

음악 특징점간의 유사도 측정을 이용한 동일음원 인식 방법

성보경*, 정명범**, 고일주***

Same music file recognition method by using similarity measurement among music feature data

Bokyung Sung*, Myoung-Beom Chung**, Ilju Ko***

요약

최근 다양한 분야에서(웹 포털, 유료 음원서비스 등) 디지털 음악의 검색이 사용되고 있다. 기존의 디지털 음악의 검색은 음악 데이터에 포함된 자체 메타 정보를 이용하여 이루어진다. 하지만 메타 정보가 다르게 작성 되었거나 작성되지 않은 경우 정확한 검색은 어렵다. 요즘 이러한 문제의 보완 방안으로 음악 자체를 이용하는 내용기반 정보 검색 기법에 대한 연구가 이루어지고 있다. 본 논문에서는 음악의 파형에서 추출된 특징 정보간의 유사도 측정을 통하여 동일음원을 인식하는 방법에 대해 논하고자 한다. 디지털 음악의 특징 정보는 단순화시킨 MFCC (Mel Frequency Cepstral Coefficient)를 이용하여 음악의 파형으로부터 추출하였다. 디지털 음악간의 유사도는 Vision 및 Speech Recognition 분야에서 사용되던 DTW (Dynamic Time Warping) 기법을 활용하여 측정하였다. 제안된 동일 음원 인식 방법의 검증을 위한 같은 장르에서 무작위 추출된 1000곡에서 시행한 500번의 검색은 모두 성공했다. 검색에 사용된 500개의 디지털 오디오는 60개의 디지털음원을 압축방식과 비트율을 다르게 조합하여 만들었다. 실험의 결과로 DTW를 이용한 유사도 측정법이 동일음원을 인식할 수 있음을 증명하였다.

Abstract

Recently, digital music retrieval is using in many fields (Web portal, audio service site etc). In existing fields, Meta data of music are used for digital music retrieval. If Meta data are not right or do not exist, it is hard to get high accurate retrieval result. Contents based information retrieval that use music itself are researched for solving upper problem. In this paper, we propose Same music recognition method using similarity measurement. Feature data of digital music are extracted from waveform of music using Simplified MFCC (Mel Frequency Cepstral Coefficient). Similarity between digital music files are measured using DTW (Dynamic time Warping) that are used in Vision and Speech recognition fields. We success all of 500 times experiment in randomly collected 1000 songs from same genre for proving of proposed same music recognition method. 500 digital music were made by mixing different compressing codec and bit-rate from 60 digital audios. We proved that similarity measurement using DTW can recognize same music.

▶ Keyword : Contents based audio retrieval, DTW, Simplified MFCC

• 제1저자 : 성보경

• 접수일 : 2008. 3. 13, 심사일 : 2008. 4. 1, 심사완료일 : 2008. 5. 24.

* 송실대학교 IT대학 미디어학과 박사과정 ** 송실대학교 IT대학 미디어학과 박사과정

*** 송실대학교 IT대학 미디어학과 조교수

※ 본 연구는 송실대학교 교내연구비 지원으로 이루어졌음

1. 서론

최근 정보처리기술과 통신네트워크기술의 발달로 많은 양의 멀티미디어 콘텐츠가 발생하게 되었다. 이에 필요한 콘텐츠를 편리하고 효율적으로 찾는 방법의 필요성이 증가 되었으며, 여러 가지 방법론들이 제안 되고 있다. 멀티미디어 콘텐츠의 한 부분인 음악 콘텐츠 역시 사용 범위가 광범위해져 콘텐츠의 효율적인 관리와 빠른 검색방법이 요구 되고 있다.

현재 사용되고 있는 음악 콘텐츠들은 거의 모두 디지털화된 형태이다. 디지털 음악 콘텐츠는 텍스트 데이터와 다르게 내용자체를 명확하게 수치적으로 표현할 수 있는 방법이 없다. 그래서 그동안 디지털 음악의 정보는 태그형태로 저장하였으며 콘텐츠는 태그 정보를 이용하여 검색 혹은 관리 되었다. 하지만 태그 데이터를 이용한 디지털 음악의 검색은 몇 가지의 한계점을 가지고 있다. 첫째, 사용자들은 태그의 정보보다 실제 음악의 내용에 더 관심을 가지지만 태그에 기록되어 있는 정보 이외의 것에 대한 검색이 어렵다. 둘째, 모든 디지털 음악의 태그 정보가 올바르게 저장되지 않을 가능성이 많다. 위와 같은 한계의 극복을 위해 디지털 음악 콘텐츠의 음악과형 자체를 통한 내용기반검색 기법이 필요하다.

