

노이즈에 강인한 음악 시작점 검출 알고리즘

김정수*, 성보경**, 구광효*, 고일주***

The Noise Robust Algorithm to Detect the Starting Point of Music for Content Based Music Retrieval System

Jungsoo Kim *, Bokyung Sung **, Kwanghyo Koo *, Ilju Ko ***

요 약

본 논문에서는 노이즈에 강인한 음악 시작점 검출 알고리즘을 제안한다. 음악의 시작점 검출은 음악을 이용한 신호처리 시스템에 있어서 일관되지 않은 입력데이터를 통한 계산낭비, 비교검색 등의 문제 해결을 위해 필요한 것이다. 특히 신호처리를 이용한 내용기반 음악 검색 시스템에서 시간의 순서로 데이터를 비교하는 시간 순차적 검색 방법에서는 더욱 필요시 된다. 시간 순차적 검색 방법은 시간의 순서로 단순 비교를 수행하기 때문에 검색의 속도가 빠르다는 장점이 있는 반면 비교하는 데이터의 시작 시간이 동일해야 하는 단점이 있다. 하지만 디지털화된 음악은 비트레이트 변환에 의한 시작 시간의 동일함을 보장할 수 없다. 따라서 본 논문에서는 검색의 전처리 단계에서 음악의 시작점을 검출함으로써 시간 순차적 검색 방법을 적용하여 고속의 검색을 수행하면서도 인식률이 낮아지지 않게 하였다. 시작점 검출은 소리를 검출할 수 있는 최소파형모형을 이용하였으며, 노이즈에 강인하기 위하여 묵음에 존재하는 노이즈는 스킵핑을 하였다. 제안한 알고리즘은 실험을 통해 시작점 검출을 미적용 한 결과보다 약 38% 성능이 향상됨을 확인하였으며 노이즈에 강인함을 검증하였다.

Abstract

This paper proposes the noise robust algorithm to detect the starting point of music. Detection of starting point of music is necessary to solve computational-waste problem and retrieval-comparison problem with inconsistent input data in music content based retrieval system. In particular, such detection is even more necessary in time sequential retrieval method that compares data in the sequential order of time in contents based music retrieval system. Whereas it has the long point that the retrieval is fast since it executes simple comparison in the order of time, time sequential retrieval method has the short point that data starting time to be compared

• 제1저자 : 김정수 교신저자 : 고일주

• 투고일 : 2009. 08. 14, 심사일 : 2009. 09. 01, 게재확정일 : 2009. 09. 14.

* 숭실대학교 IT대학 미디어학과 석박사 통합과정 ** 숭실대학교 IT대학 미디어학과 박사과정

*** 숭실대학교 IT대학 미디어학과 조교수

※ 본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음

should be the same. However, digitalized music cannot guarantee the equity of starting time by bit rate conversion. Therefore, this paper ensured that recognition rate shall not decrease even while executing high speed retrieval by applying time sequential retrieval method through detection of music starting point in the pre-processing stage of retrieval. Starting point detection used minimum wave model that can detect effective sound, and for strength against noise, the noises existing in mute sound were swapped. The proposed algorithm was confirmed to produce about 38% more excellent performance than the results to which starting point detection was not applied, and was verified for the strength against noise.

▶ Keyword : 음악 시작점 검출(Music Starting Point Detection), 최소파형모형(Minimum Wave Model : MWD), 노이즈 스킵핑(Noise Skipping)

I. 서론

일반적으로 디지털 음악파형은 음악의 시작점을 기준으로 앞부분에는 묵음 영역, 뒷부분의 실제소리 영역으로 나뉜다. 여기서 음악의 시작점은 묵음 영역과 소리 영역의 경계 지점이다. 시작점의 검출은 음악을 이용한 신호처리 시스템에 있어서 기본적인 중요 문제이다. 묵음의 길이는 음악마다 일정하지 않고 다르기 때문에 시작점을 정확하게 검출하면 순수한 음악 부분만 골라내어 사용할 수 있으므로 불필요한 묵음 영역 데이터의 계산 낭비를 막게 된다. 처리 시간이 줄어들게 되며, 묵음 길이에 상관없이 일관된 입력데이터를 가질 수 있으므로 신호처리를 이용한 비교 검색이 용이해진다.

음악에서 신호처리를 이용한 비교검색은 내용기반 음악 검색을 의미한다. 현재 멀티미디어 데이터의 급속한 보급으로 데이터베이스에 저장된 음악 정보가 증가하고 있으며, 인터넷을 이용한 음악을 수집하고 소장하는 사용자들이 증가하고 있다. 이에 따라 사용자들은 방대한 양의 음악을 효율적으로 검색할 수 있는 시스템을 요구하고 있다. 기존의 음악 검색은 제목이나 가수 등의 메타 데이터 정보를 기반으로 한 검색이 주로 이용되었다. 이러한 메타 데이터 기반 검색 시스템은 사용자가 원하는 음악의 메타 정보를 알지 못하면 검색하기 어려운 문제가 있다. 이에 따라 음악의 내용이 갖고 있는 정보를 이용하여 검색하는 내용기반 검색 기술이 중요시 되고 있다.

내용기반 음악 검색은 음악의 내용을 수학적으로 분석하여 유사도가 높은 음악 데이터를 검색하는 것으로 수학적으로 분석하는 과정에서 신호처리 기술이 사용된다. 따라서 입력되는 데이터의 시작 시간이 중요하며, 특히 음악의 시작을 기점으로 시간의 순서로 데이터를 비교하는 시간 순차적 검색 방법에서는 시간에 더욱 민감하다.

입력되는 디지털 음악은 비트레이트 변환이나 편집에 의해 시작 시간이 달라질 수 있다. 빈번한 예는 사용자가 음악 데이터의 음질을 조정하기 위해 일정한 비트레이트로 변환을 하거나, 임의적으로 사운드 편집 툴을 이용하여 묵음을 조정하는 것이다. 이러한 시작 시간 변형은 시간 순차적 검색에서 결과의 오류를 높게 된다.

