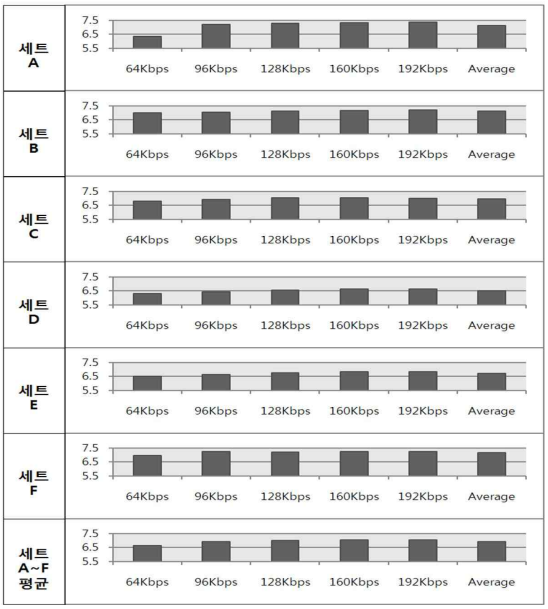

그림 3. 내용기반 사용자 음원인식 시스템의 인식속도
Fig 3. Recognition time of content based user music recognition system

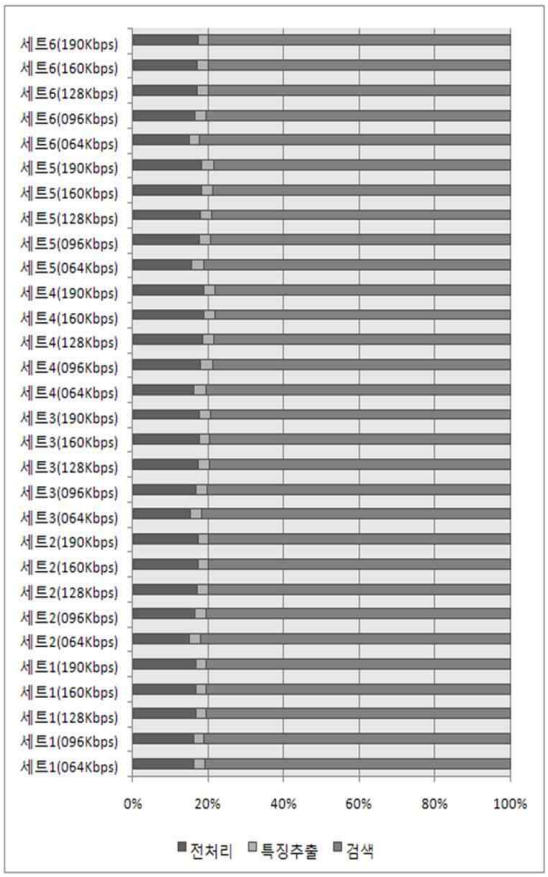

그림 4. 전처리, 특징추출, 검색 간의 연산시간 비율 비교
Fig. 4. Comparing ratio among each operation time of preprocessing, extracting feature and retrieval

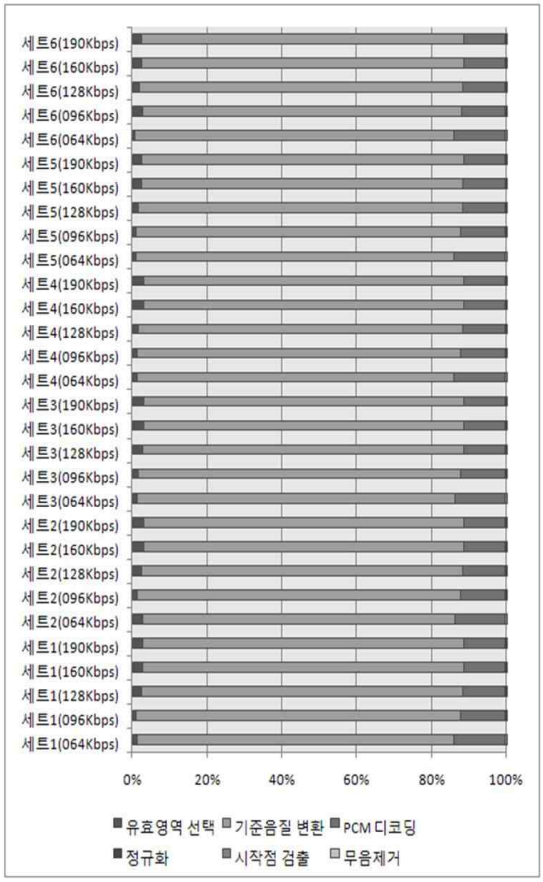

그림 5. 전처리 세부항목간의 연산시간 비율 비교
Fig. 5. Comparing ratio among each operation time of detail steps of preprocessing