본 논문에서는 음악의 파형에서 추출된 특징 정보간의 유사도 측정을 통하여 동일음원을 인식하는 방법에 대해 제안한다. 일반적인 사용자들이 직접 생성하는 디지털 음악은 디지털화 과정에서 서로 다른 규격(인코딩방식, 표본화 및 양자화 간격, 초당 비트율 등)으로 인코딩 된다. 내용기반 방법을 이용한 동일한 디지털 음악 검출을 위해서는 인코딩 규격에 상관없이 입력되는 디지털 음악과 비교해야할 디지털 음악 간의 파형간의 일치할 혹은 유사함을 측정 할 수 있어야 한다.

제안된 동일 음원 인식 방법은 디지털 음악의 파형 정보만을 이용하여 동일한 내용을 가지고 있는 디지털 음악을 음악 서버 안에서 찾아주는 것이다. 검색을 위해 입력되는 디지털 음악 클립과 음악서버 안에 있는 디지털 음악은 특징벡터 형태의 데이터로 변환된 후 유사도 측정 과정을 거친다. 유사도는 Vision 및 Speech Recognition 분야에서 이미지 와 음성의 합성 및 유사도측정에 사용되던 Dynamic Time Warping (DTW)[1-5] 기법을 이용하여 측정한다. 음성인식 분야에서 유사도 측정은 짧은 길이의 음성간의 비교에 사용되었지만, 디지털 음악 검색에서는 길이가 길고 복잡음인 음악에 대한 비교에 사용된다. 유사도 측정은 입력 음악과 서버에 저장된 음악들에서 추출된 특징데이터들 간의 절대 거리 차이 계산을 통해 이뤄진다. 다수의 서버의 음악들 중 입력된

음악과 측정된 유사도 값이 가장 작은 디지털 음악이 동일하다고 판단된다.

디지털 음악의 파형 정보는 유사도 측정을 하기 일정 한 길이로 분할되고, 각 부분으로부터 특징데이터가 추출 되어져야 한다. 파형 데이터로부터의 특징데이터 추출은 음성인식 분야에서 사람 목소리의 특징을 추출하기 위해 사용하고 있는 Mel Frequency Cepstral Coefficient (MFCC)[6-10] 과정을 간소화 시켜 사용 하였다. 이것은 인간 발성 모델을 기반으로 만들어진 것이 아닌 청각 모델을 기반으로 만들어진 것이기 때문에 음성뿐만 아니라 여러 악기들이 복합된 디지털 음악 에서도 특징벡터 추출이 가능하다. 디지털 음악에 적합하게 처리 과정을 단순화시킨 Simplified MFCC과정에 의해 만들어지는 벡터 중에서 13차까지의 벡터를 특징데이터로 사용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 디지털 음악검색 관련연구를 언급한다. 3장에서는 동일 음원 인식을 위한 제안 방법에 대해 설명한다. 4장에서는 서로 다른 규격을 가지고 있는 음악 간의 유사도 측정을 통한 동일 음원 인식 실험을 수행 하였으며 마지막으로 5장에서 결론으로 끝을 맺는다.

II. 관련 연구

내용기반 음악검색 방법은 크게 MIDI데이터 혹은 Tag데이터 같은 정보를 이용하는 방법과, 음악의 파형으로부터 추출된 정보를 이용하는 검색으로 구분되어질 수 있다.

MIDI 및 Tag정보를 이용한 경우는 MIDI데이터 자체가 곡의 빠르기, 성조 등의 음악적 요소를 가지고 있고 Tag데이터 자체가 각 음악의 객관적인 정보(음악가, 곡명, 발매시기 등)를 담고 있기 때문에 음악적 특징 검색은 용이하다. 하지만 디지털 음악 데이터 검색 에서는 이러한 방법을 통해 만족도가 높은 검색 결과를 도출시키기가 어렵다. MIDI데이터 및 Tag 데이터가 존재하는 디지털 음악에 대해서만 검색이 가능하기 때문이다.

음악의 파형으로부터 추출된 정보를 이용하는 검색 방법은 기본적으로 DSP(Digital Signal Processing) 기술을 기초로 하고 있다. MIDI를 통한 방법과는 다르게 DPS 기술은 음악 자체로부터 정보를 추출하기 때문에 파형정보를 지니고 있는 모든 음악 파일규격(MP3, OGG, WMA 등)에 적용이 가능하다. DPS 기술을 통해 음악과형으로부터 음의고저(Pitch), 음색(Timbre), 화음(Harmony) 등의 특징을 수치적인 형태로 추출하고 이것을 이용하여 검색한다.