만약 음악 내용의 구조 분석을 통하여 시작점을 검출하고 이를 이용한 내용기반 검색을 한다면 결과오류를 최소화 할 수 있다. 따라서 음악의 시작점을 검출하여 두 음악 데이터간의 시작 시간을 동일하게 맞추는 필요가 있다. 본 논문에서는 이러한 이유로 실제소리를 찾기 위해 최소파형모형 이론을 적용하고 음악 시작점을 검출하는 방법을 제안하였다. 음악은 녹음 환경의 기계적인 문제나 전기 신호의 간섭 또는 다른 신호의 간섭, 아날로그 신호의 디지털화 등 다양한 경로를 통해 노이즈가 발생할 수 있다. 시작점을 검출하는데 있어서 묵음에 존재하는 노이즈는 방해요인이 되므로 음악의 묵음에서 발생할 수 있는 노이즈의 유형을 구분 짓고, 스킵핑 하여 노이즈에 강인한 시작점을 검출하였다.

본 논문의 구성은 다음과 같다. 2장에서는 내용기반 음악 검색에 관한 기존의 연구들과 시간 순차적 검색에 대하여 설명한다. 3장에서는 시작점을 검출하기 위하여 소리를 검출할 수 있는 최소파형모형과 묵음 노이즈 스킵핑 방법에 대해 설명한다. 4장에서는 본 연구에서 제안한 알고리즘의 성능을 검증하기 위해 실험을 수행하였으며, 5장에서는 결론으로 끝맺는다.

II. 관련 연구

내용기반 음악 검색은 음악의 내용을 수학적으로 분석하여 유사도가 높은 음악 데이터를 검색하는 것이다. 음악의 내용

을 수학적으로 분석하는 과정에서는 입력 신호로부터 대상을 표현하는 공통적인 특징을 추출하고, 유사도가 높은 음악 데이터를 검색 하는 과정에서는 추출된 데이터를 이용하여 비교하는 데이터 간 유사도 값이 가장 높게 나타나는 음악을 검색한다.

데이터를 비교하여 검색하는 과정에서 음악은 비트레이트 변환 시마다 시작 시간의 동일함을 보장할 수 없기 때문에 기존 연구에서는 프레임 단위의 이동이나 타임 와핑(Time Warping) 거리 등을 이용하여 검색하였다[1]. 이러한 검색 방법은 시작 시간의 오차를 해결할 수 있지만 검색하기 위한 연산량이 증가하여 검색의 속도가 느려지는 단점을 갖고 있다. 이와 다르게 시간 순차적 검색 방법은 시간의 순서로 단순 비교만을 수행하기 때문에 빠른 검색 속도를 갖는 장점이 있는 반면, 입력되는 데이터의 시작 시간의 오차가 있는 경우에 인식이 떨어지는 단점이 있다. 본 논문에서는 입력되는 음악의 시작점을 정확하게 검출하는 알고리즘을 제안함으로써 시간 순차적 검색 방법을 적용하여 고속검색을 수행하면서도 인식이 떨어지지 않게 하였다.

2.1 내용기반 유사 검색

기존 연구에서는 비교하는 데이터 간 시간의 오차로 인해 프레임 단위의 이동 또는 정렬이나 타임 와핑 거리 등을 이용하여 내용기반 검색을 하였다. 논문 [2]에서는 신호처리 도메인에서 타임 와핑 아래 효율적인 유사 검색 방법을 제안하였다. 신호의 유형은 유사하지만 시간 도메인이 다르면 유클리드 또는 도시-블록 알고리즘을 이용한 거리 계산 값이 전혀 다른 신호로 인식하게 되므로 시계열 데이터 값에 이동을 적용하였다. 그러나 이동을 통한 유사도 측정 방식은 데이터베이스의 빈번한 접근과 계산 비용이 높다는 문제가 있으므로 FastMap 기반의 인덱스 구조 설계 기술을 사용하였다. 기존 이동 방법을 이용 하였을 때 보다 지속적인 성과를 나타내었으며, 최고 7.8배 속도 증가를 보였다. 논문 [3]도 음성 도메인에서 시계열 데이터를 다루었다. 복제된 음성 신호는 시간 축으로 다양하게 이동 될 수 있고 압축과 확대 될 수 있다. 이에 CPM(Continuous Profile Model)을 적용한 시간 정렬 방법을 제안하였고 복제된 시계열들을 동시에 분석할 수 있도록 하였다.

논문 [4]에서는 시계열 데이터의 유사 검색에 사용되어 왔던 기존의 여러 거리척도 중 유사 음악 검색 시스템과 관련이 있는 균일 스케일링 거리, 타임 와핑 거리, 스케일드 엔 워프트 매칭 거리를 정의하였다. 여기서 스케일드 엔 워프트 매칭 거리 함수는 균일 스케일링과 타임 와핑 거리를 결합한 형태로 균일 스케일링 거리[5-6]는 두 시퀀스의 거리를 계산할

때 시간 축 상으로의 전체적인 확대와 축소를 고려하는 거리 척도로 질의 시퀀스가 전체적으로 빠르게 혹은 느리게 생성될 수 있을 때에 유용한 방법이다.

타임 와핑 거리[7-8]는 두 시퀀스 간의 각 요소를 비교할 때, 요소 값을 임의의 수만큼 반복하는 것을 허용하는 변형이다. 유클리드 거리 같은 경우 시간 축이 어긋나면 시퀀스 간의 거리가 매우 커지는데 반해 타임 와핑 거리는 시퀀스 내의 각 요소가 반복되는 것을 허용하여 두 시퀀스 축의 뒤틀림을 보장하여 시간축의 어긋남에 덜 민감한 방법이다.

2.2 시간 순차적 검색

시간 순차적 검색은 본 논문에서 제안한 시작점 검출 알고리즘의 성능을 측정하기 위해 사용된 내용기반 음악 검색 시스템이다. 검색을 하기위해 사용된 특징은 MFCC(Mel Frequency Cepstral Coefficients) 이다.

MFCC[9-11]는 인간의 청각 특성을 모델링하는 방법으로 음성을 인지할 때 각 주파수 성분을 선형적으로 인지하지 않고 멜 스케일로 음성을 인지한다. 멜 스케일은 사람이 1kHz 이하에서 민감하게 듣는 것을 이용하여 1kHz 이하 부분은 세밀하게 분석하고, 그 이상은 간격을 넓게 분석하여 청각 구조에 접근시킨 방법이다. 청각 모델을 기반으로 설계된 것이기 때문에 음성뿐만 아니라 여러 소리들이 복합된 디지털 음악에서도 유효한 특징벡터를 갖는다.



