

## 최소파형모형을 이용한 음악 시작점 검출 방법

김정수\*, 정보경\*, 구광효\*, 고일주\*

### Music Starting-point Detection Method using MinWaveShape

Bung-Soo Kim\*, Bo-Kyung Sung\*, Kwang-Hyo Koo\*, Il-Ju Ko\*

#### 요 약

객관화된 메타정보를 이용하는 검색방법과는 다르게 내용기반 검색에서는 전처리된 데이터가 동일하지 않을 수 있다는 문제점이 있다. 특히 디지털 음악데이터의 경우 인코딩과정을 거칠 때 마다 미세하지만 파형의 변화가 생긴다. 이러한 변형은 타임코드를 쉬프트 시켜 동일한 데이터 검색에 어려움을 발생시킨다. 본 논문에서는 이러한 문제를 해결하고자 음악의 시작점을 검출 할 수 있는 방법을 제안하였다. 사람이 소리를 인지하는 원리는 공기가 진동하여 청각기관을 자극하기 때문이며 공기의 진동은 파동을 형성한다. 파동을 구성하는 최소파형모형의 존재 여부를 검사하여 음악의 시작점을 검출 하였다. 녹음환경 및 디지털 압축 과정으로 음을 구성하는 파형에 노이즈가 포함될 경우 음악의 시작점 검출에 방해 요인이 된다. 노이즈의 영향을 받지 않고 음악의 시작점을 검출하기 위해 노이즈가 포함된 파형의 특징을 분류하고, 이 분류를 예외 조건을 두어 해결하였다. 제안한 방법의 성능을 측정하기 위해 600개의 음원을 실험 하였으며 86%의 일치율을 보였다.

▶ Keyword :음악시작점검출, 최소파형모형, 노이즈처리

---

• 제1저자 : 김정수  
\* 숭실대학교 미디어학과

## I. 서론

현재는 컴퓨터와 통신의 발달로 멀티미디어 정보가 급속도로 확산되고 있으며, 음악 데이터를 이용한 분야 역시 그 쓰임의 다양함이나 활용 범위가 방대해짐에 따라 이미지, 비디오 등의 정보 부분과 함께 멀티미디어의 독립적인 한 분야로 다루어지게 되었다. 따라서 방대한 음악 데이터를 관리하기 위한 데이터베이스를 구축하며 이를 바탕으로 빠르고 정확하게 음악을 검색할 수 있는 방법이 필요하다. 기존의 음악 검색은 타이틀이나 아티스트와 같은 객관화된 메타 데이터 정보를 이용한 검색이 주로 이용되었으나 찾고자 하는 음악의 메타 정보를 반드시 사전에 알고 있어야 검색이 가능하다는 한계가 있어 최근 내용기반 음악 검색 시스템의 필요성이 대두되어 많은 연구가 진행되고 있다[1][2].

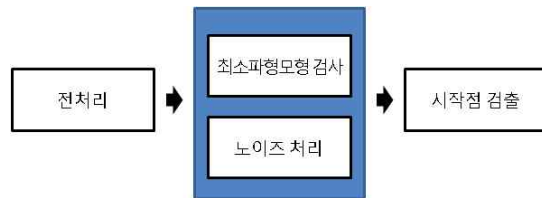
내용기반 음악 검색 시스템은 메타 데이터에 의존하던 시스템과 달리 음악 정보의 내용(contents)을 수학적으로 분석하여 유사도가 높은 음악 데이터를 찾아주는 시스템이다[3]. 이러한 시스템으로 동일한 디지털 음악을 검색하면 전처리된 데이터가 동일하지 않을 수 있다는 문제점이 있다. 특히 디지털 음악데이터의 경우 인코딩 과정을 거칠 때 마다 미세하지만 파형의 변화가 생긴다. 이러한 변형은 타임코드를 쉬프트시켜 동일한 데이터 검색에 어려움을 발생 시킨다[4]. 본 논문에서는 이러한 디지털 음악의 시간차 문제를 해결하고자 최소파형모형을 이용한 음악의 시작점 검출 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 시작점 검출을 하기위해 최소파형모형 이론과 노이즈 처리 방식을 적용한 방법을 자세히 소개한다. 3장에서는 제시한 알고리즘의 성능을 검증하기 위해 실험을 수행하였으며 마지막으로 결론으로 끝을 맺는다.

## II. 시작점 검출 방법

시작점을 검출하기 위한 기본 아이디어는 음악에서 음이 존재하는 첫 번째 점을 찾는 것이다. 음의 존재 여부 판단은 음을 최소로 구성할 수 있는 최소파형모형의 존재를 보고 결정한다. 하지만 단순히 최소파형모형만을 이용하면 음악의 시작하기 전인 묵음 부분에 노이즈가 있을 경우 방해 요인이 되어 올바른 시작점을 검출하기 힘들다. 이러한 노이즈 처리를 하기 위해 디지털 음악이 가질 수 있는 노이즈 유형을 구분하고 각 유형의 특징을 이용한다.