III. 동일 음원 인식을 위한 제안 방법

3.1 동일 음원 인식 방법의 전체적 구조

[그림 1]은 제안된 동일 음원 인식방법의 구조를 나타낸다. 동일 음원 인식 방법 구조의 흐름은 다음과 같다. 음악을 입력하면 유사도 비교에 알맞은 규격으로 전처리를 한 후 일정부분만 음악 클립으로 만든다. 음악 클립의 파형으로부터 특징데이터를 추출한다. 추출된 특징데이터는 유사도 측정 과정에서 음악 서버에 특징데이터 형태로 저장되어 있는 모든 음악데이터 각각과 유사도가 측정된다. 마지막으로 음악서버 내에서 측정된 유사도가 가장 높은 수치를 보이는 특징데이터의 음악 파일이 동일 음원으로 인식된다.

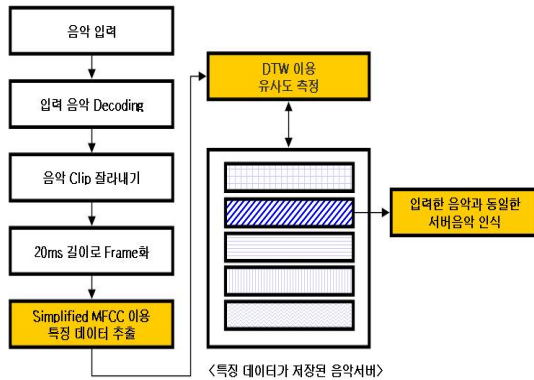


그림 1. 동일 음원 인식방법의 전체구조
Fig.1 Structure of Same music recognition method

사용자들이 사용하는 음악 파일들은 모두 같은 규격으로 디지털화 되어 있지 않다. 그러므로 전처리 과정을 통해 음악 데이터베이스에 저장되어 있는 규격으로 바꿔야 한다. 일단 각기 다른 코덱으로 압축된 것을 비압축 파형 데이터 형태로 복원해야 한다. 그리고 음악 데이터베이스와 똑 같은 규격의 양자화 간격과 부호화 정도로 입력 오디오를 변환하는 과정이 필요하다.

파형 데이터는 순간, 순간의 소리의 크기 값을 연속으로 나열한 것이다. 그래서 시간에 따른 소리의 절대 값을 의미를 가질 수 있는 형태로 변환하여야 한다. 본 구조에서는 여러 소리가 복합된 음악에 적합하게 간소화 시킨 Simplified

MFCC를 이용하여 음악의 파형으로 부터 특징 데이터를 추출한다.

동일 음원 인식을 위해서는 동일함 판단을 위한 기준이 필요하다. 이러한 기준을 위해 음악 간의 유사도 측정을 한다. 입력된 음악에서 추출된 특징데이터와 음악서버에 특징데이터 형태로 저장되어 있는 모든 음악 간의 유사도 측정을 위해 DTW 기법을 사용하였다.

3.2 음악의 특징 추출

파형 데이터는 그 자체로 음악적 의미를 지니지 못하므로 음악의 내용적인 의미를 가지는 특징 데이터 형태로 변환 시켜줘야 한다. 소리의 유사성을 측정하는 음성 인식 분야에서도 파형 형태의 음성으로부터 특징 데이터를 추출한 후 이것을 이용하여 유사성을 측정한다. 특징 데이터 추출은 [그림 2]와 같이 입력 음악클립을 10ms 길이씩 중복되는 부분이 생기도록 20ms 크기의 프레임으로 자르고, 각각의 프레임에서 특징 데이터를 추출한다.