그림 1. MFCC 특징 추출 과정
Fig. 1. MFCC Feature Extraction Process

그림 1은 MFCC의 특징 추출 과정이다. 입력된 신호의 프레임에 대해서 pre-emphasis 과정과 해밍 윈도우를 거친 후 FFT(Fast Fourier Transform)를 수행한다. 여기서 FFT 처리는 시계열 데이터의 파형 신호를 주파수 영역으로 변환시키는 과정이다. 멜 필터 뱅크는 FFT 처리된 결과를 가지고 멜 스케일 필터 값을 곱해주는 과정으로 필터 간 경계 부분의 불연속성을 줄이기 위하여 일반적으로 삼각형 필터를 중첩시켜 사용한다. 로그화 처리는 필터를 통과한 값에 로그를 취하는 것으로 인간의 귀가 소리의 크기에 대해 로그 함수로 느끼기 때문이다. 마지막 DCT 처리는 필터 뱅크의 출력 간의 상관관계를 없애주고 파라미터의 특징을 모아주는 역할을 한다. 특징은 필터 뱅크의 개수만큼 결과가 나오며 본 논문에서는

13차 특징 벡터 계수를 사용하였다.

추출된 특징 벡터는 차원이 높아 데이터 량이 방대하기 때문에 비교를 하기에 용이하지 않다. 차원을 줄이기 위하여 차원의 값을 합하여 일차원을 갖는 특징 데이터로 구성한다. 즉, 시간의 순서로 음악의 특징 경향이 나타날 수 있도록 한 것이다. 검색은 이러한 과정을 거쳐서 나온 특징 데이터를 이용하며 시간의 순서로 도시-블록 거리 값을 구해 거리 값이 가장 적게 나오는 음악을 검색한다.

III. 노이즈에 강인한 음악 시작점 검출

노이즈에 강인한 음악 시작점 검출 알고리즘은 그림 2와 같은 과정을 갖는다. 디지털 음악을 입력데이터로 하여 일정한 프레임 단위로 최소파형모형 검사를 수행한다. 최소파형모형 검사는 소리가 존재하는 프레임을 검사하는 것으로, 묵음과 소리의 프레임을 구분 짓는다. 따라서 음악의 시작점을 검출할 수 있게 된다. 하지만 음악의 시작점을 검출하기 이전에 음악의 시작 부분에 노이즈가 존재하면 노이즈를 시작점으로 검출하게 되는 문제가 발생한다. 따라서 최소파형모형 검사를 수행한 후 바로 시작점을 검출하지 않고 후보군으로 남겨둔 후 노이즈 스킵핑을 적용한 다음에 시작점을 검출한다. 이러한 과정은 노이즈에 강인한 음악의 시작점 위치를 결과로 얻는다.

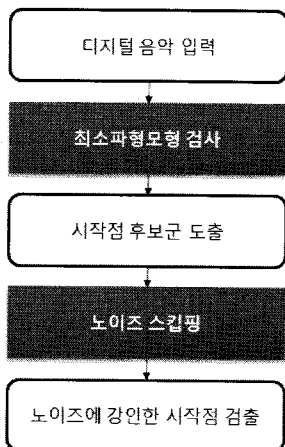


그림 2. 알고리즘 전체 구조도
Fig. 2. Algorithm Overall Flow Chart

3.1 최소파형모형 검사

최소파형모형 검사는 음악의 시작점을 검출하기 위하여 묵음 영역과 실제소리 영역을 구분 짓는 역할을 한다. 검사는

일정한 프레임 단위로 분할하여하며 단위는 음파의 길이 변화에 대해 청각적인 구별을 할 수 있는 5ms를 사용하였다[12]. 검사된 프레임 내에 최소파형모형이 존재할 경우 소리를 갖는 프레임으로 구분되며, 존재하지 않을 경우 묵음 프레임으로 구분된다. 이러한 구분을 통해 소리를 갖는 프레임들은 시작점 후보군으로 도출된다.

사람이 소리를 인지하는 원리는 공기가 진동하여 청각기관을 자극하기 때문이며 공기의 진동은 파동을 형성한다. 파동은 매질내의 한 점에서 생긴 매질의 진동 상태가 매질을 통해서 주기적으로 퍼져나가는 현상을 말하며 공간상의 한 점에서 서로 순환적으로 변환되는 에너지가 존재하게 된다. 파동의 주기는 일반적으로 사인 또는 코사인처럼 규칙적으로 퍼져 나가지만 음악의 파동은 다양한 악기의 소리와 음성이 섞여 있기 때문에 불규칙한 주기의 파동을 가진다.

최소파형모형은 불규칙한 주기를 가지는 파동의 파형에서 소리를 최소단위로 구성할 수 있는 반 주기 파형을 말한다. 보통 파형의 최소단위는 한 주기인데 반해 반 주기를 사용하는 것은 음악에서 파형은 불규칙한 주기의 파동을 가지므로 한 주기를 최소단위로 보지 않고 한 부호의 에너지 값을 가지는 반 주기 파형을 최소 단위로 사용한다. 여기서 부호는 파형의 주기에서 0을 기준으로 음수와 양수의 값을 의미하며 에너지는 파형의 세기를 나타낸다.

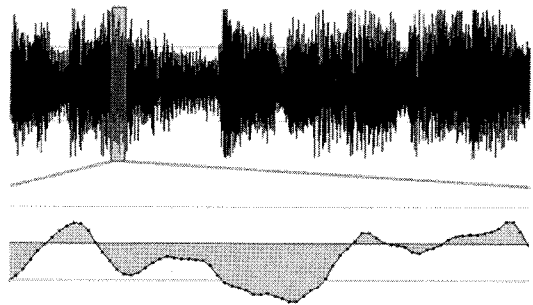


그림 3. 반 주기를 갖는 최소파형모형의 예
Fig. 3. Example of Minimum Wave Model that has half-period

음악에서 존재하는 모든 반주기 파형들은 그 자체만으로는 최소파형모형이 될 수 없고, 반주기 파형을 형성하는 샘플의 개수가 설정한 임계치 개수 보다 많을 경우 최소파형모형이 된다. 임계치 개수는 TH_N 으로 나타내며, TH_N 보다 많은 샘플을 갖는 반주기 파형모형을 최소파형모형이라 한다. 그림 3은 음악의 불규칙한 파동에서 최소파형모형을 보여준다. 부호에 따라 양과 음의 에너지를 갖는 최소파형모형으로 구분되

며, TH_N 을 1로 설정하였을 때 그림에서는 양의 에너지를 갖는 최소파형모형이 3개, 음의 에너지를 갖는 최소파형모형이 2개 검사된다.