본 논문에서 제시하는 음악의 시작점 검출 방법 흐름은 그림 1과 같다. 처음으로 디지털 음악을 입력하면 전처리 과정을 거친다. 전처리 과정에서는 압축된 음원을 PCM으로 디코딩하며 스테레오 채널을 모노 채널로 바꾸어 준다. 그 다음 시작점을 검출하기 위해 최소파형모형 검사와 노이즈 처리를 조합하여 사용한다.



<그림 1> 음악 시작점 검출 방법 흐름

### 2.1. 음을 구성하는 최소파형모형

사람이 소리를 인지하는 원리는 공기가 진동하여 청각기관을 자극하기 때문이며 공기의 진동은 파동을 형성한다. 파동은 매질 내의 한 점에서 생긴 매질의 진동 상태가 매질을 통해서 주기적으로 퍼져나가는 현상을 말하며 공간상의 한 점에서 서로 순환적으로 변환되는 에너지가 존재하게 된다. 파동의 주기는 일반적으로 사인파처럼 규칙적으로 퍼져 나가지만 음악의 파동은 다양한 악기의 소리와 음성이 합성되어 있기 때문에 불규칙한 주기의 파동을 가지고 있으며 파동의 에너지는 소리의 세기를 나타낸다.

최소파형모형은 음악의 불규칙한 주기를 가지는 파동에서 한 부호의 에너지 값을 가지는 반주기 파형을 나타내며, 부호의 에너지 값은 파형의 주기에서 0점을 기준으로 음수와 양수의 값을 의미한다. 반주기 파형은 바로 최소파형모형이 될 수 없고 반주기 파형을 형성하는 샘플의 개수가 일정 수 이상으로 커야 최소파형모형이 된다. 이후 최소파형모형을 형성할 수 있는 최소 샘플의 개수는  $N$ 으로 표현한다. 그림 2에서는  $N$ 을 5로 설정하였을 때  $N$ 보다 큰 반주기를 가지는 최소파형모형이 7개가 존재하는 것을 보여준다. 여기서 사용된 샘플의 단위는 아날로그 음악의 파동을 디지털 음악으로 변환하기 위해 시간에 따른 에너지를 표본화한 수치이며 보통 CD음질의 음원인 경우 초당 44100의 샘플을 갖는다. 이렇게 반주기 파형으로 형성된 최소파형모형은 음을 구성하는 최소 단위로 사용된다.



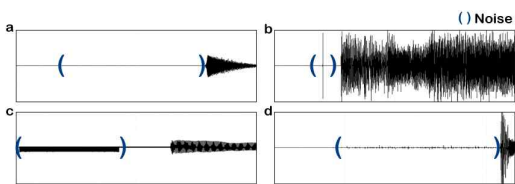
<그림 3> N보다 큰 반주기를 가지는 최소파형모형 (N=5)

최소파형모형이 음일 확률을 높여주기 위해 최소파형모형 검사는 샘플 단위로 수행을 하지 않고 샘플의 묶음인 프레임 단위로 수행한다. 샘플 단위로 수행을 하면 최초 하나의 최소파형모형을 보고 시작점을 결정하지만 프레임 단위로 수행하게 되면 프레임 안에 속한 최소파형모형을 전부 보고 결정하기 때문에 음일 확률이 높아진다. 프레임 단위 설정은 사람이 소리를 최소한으로 인식할 수 있는 단위로 음성인식 분야에서 많이 사용된 20ms (880sample)로 하였다. 시작점을 검출하기 위해 프레임 단위로 최소파형모형이 존재하는지 검사하며 첫 번째로 찾은 프레임이 음악의 시작이다.

## 2.2. 노이즈 처리

노이즈는 소음 또는 잡음을 의미하며 대개 녹음을 하거나 아날로그 음원을 디지털로 변환하는 과정에서 발생한다. 디지털 음원에 노이즈가 포함될 경우 노이즈도 소리를 구성하는 파동과 에너지를 가지고 있어 최소파형모형이 존재한다. 그래서 노이즈가 포함된 디지털 음원은 노이즈를 시작점으로 찾게 되는 문제가 발생한다.

노이즈의 영향을 받지 않고 음악의 시작점을 검출하기 위해 노이즈가 포함된 파형의 특징을 분류하고 이 분류를 이용한 예외 조건을 두어 해결한다. 실험에 사용된 음악을 조사하여 노이즈의 유형을 구분하였으며 그림 3과 같이 디지털 음악 앞부분에 생성된 노이즈 파형은 크게 4가지 유형으로 나타난다.