파형 데이터로부터 특징 데이터를 추출해주는 다양한 계수들이 있다. Short Time Fourier Transform (STFT)[11] 값의 스펙트럼 중심을 뜻하며, 스펙트럼의 형태를 측정하는 방법중의 하나인 Spectral Centroid[12], 스펙트럼의 형태와 낮은 주파수 영역에 신호의 에너지가 얼마나 집중되어 있는지를 보여주는 Spectral Rolloff[13], 연속된 스펙트럼 분포에서 정규화된 절대값들 간의 차이를 제공해서 구할 수 있으며 스펙트럼 변화의 양을 계산할 수 있는 spectral flux[14], 오디오 신호의 리듬 정보를 수치적으로 산출하는 계수로서 Wavelet변환 후 대역 별 상관도를 구하여 Beat 히스토그램을 만들고 Beat정보를 추출하는 Beat Histogram[15], 인간의 발성 모델에 입각해서 음성 신호를 부호화 하는 방법으로 오디오 파형의 샘플값에서 필터 계수를 구하여 성대에서 입, 코까지 성도 특성을 8~12차의 전극형(All-pole) 필터에 근사 시키는 방법인 Linear Predictive Coding(LPC)[16-17] 및 인간 청각 특성을 모델링 하는 방법으로 오디오 신호의 절대값 스펙트럼을 log scale한 후 FFT bin을 그룹화하여 인간의 청각 특성에 맞는 Mel-Frequency 스케일로 변환한 MFCC[6-10] 등이 있다.

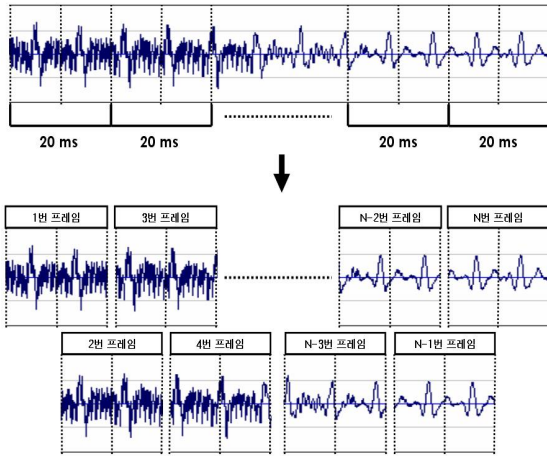


그림 2. 입력된 음악 클립의 프레임화
Fig.2 Inputted music clip is divided as frame

MFCC 는 본래 음성인식 분야에서 사람 목소리에서 특징을 추출하기 위해 사용되었다. 이것은 인간 발성 모델을 기반으로 한 것이 아니라 청각 모델을 기반으로 만들어졌다. 그래서 원래 사용되던 단채널의 음성뿐만 아니라 여러 악기들이 복합된 음악 데이터에서도 특징 벡터 추출이 가능하다고 가정하고 음악의 특징을 추출하는 것으로 사용하였다. 본 논문에서는 음성에 적합하게 설계된 것을 음악에 적합하게 사용하기 위해 몇 가지 단계를 축소하여 간소화 시킨 MFCC를 사용하였다. 그리고 간소화된 MFCC를 이용하여 추출되는 특징 벡터 중에서 1차에서13차까지를 실험 데이터로 사용 하였다.

[그림 3]은 음성인식 분야에서 사용하던 일반적인 MFCC 과정과 음악에서 특징 데이터 추출을 위해 간소화 시킨 Simplified MFCC과정을 보여준다. [그림 3]에서 (b)는 본 논문의 실험을 위해 기존의MFCC과정을 간소화 시킨 것이다. 일반적으로 사용되는 과정에서 전처리 과정, 프리-엠퍼시스 과정, 윈도우링 과정이 생략 되었다. 이 세 과정은 소리의 파형이 사람 목소리뿐만 아니라 이루어진 경우 잡음과 불필요한 무음 부분을 제거하고 음성 영역을 강조하기 위해 사용 되던 단계이다. 하지만 본 논문의 유사도 측정에 사용되는 음악 데이터는 여러 가지 소리들이 복합적으로 녹음된 상태이다. 즉 음성 부분만을 강조하는 과정이 불필요 하므로 이 과정을 제거하고 단순화 시킨 MFCC 과정을 사용하였다.

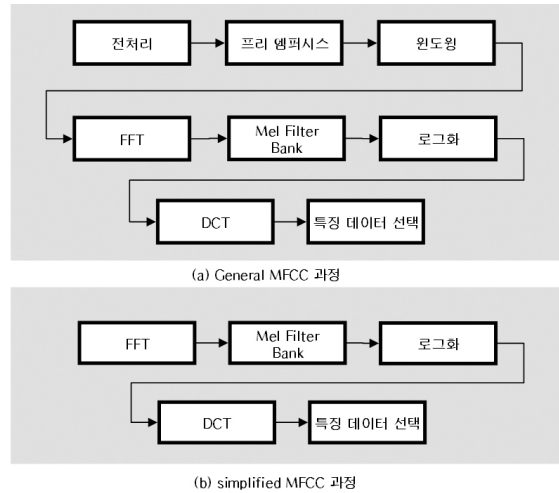


그림 3. 음악의 특징을 나타내는 특징데이터 추출과정
Fig. 3 Feature data that have musical characteristic extracting sequences