3.2. 노이즈 스킵핑

디지털 신호에서 노이즈는 일반적으로 처리에 필요하지 않고 바라지 않는 정보를 말한다. 녹음 환경의 기계적인 문제나 전기 신호의 간섭 또는 다른 신호의 간섭, 아날로그 신호의 디지털화 등 다양한 경로를 통해 노이즈가 일어날 수 있다. 특히 시작점을 검출하는데 있어서 묵음 영역에 존재하는 노이즈는 방해요인이 된다.

소리와 같이 노이즈도 파형과 에너지를 가진다. 따라서 최소파형모형이 검사되며, 묵음 영역에 노이즈가 존재하면 최소파형모형을 이용한 올바른 시작점을 검출하기 힘들다. 따라서 노이즈의 영향을 받지 않고 스킵핑 할 수 있어야 한다. 노이즈에 강인한 음악 시작점을 검출하기 위하여 노이즈 파형이 가지는 특징을 구분 짓고, 측정 대상이 되는 최소파형모형과 후보군 프레임이 노이즈의 특징을 보인다면 스킵핑 한다. 묵음 영역 노이즈는 실험에서 사용된 디지털 음악들을 통해 노이즈의 특징을 분류하였고, 그림 4와 같이 4가지 형태의 유형으로 구분된다.

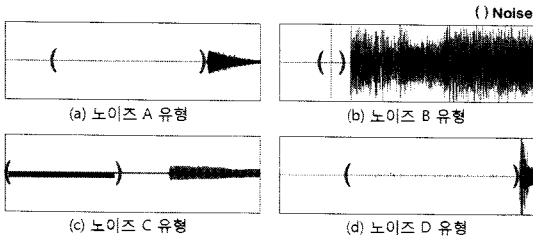


그림 4. 묵음 영역 노이즈 파형의 4가지 유형
Fig. 4. Four Types of Noise Waves

묵음 영역 노이즈 파형의 특징은 다음 표 1과 같다. 표 안의 TH_A , TH_B , TH_C , TH_D 는 각 노이즈의 임계치 값을 의미한다.

표 1. 묵음 영역 노이즈 파형의 특징
Table 1. The Characteristics of Noise Waves

특징 파형	에너지 강도	파형의 성질	파형의 부호 비율	최소파형모 형 개수
소리	높음	연속	비 편중	적음
노이즈 A	$> TH_A $	$> TH_B$	$> TH_C$	$< TH_D$
노이즈 B	높음	연속	편중	적음
노이즈 C	$> TH_A $	$> TH_B$	$< TH_C$	$< TH_D$
노이즈 D	높음	연속	비 편중	많음
	$> TH_A $	$> TH_B$	$> TH_C$	$> TH_D$

노이즈	높음	불연속	비 편중	적음
B	$> TH_A $	$< TH_B$	$> TH_C$	$< TH_D$
노이즈 C	높음	연속	편중	적음
	$> TH_A $	$> TH_B$	$< TH_C$	$< TH_D$
노이즈 D	높음	연속	비 편중	많음
	$> TH_A $	$> TH_B$	$> TH_C$	$> TH_D$

일반적인 소리 파형의 특징은 사람의 귀로 인지 가능한 $|TH_A|$ 이상의 에너지 강도 가지고 있으며, 연속되는 성질을 가지므로 TH_B 이상 프레임에 최소파형모형이 검사된다. 음악의 파형은 불규칙하더라도 주기의 특성을 갖고 있기 때문에 검사되는 최소파형모형의 부호 비율이 한쪽으로 크게 TH_C 이상 편중되지 않으며 최소파형모형의 개수는 비교적 낮은 TH_D 미만의 개수를 보인다. 이와 다르게 노이즈 파형은 소리와는 다른 특징을 보인다.

노이즈 A는 에너지 강도가 매우 낮아 사람의 귀로 인지하기 힘든 $|TH_A|$ 미만의 에너지 강도를 가진다. 노이즈 B는 파형의 성질이 불연속적이므로 TH_B 미만의 연속되지 않는 프레임을 갖는다. 노이즈 C는 최소파형모형의 부호 비율이 한쪽으로 TH_C 미만 편중되는 특징을 보인다. 노이즈 D는 최소파형모형의 개수가 소리의 특징과 비교하였을 때 상대적으로 TH_D 보다 많은 특징을 보인다.

이러한 노이즈는 스킵핑 하는 시점에 따라 크게 두 가지로 나뉜다. 첫 번째는 프레임 내부 스킵핑으로 최소파형모형 검사 시 수행한다. 노이즈 A가 이에 해당되며, 프레임 내부 최소파형모형을 검사하면서 반주기 파형을 구성하는 샘플이 노이즈인지 판단하고 스킵핑 한다. 두 번째는 프레임 외부 스킵핑으로 최소파형모형 검사를 수행한 후 시작점 후보군이 도출되었을 때 수행한다. 노이즈 B, C, D가 이에 해당되며 시작점 후보군 프레임들을 대상으로 노이즈 프레임인지 판단하여 스킵핑을 한다.

프레임 내부와 외부 노이즈를 스킵핑 하는 방법은 다음과 같다. 먼저 프레임 내부 노이즈 스킵핑은 최소파형모형 검사 시 수행되므로 최소파형모형을 검사하기 위하여 샘플의 개수 TH_N 을 카운트하는 과정에서 수행된다. 샘플의 에너지 강도 값이 $|TH_A|$ 미만이면 해당 샘플은 노이즈 A로 판단하고 스킵핑하여 카운트되지 않는다. 이 경우 최소파형모형 검사에서 몇몇 파형의 샘플이 스킵핑 되더라도 프레임 내에 최소파형모형이 존재 한다면 후보군 프레임으로 될 수 있다.