<그림 3> 노이즈 파형의 특징

(a)와 같은 노이즈 파형의 경우는 그림으로 보았을 때 노이즈의 식별이 불가능하지만 시작점을 검출하기 위해 최소파형모형의 존재를 검사하면 묵음 부분에 시작점을 검출하게 된다. 이는 귀로 인지하기 힘들 정도로 낮은 에너지의 파형이

묵음 부분에 존재하기 때문이다.

(b)와 같은 노이즈 파형의 경우는 귀로 인지할 수 있는 에너지를 가지며 시간 축으로 고루 분포되지 않고 단 한 점에서 발생하는 노이즈이다.

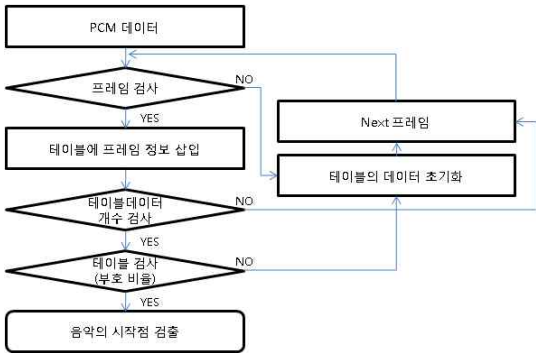
(c)와 같은 노이즈 파형의 경우는 귀로 인지할 수 있는 에너지를 가지며 시간 축으로 길게 연장된 특징을 보이고 에너지의 값이 특정 한 부호의 값으로 치중되어 있다.

(d)와 같은 노이즈 파형의 경우는 귀로 인지할 수 있는 에너지를 가지며 시간 축으로 고루 분포 되어 있다.

위와 같은 특징을 이용하여 노이즈의 영향을 받지 않고 시작점을 검출하기 위해서는 다음과 같이 한다. (a)의 노이즈는 최소파형모형을 결정짓는 N을 측정할 때 에너지가 낮은 샘플이라면 해당 샘플을 N의 개수에 포함 시키지 않으므로써 낮은 에너지를 갖는 노이즈의 영향을 받지 않게 한다. (b)의 노이즈는 프레임 검사를 할 때 한 개의 프레임만 검사를 하는 것이 아니라 두 프레임 이상을 연달아 최소파형모형이 존재하는지 검사하여 처리한다. 일반적으로 음악이 재생되면 파형은 시간 축으로 계속 존재하게 되지만 (b)와 같은 노이즈는 파형이 한 점에서만 존재하고 다른 부분에서는 파형이 존재하지 않기 때문에 연달아 검사하면 노이즈인지 구분 할 수 있다. (c)의 노이즈는 최소파형모형을 프레임 단위로 검사할 때 음의 부호 값을 가지는 최소파형모형 수와 양의 부호 값을 가지는 최소파형모형 수를 세고, 부호의 비율이 한쪽으로 치중되어 있다면 노이즈로 보고 처리한다. (d)의 노이즈의 경우는 파형이 고루 분포되어 있고 사람이 인지할 수 있는 에너지의 값을 가지고 있어 실제 음을 가지는 파형의 모형과 유사하다. 그래서 (d)의 노이즈는 최소파형모형 특성을 이용하여서는 처리하기 힘들다.

## 2.3. 음악의 시작점 검출

음악의 시작점을 검출하기 위해 최소파형모형의 검사와 노이즈 파형을 처리하기 위해 프레임을 검사하였다. 이 때 최소파형모형과 노이즈는 각각 따로 검사하는 것이 아니라 조합하여 검사한다. 그림 4는 조합된 알고리즘의 순서도를 나타내며 순서의 구성을 살펴보면 다음과 같다.



<그림 4> 최소파형모형 검사와 노이즈 처리가 조합된 알고리즘 순서도

먼저 PCM 데이터를 프레임 단위로 검사한다. 첫 번째 프레임부터 검사를 수행하며 조건에 맞지 않을 경우 다음 프레임으로 넘어 간다. 프레임을 검사 할 때는 최소파형모형의 존재 여부를 살펴봄과 노이즈(a)의 처리를 같이 수행한다. 이렇게 검사하여 나온 결과가 TRUE 이면 프레임 검사 테이블에 프레임 샘플 위치와 부호에 따른 최소파형모형의 수를 삽입한다. 테이블에 데이터를 삽입하는 중 프레임 검사가 FALSE 이면 테이블의 데이터는 초기화 된 후 다음 프레임부터 다시 검사한다. 이처럼 프레임 검사 테이블의 데이터는 노이즈(b)의 처리를 위해 검사 결과가 연달아 만족할 때까지 담는다. 표 1은 프레임 검사 테이블에 삽입된 결과를 보여준다. 테이블에 데이터가 전부 채워지면 마지막으로 양의 파형수와 음의 파형 수의 비율을 검사하고 만족한다면 삽입된 테이블의 프레임 위치 첫 번째 값이 음악의 시작점이 된다.