음악에서 특징벡터 추출을 위한 Simplified MFCC의 첫째 과정은 FFT이다. FFT처리는 시간 영역의 파형 신호를 주파수 영역으로 변환시키는 과정이다. 주파수 영역은 신호의 각 성분들을 알 수 있기 때문에 시간 영역에 비해 신호의 해석 및 처리가 용이한 장점이 있다. Mel Filter bank처리는 FFT처리된 결과를 가지고 멜 스케일 필터 값을 곱해주는 과정이다. 이 필터는 보통 사람은 1kHz 이하에서 잘 듣는다는 것을 이용 하여 1kHz 이하 부분은 촘촘히 분석하고 그 이상은 간격을 넓게 분석하여 좀 더 청각구조에 접근시킨 필터이다. 로그화 처리는 필터를 통과한 값에 로그를 취하는 것이다. 우리의 귀가 소리의 크기에 대해 로그 함수로 느끼기 때문에 로그화 과정을 거친다. Discrete Cosine Transform(DCT) 처리는 필터 뱅크의 출력 간의 상관 관계를 없애주고 파라미터의 특징을 모아주는 역할을 한다. 하나의 프레임에 MFCC를 적용할 경우 Mel Filter Bank의 뱅크 갯수 만큼 특징데이터 값이 나온다. 특징 추출 처리는 필터 뱅크 수만큼 나온 특징 데이터에서 작은 차수부터 필요한 갯수 만큼 선택하는 것이다.

3.3 유사도 측정 방법

대부분의 사람은 동일한 단어를 발성할 경우 발성 시마다 발성시간 길이가 변화한다. 두 발성의 파형은 완벽히 일치하지 않지만 소리의 내용은 다르지 않다. 음악에서는 디지털 규

격을 변경할 때 이와 비슷한 현상이 발생된다. 같은 파형을 다른 규격으로 디지털화 할 때 사람에게서는 동일하게 들리나 동일 시간대의 파형의 모양이 변함 [그림 4]에서 볼 수 있다. 하지만 이것 역시 내용적으로 완전히 다르다고 할 수 없다. 이러한 것을 고려하여 동일 음원을 인식 하기 위해서는 일정 범위 내에서 길이와 파형 값이 차이 나더라도 두 파형이 동일하다고 판단할 수 있는 유사도 측정법이 필요하다.

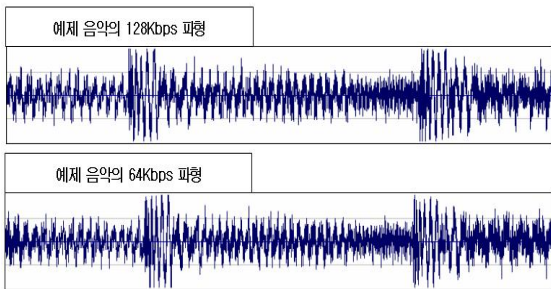


그림 4. 디지털규격이 다른 동일음악 파형의 비교
Fig. 4 This is a comparing of same music waveform graphs that have different digital spec

본 논문의 실험에서는 추출된 특징 데이터간의 유사도를 수치적으로 측정하기 위한 방법으로 DTW[3]를 사용한다. DTW는 과거 음성인식 분야에서 음성간의 유사성 측정을 위한 방법의 하나로 사용되던 것이다. 음성인식을 위해서는 입력되는 음성을 데이터베이스에 저장된 데이터와 유사도를 측정해야 한다. 본 논문의 실험은 음성데이터 간의 비교가 아니라 복합 사운드 형태의 디지털 음악 간의 비교지만 DTW를 이용하여 유사도를 측정 하였다. 유사도는 두 음악이 얼마나 절대 거리차이를 보이는지 표현한 것이다.

[그림 5]는 입력된 음악의 특징 데이터 열과 서버에 저장된 음악 중에서 한 개 음악의 특징 데이터 열이 정합되어지는 함수를 그림으로 나타내고 있다. 이와 같이 서로 다른 두 개의 자료에서 최적의 정합 경로를 동적으로 찾아 두 자료를 서로 비교하는 방법이 DTW 이다.