표 1. 음원 인식 속도의 예 (세트: F, 곡명: Insomnia, 발매: 2009년 2월, 음악가: 휘성)(단위: Sec)
Table 1. Example for music recognition time (Set: F, Title: Insomnia, Release: Feb of 2009, Artist: WheeSung)

구분 / 음질	전처리						특징추출			검색
	유효영역 선택	기준음질 변환	PCM 디코딩	정규화	시작점 검출	무음제거	MFCC 추출	SUM 연산	Delta 연산	Query 응답
64 Kbps	0.005921	0.880052	0.145831	0.001784	0.000127	0.001336	0.202403	0.001245	0.000001	5.656874
96 Kbps	0.027586	1.017288	0.142115	0.001758	0.000125	0.001295	0.191298	0.001122	0.000001	5.744317
128 Kbps	0.011443	1.067355	0.141628	0.001726	0.000124	0.001341	0.201474	0.001102	0.000001	5.649919
160 Kbps	0.022238	1.069598	0.143675	0.001755	0.000124	0.001199	0.19403	0.001162	0.000001	5.62885
192 Kbps	0.021089	1.109546	0.142146	0.001851	0.00012	0.001374	0.196232	0.001109	0.000001	5.681328

표 1. 은 음악 한곡의 인식 속도 결과치의 예시를 나타내고 있다. 음원 인식 과정은 크게 전처리, 특징 추출, 검색 과정으로 구분되고 각각의 과정은 전장에서 나타낸 세부 기능들로 구성된다. 전처리, 특징 추출 과정에 비해 서버에서 인식 쿼리에 대해 응답하는 검색 과정의 시간이 많이 걸린다. 특징 추출의 경우 음질에 상관없이 거의 일정한 시간이 소요됨을 나타낸다. 전처리의 경우 음질별로 음원의 일정 부분만을 취하는 과정과 기준 음질로 변환 하는 과정에서 기준 음질에 가까울수록 근소하게 속도의 차이가 남을 나타내고 있다. 이와 같은 특징은 실험 세트별로 음원 인식 속도의 평균과도 유사함을 나타낸다.

그림 4. 는 전체 인식 시간에서 전처리, 특징추출, 검색 간의 연산시간의 비율을 나타내고 있다. 본 논문에서 중점을 두고 있는 전처리, 특징 추출 과정은 20% 내외의 비중을 나타내고 있다. 이는 평균 1.5초 이내의 연산시간을 의미한다.

그림 5. 는 음질 및 세트별로 가변적인 속도를 나타내는 전처리 과정의 세부 항목별 연산시간 비율을 나타내고 있다. 전처리 과정에서 mp3 음악의 변환 처리가 다른 항목에 비해 상대적으로 높은 처리 시간을 소요하고 있다. 실제 인식 성능에 영향을 미치는 시작점 검출 및 무음 제거 알고리즘의 경우는 빠른 처리 속도를 나타내고 있다.

V. 결론

본 논문에서는 온라인 음악추천 및 관리 서비스를 위한 사용자 음원 인식 방법을 제안하였다.

실험결과를 통해 각 세트에서 음질변수에 따라 인식과정상에서 특별히 절대적인 패턴이 나타나지 않음을 확인하였다. 전체 실험 결과에 대한 평균을 통해 장기간 실험된 평균적인 결과는 음질에 상관없이 일정한 인식률로 수렴하는 것을 보여주고 있다. 제안하는 방식을 통한 음원인식에서 음원의 전처리 및 특징 추출 과정이 최적화되지 않은 상태이지만 1초대의 수행 시간으로 보아 향후 실시간 서비스에 적합할 만큼의 기술로 발전할 가능성을 보여주고 있다. 인식율에 직접적인 영향을 주지 않지만 인식속도에 영향을 미치는 mp3 디코딩 및 서버Database 검색 시간은 최적화된 프로그래밍을 통해 개선해야할 부분으로 남겨지고 있다.

그럼에도 불구하고 본 논문을 통한 제안한 음원 인식 방법은 사용자의 음악을 온라인 상에서 자동으로 인식할 수 있는 기술로서, 다양한 응용분야의 기반기술로 활용 가능함을 충분히 보였다.

제안하는 음원인식 기술의 직접적인 적용분야는 온라인 사용자 음원 인식 서비스 분야이며, 넓게는 음원이 활용되는 콘텐츠 분야에서 중복제거 목적으로 적용될 수 있을 것이다.

향후 개선 및 연구과제는 다음과 같다. 우선적으로 현재보다 높은 인식률을 위한 특징 추출 알고리즘에 대한 연구이다. 차후 검색속도 단축 및 일관적인 인식 속도를 유지하기 위해 요구되는 음원 특징 Indexing 기법으로 연구 영역을 확장하는 것이다.