프레임 외부 노이즈 스킵핑은 도출된 시작점 후보군 프레임을 이용하여 스킵핑 여부를 결정한다. 노이즈 B를 위해 한 개의 프레임만 측정하는 것이 아니라 연속된 TH_B 이상의 프

레이를 측정하여 판단한다. 음의 불연속성은 파형이 연속되어 있지 않고 한 부분에만 집중되어 있기 때문이다. 연속된 두 개 이상의 프레임에서 단 한 프레임이라도 최소파형모형이 검사되지 않으면 스킵핑 한다. 노이즈 C는 연속되는 프레임에서 최소파형모형을 부호에 따라 개수를 따로 합산하고 부호비율을 구하여 한쪽으로 크게 TH_C 미만 치중되면 스킵핑 한다. 노이즈 D는 연속되는 프레임에서 최소파형모형 총 개수가 소리의 특징과 비교하였을 때 상대적으로 높은 TH_D 이상의 개수를 보이면 스킵핑 한다.

3.3 노이즈에 강인한 음악 시작점 검출

노이즈에 강인한 음악 시작점 검출은 최소파형모형 검사와 노이즈 스킵핑을 함께 적용한다. 최소파형모형 검사를 통해 실제 소리를 갖는 시작점 후보군 프레임을 도출하고 프레임 내부와 외부 노이즈 스킵핑을 적용하여 노이즈에 강인한 시작점을 검출한다. 그림 5는 시작점 후보군 프레임을 도출하는 순서도이며, 그림 6은 프레임 외부 노이즈 스킵핑을 적용하여 노이즈에 강인한 시작점을 검출하는 순서도이다.

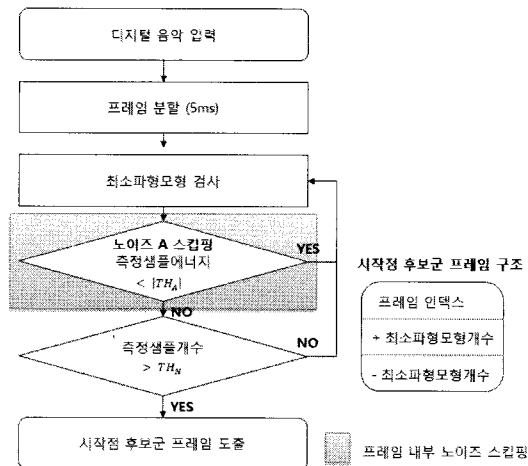


그림 5. 시작점 후보군 프레임 도출 순서도
Fig. 5. Flow Chart of Starting Point Candidates Group Frame Deduction

시작점 후보군 프레임을 도출하는 과정은 다음과 같다. 디지털 음악을 입력데이터로 하고 5ms를 한 프레임 단위로 이동하면서 최소파형모형 검사를 한다. 최소파형모형 검사는 최소파형모형 검사와 프레임 내부 노이즈 스킵핑을 적용하여 검사한다. 최소파형모형 검사에서 반주기 파형의 샘플 개수를 카운트하며 이때 카운트 되는 샘플의 에너지가 설정한 $|TH_A|$ 값 이상이면 해당 샘플은 스킵핑 된다. 검사된 샘플의 개수가

최소파형모형을 이루는 TH_N 보다 많다면 반주기 파형은 최소파형모형이 되며 해당 프레임은 최소파형모형이 존재하므로 시작점 후보군 프레임이 된다. 시작점 후보군 프레임은 프레임 인덱스, +최소파형모형 개수, -최소파형모형 개수 등 세 가지의 정보를 담은 구조를 갖는다. 여기서 프레임 인덱스는 프레임의 연속된 번호를 의미하며, +와 -는 최소파형모형에서 반주기 파형이 가지는 부호를 의미한다.

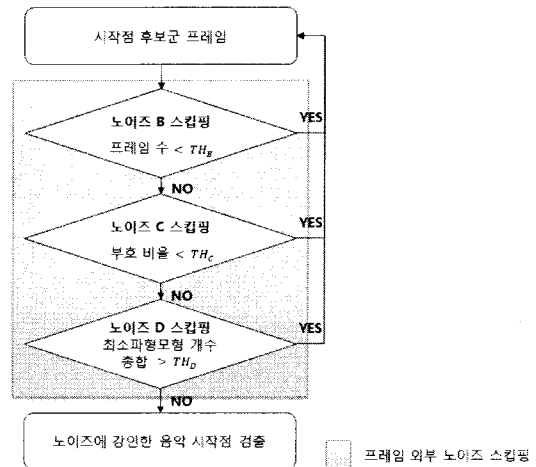


그림 6. 노이즈에 강인한 음악 시작점 검출 순서도
Fig. 6. Flow Chart of Music Starting Point Strong against Noise

도출된 시작점 후보군 프레임을 이용하여 노이즈에 강인한 음악 시작점을 검출하는 과정은 다음과 같다. 시작점 후보군 프레임에 외부 노이즈 B, C, D 스킵핑을 순서대로 적용하여 시작점을 검출한다. 프레임 외부 노이즈 스킵핑을 위하여 시작점 후보군 프레임의 구조가 이용된다. 먼저 노이즈 B 스킵핑에서는 프레임 인덱스를 이용하여 연속되어 있는 후보군 프레임을 검색한다. 설정한 TH_B 값 보다 연속되는 후보군 프레임의 수가 적다면 스킵핑 된다. 노이즈 C 스킵핑에서는 부호 비율을 구하기 위하여 +와 -의 최소파형모형 개수가 이용된다. 연속되는 프레임 마다 부호 별 최소파형모형 개수를 따로 합산하여 부호 비율을 계산한다. 계산된 값은 설정한 TH_C 값과 비교하여 스킵핑 여부를 결정한다. 노이즈 D 스킵핑에서는 최소파형모형 개수를 총 합한 것으로 설정한 TH_D 값과 비교하여 스킵핑 여부를 결정한다. 각 스킵핑 단계마다 프레임이 노이즈로 판단되면 시작점 검출에서 제외되며 다음 도출된 후보군 프레임을 다시 검사하여 스킵핑이 적용되지 않을 때까지 반복한다. 연속된 후보군 프레임이 시작점 프레임으로 판단되면 해당 첫 프레임의 인덱스는 음악의 시작점으로 검출된다.