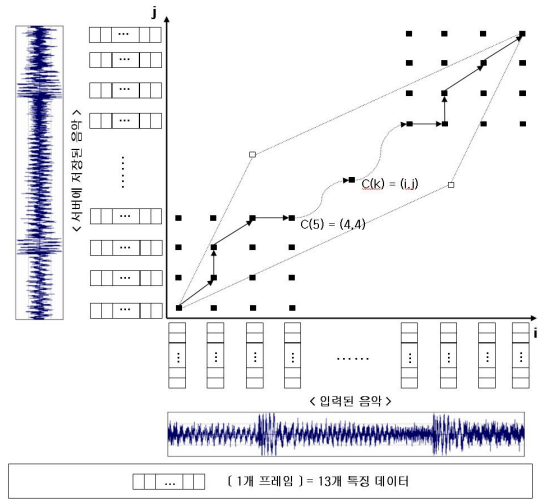


그림 5. DTW를 이용한 디지털 음악의 특징 데이터 간 유사도 측정
Fig. 5 This is a similarity measurement between feature data of two digital music by DTW

아래 식은 DTW를 계산하는 식이다. $D(A,B)$ 에서 A,B는 입력되는 두 음원 이고, 거리차이를 구하는 것이 함수 D 이며 DTW를 의미한다.

$$D(A,B) = \min_F \left[\frac{\sum_{k=1}^j d(c(k)w(k))}{\sum_{k=1}^j w(k)} \right] \dots\dots\dots (식 1)$$

유사도 측정을 위해 입력되는 두 음원은 특징데이터의 열로 표현할 수 있다. 두 음원 내에서 시간 코드가 차이가 난다고 한다면 두 음원 을 하나의 시간 축으로 정합 시켜야 한다. 이것을 해주는 함수를 와핑(Warping) 함수라 하고, $c(k)$ 가 이에 해당된다. k 는 와핑 함수에서 열로 표현되는 음원의 포인트 개수를 의미한다. $w(k)$ 는 가중치 계수로 와핑 함수의 탄력 있는 특성을 유도하는데 도입되며 적절한 와핑 함수를 찾는데 이용한다. 식에서 우변의 분모는 와핑 함수들의 가중치 합을 의미한다. 함수에서 포인트 개수 k 에 의한 영향을 보상하기 위해 사용된다.

IV. 실험 및 결과

실험에서 사용하는 음악 서버는 동일 장르에서 무작위 추출된 1000개의 음악으로 구성되었다. 음악 서버에 저장된 음

악들은 22050Hz로 양자화 되고, 16bits로 부호화 되어 저장 되었다.

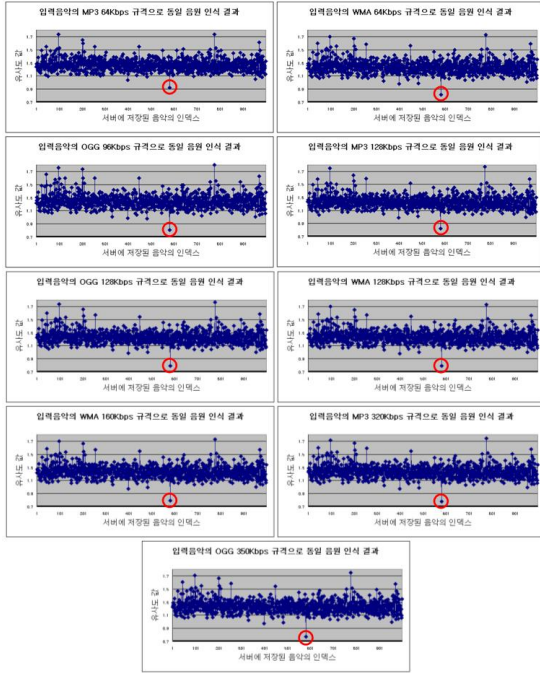


그림 6. 동일한 곡을 서로 다른 9가지 규격으로 변환 후 동일음원 인식 실험한 결과

Fig. 6 These are result graphs that experiment with 9 different spec music that are made by one input music

인식 명령을 위해 사용한 음악은 60곡이며 이 음악은 음악 서버의 1000곡 내에 존재하는 음악이다. 60개의 음악은 파형변화에 강인함에 대한 증명을 실험하기 위해 각각 60개의 음악은 압축방식, 초당 비트율을 다르게 조합한 9가지의 다른 규격으로 변환하여 총 500개의 음악으로 만들었다. 9가지 규격은 MP3 (64kbps, 128kbps, 320kbps), OGG (96kbps, 128kbps, 350kbps), WMA (64kbps, 128kbps, 160kbps) 이다.