참고문헌

[1] ID3 tag version 2.3.0, *http://www.id3.org/id3v2.3.0.html*

[2] 성보경, 김정수, 고일주, "음악특징점간의 유사도 측정을 이용한 동일음원 인식 방법," 한국컴퓨터정보학회논문지, 제 13권, 제 3호, 99-106쪽, 2008년 5월.

[3] R. Zhou and J. D. Reiss, "Music Onset detection combining energy-based and pitch-based approaches," Proc. MIREX Audio Onset Detection Contest, 2007.

[4] A. J. eronen, V. T. Peltonen, J. T. Tuomi, A. P. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho and J. Huopaniemi, "Audio-Based Contest Recognition," IEEE Trans. Audio, Speech and Language processing, vol. 14, no. 1, pp. 321-329, Jan. 2006.

[5] A. Harma, UK.Laine, "A comparison of warped and conventional linear predictive coding," IEEE Transactions on Speech and Audio Processing, Vol.9, No.5, pp.579-588, 2001.

[6] J. Makhoul, "Linear prediction-A tutorial overview," Proceeding of IEEE, Vol.63, No.4, pp.561-580, 1975.

[7] B.J. Shannon, K.K. Paliql, "A comparative study of filter bank spacing for speech recognition," Proceedings of International Micro-electronic engineering research conference, pp.1-3, Brisban, Austria, 2003.

[8] F. Zheng, G. Zhang, "Integrating the energy information into MFCC," 6th International Conference of Spoken Language Processing, Vol.1, pp.389-392, Beijing. China, 2000.

[9] Z. Jun, S. Kwong, W. Gang, Q.Hong. "Using Mel-Frequency Cepstral Coefficients in Missing

Data Technique," EURASIP Journal on Applied Signal Processing, Vol.2004, No.1, pp.340-346, 2004.

[10] M. Xu, NC. Maddage, C. Xu, M. Kankanhalli, Q. Tian, "Creating audio keywords for event detection in soccer video," Multimedia and Expo. ICME03 Proceedings, Vol.2, pp.281-284, 2003.

[11] R. Vergin, Oapos, D. Shaughnessy, A. Farhat, "Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition," IEEE Transactions on Speech and Audio Processing, Vol.7, No.5, pp.525-532, 1999.

[12] JJ. Burred, A. Lerch, "A hierarchical approach to automatic musical genre classification," Proceeding of 6th International Conference on Digital Audio Effects, pp.DAFX1-DAFX4, London, UK, 2003.

[13] L. Lu, HJ. Zhang, H. Jiang, "Content analysis for audio classification and segmentation," IEEE Transactions on Speech and Audio Processing, Vol.10, No.7, pp.504-516, 2002.

[14] E. Scheirer, M.Slaney, "Construction and evaluation of a robust multifeature," Acoustics, Speech, and Signal Processing ICASSP-97 IEEE, Vol.2, pp.1331-1334, 1997.

[15] SH. Nawab, TF. Quatieri, "Short-time Fourier transform," Advanced topics in signal processing of Prentice Hall Signal Processing Series, pp.289-337, 1987.

[16] B.K. Sung, I.J. Ko, "A Practical Method for Digital Music Matching Robust to Various Sound Qualities," Proceedings of World Academy of Science, Engineering And Technology (WASET) Vol.60, Conference on Computer, Electrical, Systems Science and Engineering 2009 (CESSE2009), pp.309-314, Bangkok, Thailand, December 2009.

[17] B.K. Sung, I.J. Ko, "Effective Digital Music Retrieval System through Content-based Features," Proceedings of World Academy of Science, Engineering And Technology (WASET) Vol.50, International Conference on Computer, Electrical, Systems Science and Engineering 2009 (ICSSE2009), pp.721-726, Penang, Malaysia, February 2003.

[18] 김정수, 성보경, 구광효, 고일주, "노이즈에 강인한 음악 시작점 검출 알고리즘," 한국컴퓨터정보학회논문지, 제 14권, 제 9호, 95-104쪽, 2009년 9월.

저 자 소 개

성 보 경

2006:숭실대학교 미디어학과(공학사)
2007:숭실대학교 미디어학과(공학석사)
2007~현재: 숭실대학교 미디어학과 박사과정 재학
관심 분야 : MIR, SNS, Social Web, Entertainment Technology

고 일 주

1992: 숭실대학교 전산학과(공학사)
1994: 숭실대학교 전산학과(공학석사)
1997: 숭실대학교 전산학과(공학박사)
2003~현재: 숭실대학교 글로벌미디어학부 부교수
관심분야 : 콘텐츠, 인공감정, 인간-로봇 인터페이스