IV. 실험 및 결과

실험에 사용된 음악 데이터베이스는 인터넷 음악 전문 사이트와 CD를 이용하여 국내외 아티스트 앨범단위 500곡을 무작위로 수집하였다. 수집된 음악 형식은 MP3 파일이며, 검색을 수행하기 위해 서버 군과 쿼리 군으로 나누었다. 서버 군의 음악은 수집하고 나서 변환되지 않은 상태로 다양한 음질과 채널을 가지고 있으며, 검색 수행 시 비교되는 군으로 특징을 추출하고 데이터베이스에 등록하였다. 쿼리 군의 음악은 64kbps의 음질과 스테레오 채널로 변환하였으며, 검색 수행 시 질의로 사용되는 군으로 데이터베이스에 등록이 되지 않는다. 검색 방법은 쿼리 군의 음악을 입력 데이터로 하여 데이터베이스에 등록된 서버 군의 음악과 비교를 수행하였다. 비교는 변환한 쿼리 500개의 곡과 서버 500개의 곡을 이용하였다. 검색의 결과는 특징간의 거리 값이 가장 적은 수치를 보이는 음악으로 하며 검색의 판단은 파일명을 비교하였다. 즉, 유사도가 가장 높은 파일을 검색하고 파일명이 같을 경우 검색 성공이며, 다르면 검색 실패로 나타내었다.

4.1 시작점 검출 성능 검증

본 논문에서 제안한 시작점 검출 알고리즘의 성능을 검증

하기 위하여 시작점을 검출하지 않고 검색한 결과와 시작점을 검출하고 검색한 결과를 비교하였다.

검색의 결과는 시작점 검출 알고리즘의 성능과 동일한 성능을 갖는다. 이를 증명하기 위하여 검색의 실패 요인 중 한 가지를 배제하고 실험하였다. 검색의 실패 요인은 두 가지이다. 첫째는 검색 알고리즘의 오류로 인한 실패이고, 둘째는 시작점 검출 오류로 인한 실패이다. 여기서 검색 알고리즘의 오류로 인한 검색의 실패를 배제시켜 오직 시작점 검출의 오류로 인한 검색실패만 나타나게 하였다. 따라서 서버 군 음악과 쿼리 군 음악의 시작점을 수동으로 정확히 맞추어 주고 검색의 결과가 100%임을 보이는 실험을 하였다. 이러한 실험 설정은 본 실험 수행 시 검색의 결과가 실패로 나타나는 경우 시작점의 불일치만을 의미하기 때문에 검색의 성공률과 시작점 검출의 성공률은 동일하게 된다.

실험결과 시작점 검출을 적용하지 않은 검색 결과는 57.4%, 시작점 검출을 적용한 검색 결과는 95.2%로 미적용한 검색 결과보다 37.8% 높은 성능을 보였다. 검색 결과는 시작점 검출의 성능과 동일하므로 본 논문에서 제시한 음악 시작점 검출 알고리즘의 성능은 95.2%이다.

검색 성능 향상의 주된 요인은 입력데이터 시작시간의 동일함이다. 쿼리 군과 같은 64kbps의 비트레이트 변환은 미세하지만 음악의 시작 시간을 다르게 한다. 비교 실험에 사용된 특징 데이터들은 시간의 순서로 도시-블록 거리 값을 다음 수

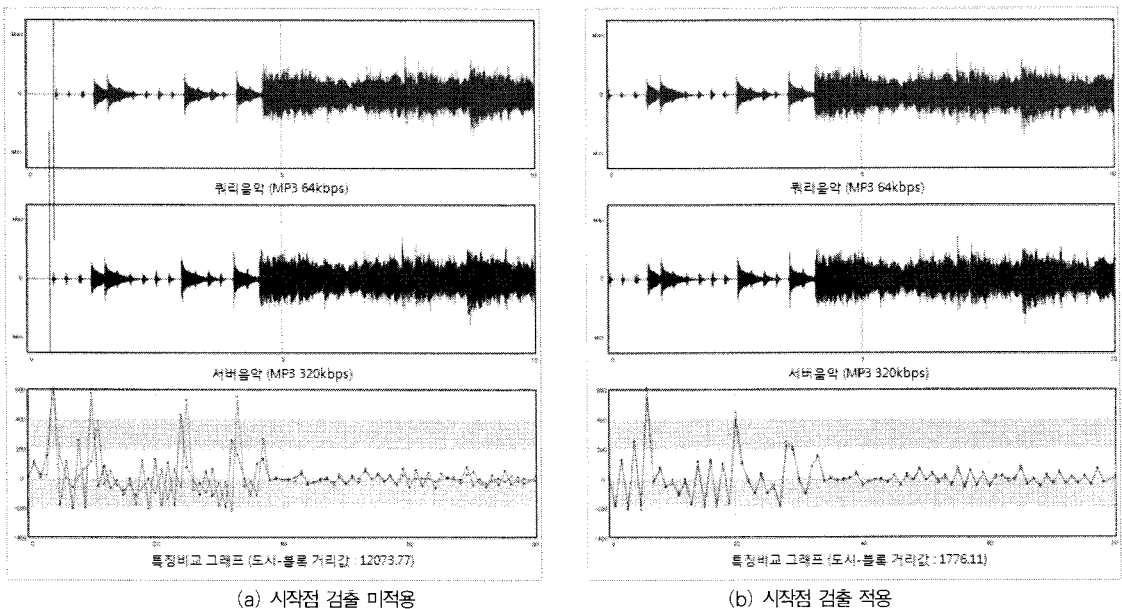


그림 7. 시작점 검출 알고리즘 적용 유무의 비교 실험 결과

Fig. 7. Results of comparison and experiment on whether starting point detection algorithm is applied

식(1)과 같이 구한다.

$$City-BlockDistance = \sum_{i=1}^N |S_i - Q_i| \dots\dots\dots (1)$$

여기서 S 는 데이터베이스에 등록된 서버 군의 특징벡터를, Q 는 쿼리 군의 특징벡터를, N 은 비교하는 특징벡터 개수를, i 는 특징벡터 개수에 대한 인덱스이다. 식(1)에서 시간의 변화는 검색에 영향을 미친다.