제한한 동일 음원 인식 구조는 동일한 규격으로 디지털화된 음악 간의 인식뿐만이 아니라 다른 규격으로 디지털화 되어 있어도 동일 음원 인식을 위한 것이다. 실험내용은 인식 명령을 위한 음악 1곡을 9가지의 다른 규격으로 디지털화 하여 실험을 하여도 동일 음원 인식을 할 수 있다는 것을 증명하는 것이다. [그림 6]은 하나의 음악을 다른 규격을 가지는 9가지의 음악 파일로 변환한 후 인식을 위해 서버의 음악들과 측정할 유사도 값을 그래프로 표현한 것이다.

유사도는 입력된 짧은 길이의 음악 클립과 서버에 존재하

는 모든 음악들 사이에서 일대 다 형태로 측정된다. 입력 음악 클립은 짧은 길이 이므로 원본 형태의 음악이 저장된 서버의 음악과 유사도를 측정할 때는 순차적인 슬라이드 형태로 진행된다. 이때 슬라이딩 형태로 진행되는 매 순간 순간마다 유사도 값이 측정 된다. 그 중 가장 작은 값이 입력된 음악클립과 서버 음악 한곡간의대표 유사도 값이 된다. 서버의 모든 음악은 입력음악과의 사이에서 측정된 유사도 값을 가지고 되고 이중 가장 큰 수치를 보여주는 유사도 값의 음악이 동일음악으로 인식된다.

각 그래프의 세로축은 입력된 음악 클립과 유사도가 측정된 음악 서버에 있는 1000개 음악의 대표 유사도 값이다. 유사도가 높을수록 유사도 수치는 0에 가까워진다. 각 그래프의 가로축은 음악 서버에 저장된 1000개 음악의 인덱스이다. [그림 6]에 사용된 음악은 음악 서버의 581번째 음악과 같은 노래이다. 모든 서브 그래프의 581번째 음악에서 유사도 값이 다른 서버의 음악들 보다 0에 가깝게 나왔다. 이것은 디지털 규격 변화에 따른 파형의 변형에 상관없이 동일 음원을 인식함을 보여준다. 유사도 그래프에서 0에 일치하지 않게 나오는 이유는 음악들 간에 유사도 비교를 조밀하게 비교하지 않았기 때문이다.

실험은 위와 같은 방식으로 60곡을 이용하여 총 500번의 동일 음원 인식 실험을 실시하였다. 전체 실험에서 동일 음원 인식 결과는 <표 1>과 같이 입력된 음악클립과 동일한 음악을 음악 서버로부터 찾은 성공률이 100% 임을 보였다. 유사도 측정을 위해 진행한 순차적 슬라이드 비교를 본 실험보다 덜 조밀하게 할 경우에는 입력된 음악 클립과 동일한 부분의 비교가 발생되지 않을 수 있으며 이를 통해 실패율이 발생하여 100%이하의 성공률을 보일 수 있다.]

표 1 동일 음원 인식 결과

Table 1 Result of Same music recognition experiment

규격 구분	실험 파일	성공 파일	성공율
64Kbps MP3	55 개	55 개	100%
64Kbps WMA	55 개	55 개	100%
96Kbps OGG	55 개	55 개	100%
128Kbps MP3	55 개	55 개	100%
128Kbps WMA	56 개	56 개	100%
128Kbps OGG	56 개	56 개	100%
160Kbps WMA	56 개	56 개	100%
320Kbps MP3	56 개	56 개	100%
350Kbps OGG	56 개	56 개	100%
전체	500개	500개	100%

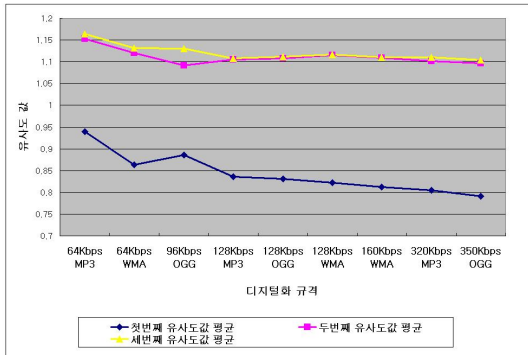


그림 7. 디지털 규격별 첫 번째, 두 번째, 세 번째 유사도 값 평균
 Fig. 7 These are average value graphs of first, second and third similarities in each digital specs

[그림 7] 은 총 500개의 음악에서 각 디지털 규격별 첫 번째, 두 번째, 세 번째 유사도 값의 평균을 그래프로 나타낸 것이다. 첫 번째 유사도 평균값이 두 번째, 세 번째 유사도 평균값과 큰 차이를 보인다. 이는 측정된 유사도 값이 변별력이 있는 수치임을 보인다.