<표 1> 프레임 검사 테이블

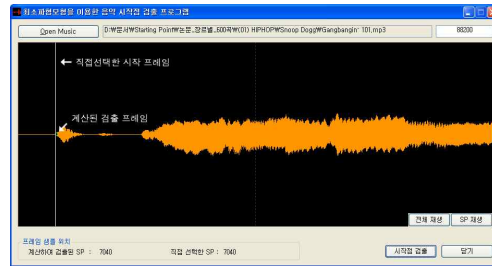
프레임 시작 위치 (단위 : 샘플)	음의 부호를 갖는 최소파형모형 개수	양의 부호를 갖는 최소파형모형 개수
4400	5	4
5280	7	9
6160	8	6

### III. 실험 & 결과

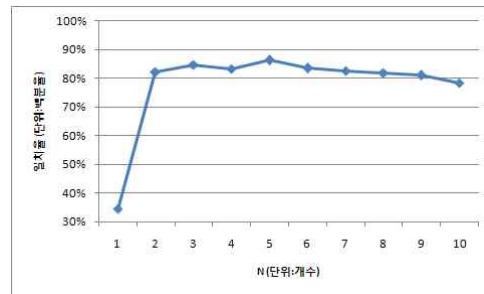
실험에서 음악의 시작 검출을 위해 사용한 디지털 음원은 44100 hz, 16bit로 샘플링 되었고, 제안한 방법의 성능을 측정하기 위해 600개의 음악으로 실험하였다.

실험은 최소파형모형을 형성 할 수 있는 N의 개수를 알기 위해 N을 1개에서 10개까지 순서대로 설정하고, 개수마다

측정한 시작 위치를 프레임 단위로 검출 하였다. 검출된 결과는 직접 선택한 시작 프레임과 비교를 하였으며, 정확한 프레임 단위로 선택하고자 그림 5와 같은 응용프로그램을 제작 하였다. 응용프로그램에서는 음악의 시작 파형을 확대해서 보여주며 시작점을 프레임 단위로 직접 선택하고 비교할 수 있는 기능을 가진다. 본 논문에서 제안한 시작점 검출 방법의 정교한 성능을 검증하기 위해 검출 오차의 범위는 음을 최소한으로 인식할 수 있는 단위인 1프레임(20ms)으로 하였다.



<그림 5> 시작점 추출 프로그램 실행의 예



<그림 6> N 개수에 따른 성능 측정 그래프

위의 그림 6은 최소파형모형을 가질 수 있는 샘플의 개수 N에 따른 성능 측정 그래프 이다. 그래프의 가로축은 N의 개수를 나타낸 것이며 세로축은 직접 선택한 검출 프레임과 비교해서 나온 일치율이다. N의 개수가 1개로 이루어진 파형은 결과에서도 알 수 있듯이 매우 낮은 일치율을 보이며 N의 개수가 2개 이상으로 이루어진 파형은 높은 일치율을 보인다. 즉 N이 1개로 구성된 반주기 파형은 최소파형모형이라 볼 수 없고 적어도 2개의 샘플로 구성된 반주기 파형을 최소파형모형이라 볼 수 있다. 그리고 6개 이후부터는 점점 일치율이 낮아진다. 이런 점들을 고려하여 최소파형모형을 형성하는 최소 샘플의 개수는 2개에서 6개 사이이다. 본 논문에서는 가장 일치율이 높게 나온 N=5를 최소파형모형으로 결정하였고 86%의 일치율을 보였다.

#### IV. 결론

본 논문에서는 내용기반 음악 검색에서 디지털 음원간의 시간차 문제를 해결하고자 음악의 시작점을 검출하는 방법을 제안하였다. 시작점을 검출하기 위해 음을 구성하는 최소단위인 최소파형모형을 이용하였고 노이즈의 영향을 받지 않고 검출하기 위해 특징을 분류하고 특징에 따른 예외 조건을 두어 해결하였다.

논문에서 제시한 시작점 검출 방법은 86%의 높은 일치율을 보였으며, 본 연구의 결과로 내용기반 음악 검색 시스템에서 성능을 향상시킬 수 있을 것으로 기대된다.

#### 참고문헌

- [1] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney, "Content-Based Music Information Retrieval: Current Directions and Future Challenges", *Proceedings of the IEEE*, Vol. 96, 2008.
- [2] Lihui Guo, Xin He, Yaxin Zhang, Yue Lu, "Content-based retrieval of polyphonic music objects using pitch contour", *Acoustics, Speech and Signal Processing*, 2008.
- [3] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio", *IEEE Multimedia*, 3(2), 1996.
- [4] Bo-Kyung Sung, "A method of same music extracting between digital music by using similarity measurement", 2007.