그림 7은 시작점 검출 알고리즘의 비교 실험을 보여준다. 시간의 단위로 한 개 이상 특징 데이터의 시간 오차가 발생하게 되면, 그림 7의 (a)와 같이 특징비교 그래프가 어긋나게 되고 도시-블록 거리를 계산하는 값도 증가하여 검색의 실패로 나타난다. (a)는 도시-블록 거리 값이 12073.77으로 측정되었으며, 검색실패로 나타났다. 이러한 특징 데이터의 시간 오차를 해결하고자 시작점 검출 알고리즘을 그림 (b)와 같이 쿼리음악과 서버음악에 적용하였다. (b)는 도시-블록 거리 값이 1776.11으로 (a)의 거리 값과 비교하였을 때 보다 약 7배 정도 낮은 수치를 보였으며, 검색성공으로 나타났다. 이처럼 입력데이터의 시작시간을 동일하게 함으로써 검색 성능이 향상되었다.

4.2 노이즈에 강인함 검증

노이즈에 강인한 시작점 검출을 검증하기 위하여 노이즈 스킵핑을 미적용한 시작점 검출과 노이즈 스킵핑을 적용한 시작점 검출을 비교 실험 하였다. 노이즈 스킵핑의 미적용은 최소파형모형 검사만을 이용하여 시작점 검출을 하였고, 노이즈 스킵핑의 적용은 최소파형모형 검사와 프레임 내부와 외부 스킵핑을 적용하여 검출 하였다. 그림 8은 노이즈 스킵핑을 적용하지 않고 최소파형모형만을 이용해 검출한 실험 결과를 나타낸다. 실험은 최소파형모형 N의 임계치 값을 측정하기 위하여 N을 1에서 10까지 순서대로 설정하였다. 그림 8에서 보듯이 N의 임계치 값이 3일 때 91.8%로 가장 높은 검출성공률을 보였다. 즉, 노이즈 스킵핑을 미적용한 시작점 검출의 성능은 91.8%이다.

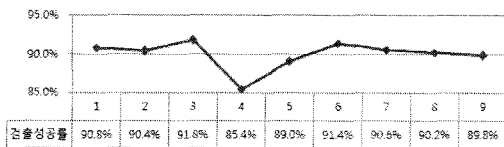


그림 8. 최소파형모형 TH_N 실험 결과
Fig. 8. Minimum Wave Model TH_N Experiment Results

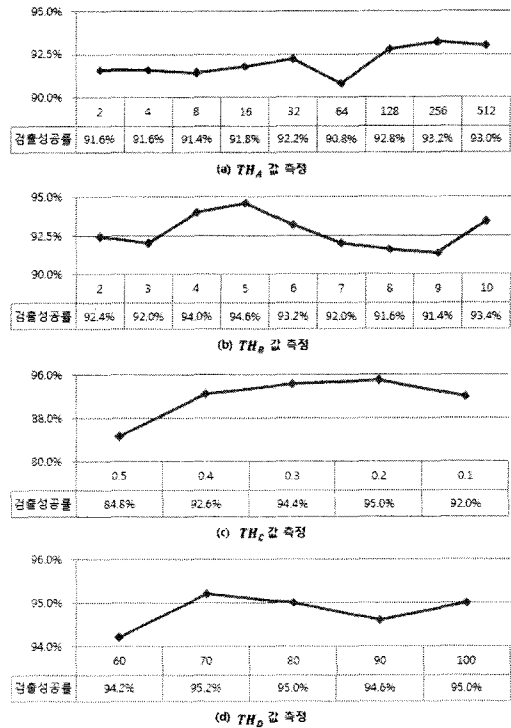


그림 9. 노이즈 임계치 값 실험 결과
Fig. 9. Noise Threshold Value Experiment Results

노이즈 스킵핑의 측정은 프레임 내부에서 외부로 노이즈 단계별로 수행하였다. 임계치 값을 결정하는 실험도 같은 단계로 수행하였다. 실험에서는 한 단계의 임계치 값이 결정되면 다음 단계에 이전 단계의 임계치 값을 반영하고 실험하였다. 노이즈 A에 해당하는 TH_A 값이 결정되면 TH_B 값을 결정하는 실험에 반영하였다. 마찬가지로 TH_C 값의 실험에서는 이전에 결정된 TH_A 와 TH_B 의 값을 모두 반영하여 실험하였다. 그림 9는 이러한 실험의 순서로 모든 노이즈 임계치 값에 대한 검출성공률을 보여준다.

그림 9에서 그래프의 x축의 값은 노이즈별 임계치 값을 의미한다. TH_A 는 2에서 512의 에너지 강도 값을 가지며, 에너지 강도의 세기는 이전 강도에서 2배씩 커진다. 이는 로그스케일을 취한 것으로 사람의 귀가 소리를 인지할 때 로그스케일 단위로 인지하기 때문이다. TH_B 는 2개 이상의 연속된 프레임을 2개에서 10개 사이의 값으로 실험하였다. TH_C 는 최소파형모형의 부호비율로 처음 0.5에서 0.1까지 비율 수치를 낮춰가면서 실험하였고, TH_D 는 최소파형모형의 개수로 60개부터 100개의 값으로 실험하였다.

표 2. 노이즈 스킵핑 실험 결과
Table 2. Noise Skipping Experiment Results

노이즈 스킵핑	미적용 TH_N	적용 TH_A	적용 TH_B	적용 TH_C	적용 TH_D
임계치 값	3	256	5	0.2	70
검출 성공률	91.8%	93.2%	94.6%	95.0%	95.2%

표 2와 같이 노이즈 스킵핑을 적용한 실험 결과는 다음과 같은 임계치 값에서 높은 검출성공률을 보였다. TH_A 값은 256의 설정 값에서 93.2%로 다른 임계치 설정 값보다 높게 나타났다. TH_B 값은 5개의 연속된 프레임을 측정하였을 때 검출성공률이 94.6%로 가장 높게 나타났다. TH_C 는 부호비율이 0.2에서 95.0%로 가장 높게 나타났으며, TH_D 값은 70 개로 설정하였을 때 95.2%로 가장 높게 나타났다. 이러한 결과는 노이즈 스킵핑 단계를 거칠 때마다 검출 성능의 향상됨을 보이며, 노이즈 스킵핑을 미적용한 결과보다 3.4% 향상으로 노이즈에 강인함을 검증하였다.