V. 결론

본 논문에서는 내용기반 음악 검색을 위한 동일 음원 인식 방법을 제안 하였다. 제안된 동일 음원 인식은 DTW 기법을 통한 유사도 측정법을 통해 구현되었다. 입력된 디지털 음악은 Simplified MFCC를 이용하여 파형 데이터에서 특징 데이터를 추출 하였고, 두 음악의 특징 데이터들 간의 유사도를 측정 하였다. 유사도 측정법을 통하여 다른 규격으로 디지털화된 음악 사이에서도 동일함을 인식할 수 있었다.

기존의 기술을 이용하여 동일음악을 인식하기 위해서는 음악콘텐츠의 부가적인 정보(Tag정보 및 파일이름 등) 을 사용해야만 가능하다. 하지만 본 논문에서 제안된 음악 특징점간의 유사도 측정을 이용한 동일음원 인식방법은 디지털 음악콘텐츠에 음악관련 정보를 추가적으로 삽입하지 않아도 음악콘텐츠의 파형 정보 자체만으로도 동일음원을 인식할 수 있게 한다.

다른 규격으로 디지털화된 음악 간의 유사도 측정법을 통한 동일 음원 인식 방법은 내용기반 오디오 검색을 위해 필요한 기술이다. 대부분의 사용자가 음악 검색을 위해 입력하는 음악조각 들은 모두 같은 조건으로 디지털화 되어 있지 않을 가능성이 높기 때문이다.

향후 개선 및 연구과제로는 유사도 측정법을 이용하여 동

일 음원의 인식뿐만 아니라 비슷한 일정 기준에 기반 하여 비슷한 분위기의 음악까지 인식하는 방법에 대한 연구를 들 수 있다.

참고문헌

- [1] E.J. Keogh, M.J. Pazzani, "Derivative Dynamic Time Warping", First SIAM international Conference on Data Mining, Chicago, USA, 2001.
- [2] J.C. Brown, A. Hodgins-Davis, P.J.O. Miller, "Classification of vocalizations of killer whales using dynamic time warping", Journal of the Acoustical Society of America 119, Vol.119, No.3, pp.EL34-EL40, 2006.
- [3] A.M. Youssef, T.K. Abdel-Galil, E.F. El-Saadany, M.M.A. Salama, "Disturbance classification utilizing dynamic time warping classifier", IEEE Transactions on Power Delivery, Vol.19, No.1, pp.272-278, 2004.
- [4] A. Pikrakis, S. Theodoridis, d. Kamarotos, "Recognition of isolated musical patterns using context dependent dynamic time warping", Speech and Audio Processing of IEEE Transactions on, Vol.11, pp.175-183, 2003.
- [5] E. Keough, C.A. Ratanamahatana, "Exact indexing of dynamic time warping", Proceedings of the 28th international conference on Very Large Data Bases, pp.406-417, HongKong, China, 2002.
- [6] 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.
- [7] 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.
- [8] 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.

[9] 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.

[10] 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.

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

[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] 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.

[14] 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.

[15] G. Tzanetakis, G. Essl, P. Cook, "Human perception and computer extraction of musical beat strength", Proceeding of 5th International Conference on Digital Audio Effects, pp.DAFX257-DAFX261, Hamburg, Gemany, 2002.

[16] 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.

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

저 자 소개



성 보 경

2006년 : 숭실대학교 미디어학부 졸업 (공학사)
 2007년 : 숭실대학교 미디어학과 졸업(공학석사)
 2007년~현재 : 숭실대학교 미디어학과 박사과정
 관심 분야 : 디지털 음악 분석, 음악 검색, 감성공학



정 명 범

2004년 : 숭실대학교 미디어학부 졸업 (공학사)
 2006년 : 숭실대학교 미디어학과 졸업(공학석사)
 2006년~현재 : 숭실대학교 미디어학과 박사과정
 관심 분야 : 디지털 신호처리, 감성인식, 콘텐츠 공학



고 일 주

1992년 : 숭실대학교 전산학과(공학사)
 1994년 : 숭실대학교 전산학과(공학석사)
 1997년 : 숭실대학교 전산학과(공학박사)
 2003년 ~ 현재 : 숭실대학교 미디어학부 조교수
 관심 분야 : 콘텐츠, 정보검색, 감성공학