V. 결론

시간 순차적 검색 방법을 통한 내용기반 음악 검색은 비교하는 데이터 간 시작 시간이 동일하지 않으면 검색의 오류로 나타났다. 본 논문에서는 이러한 시작 시간의 오차를 검색의 전 단계에서 해결하고자 음악의 시작점을 검출하여 시작 시간을 일관성 있게 하였다. 음악의 시작점을 검출하기 위하여 최소파형모형을 프레임 단위로 검사하였고, 묵음 영역에 존재하는 노이즈는 스킵핑 하여 노이즈에 강인한 시작점을 검출하였다. 음악의 시작점을 시간 순차적 검색 방식에 적용시켜 본 결과, 시작점을 검출하지 않고 적용시킨 것 보다 37.8% 성능 향상이 있었다.

노이즈에 강인함을 검증하기 위하여 입력되는 음악에 노이즈 스킵핑을 적용하였다. 스킵핑의 적용은 묵음 영역에 존재하는 노이즈를 대상으로 하였으며 실험에 사용된 디지털 음악을 통해 4가지 유형으로 구분하였다. 각 노이즈 유형마다 특징을 분석하였고 최소파형모형을 이용한 스킵핑 조건을 두었다. 또한 조건에 따른 최적의 성능을 구하기 위하여 노이즈 조건에 따른 임계치 값을 설정하고 각각 실험하였다. 실험결과 노이즈 스킵핑 단계를 거칠 때마다 성능이 높아짐을 확인하였으며, 노이즈 스킵핑 단계를 거쳐 총 3.4% 성능 향상이 있었다. 따라서 제안한 시작점 검출 알고리즘은 노이즈에 강

인하다.

제안한 알고리즘은 음악을 이용한 분석시스템 설계 시 전처리 단계로 유용하게 사용될 것으로 기대된다. 처리에 필요하지 않은 묵음영역의 계산 낭비를 막게 되며, 후행 되는 분석시스템의 인식 향상을 가져오기 때문이다. 본 논문의 실험은 한정된 음악 내에서 수행되어 제안한 노이즈의 특성 이외에 다른 특징을 보이는 노이즈가 존재할 가능성이 있다. 따라서 시작점 검출 알고리즘에 있어서, 노이즈에 강인함을 더하기 위하여 묵음영역에 존재하는 노이즈 유형의 구분 및 분석을 향후 과제로 남긴다.

참고문헌

- [1] 정명범, 성보경, 고일주, "내용 기반 음악 검색의 문제점 해결을 위한 전처리," 한국컴퓨터정보학회논문지, 제 12권, 제 6호, 97-104쪽, 2007년 12월.
- [2] B.-K. Yi, H.V. Jagadish, and C. Faloutsos, "Efficient Retrieval of Similar Time Sequences under Time Warping," In Proceedings of ICDE, pp.201-208, February 1998.
- [3] J. Listgarten, R. M. Neal, S. T. Roweis, and A. Emili, "Multiple alignment of continuous time series," In Advances in Neural Information Processing Systems 17, pp.817 - 824, 2005.
- [4] 이해환, 심규석, 박형민, "HummingBird: 향상된 스케일드엔워프 매칭을 이용한 유사 음악 검색 시스템," 한국정보과학회논문지, 제 34권, 제 5호, 409-419쪽, 2007년 10월.
- [5] A.W.Fu, E.J.Keogh, L.Y.H.Lau and C. Ratanamahatana "Scaling and Time Warping in Time Series Querying," In Proc. of the VLDB Conference, pp.649-660, 2005.
- [6] Waiyawuth Euachongprasit and Chotirat Ratanamahatana, "Accurate and efficient retrieval of multimedia time series data under uniform scaling and time warping," pp.100 - 111, 2008.
- [7] Dixon, S., Widmer, G., "MATCH: a music alignment tool chest," Proceedings of the International Conference of Music Information Retrieval, pp.492 - 497, 2005.
- [8] M. M. Ruxanda and C. S. Jensen, "Efficient similarity retrieval in music databases," In

Proc. of COMAD, pp.56-67, 2006.

- [9] Shannon B.J., Paliqi K.K., "A comparative study of filterbank spacing for speech recognition," Proc. International Microelectronic engineering research conference, Brisbane, Austria, 2003.
- [10] F. Zheng, G. Zhang, "Integrating the energy information into MFCC," Proc. International Conference of Spoken Language Processing, Beijing, China, 2000.
- [11] Z. Jun, S. Kwong, W. Gang, Q. Hong, "Using Mel-Frequency Cepstral Coefficients in Missing Data Technique," EURASIP Journal on Applied Signal Processing, pp.340-346, 2004.
- [12] 양병곤, "인간의 청각 척도에 관한 고찰," 음성과학, 제 2권, 125-134쪽, 1997년.

저자 소개



김 정 수
 2008: 숭실대학교 미디어학부 (공학사)
 2008 - 현재:
 숭실대학교 미디어학과 석박사 통합과정
 관심분야: 콘텐츠, 오디오 신호처리,
 내용기반 검색



성 보 경
 2006: 숭실대학교 미디어학부 (공학사)
 2007: 숭실대학교 미디어학과 (공학석사)
 2007 - 현재 :
 숭실대학교 미디어학과 박사과정
 관심분야: 디지털음악 분석, 음악검
 색, 내용기반 학습



구 광 호
 1999: 국립경상대학교 한문학과
 (문학사)
 2005 - 현재:
 숭실대학교 미디어학과 석박사 통합과정
 관심분야: 추천시스템, 미디어인식, 정
 보검색



고 일 주
 1992: 숭실대학교 전산학과 (공학사)
 1994: 숭실대학교 전산학과 (공학석사)
 1997: 숭실대학교 전산학과 (공학박사)
 2003 - 현재 :
 숭실대학교 미디어학부 조교수
 관심분야: 콘텐츠, 영상처리, 감성공학