특집논문-06-11-1-01

음악요약 생성에 관한 연구

김 성 탁[a)‡], 김 상 호[a)], 김 회 린[a)], 최 지 훈[b)], 이 한 규[b)], 홍 진 우[b)]

A Study on Music Summarization

Sungtak Kim[a)‡], Sangho Kim[a)], Hoirin Kim[a)], Ji Hoon Choi[b)], Hankyu Lee[b)], and Jinwoo Hong[b)]

요 약

음악요약이란 주어진 음악 컨텐츠에서 가장 중요하고 특징적인 한 부분이나 여러 부분들을 제공하는 것을 말한다. 음악요약 기술에는 크게 두 가지 종류의 음악요약을 위한 기술들이 연구되고 있다. 음악 컨텐츠내에서 반복되는 구간을 음악요약으로 제공하는 기술과 특징이 다른 부분들의 일정구간을 모두 제공하는 기술이 있다. 본 논문에서는 두 가지 종류의 음악요약을 제공하는 알고리즘들을 제안하고 평가하였다. 반복되는 구간을 음악요약으로 제공하는 다중 레벨 벡터양자화를 이용한 알고리즘은 고정된 길이와 최적의 길이를 가지는 음악요약을 제공하는 알고리즘들을 객관적인 방법으로 성능을 평가 하였고, 음악 내에서 특징이 다른 부분들을 일정부분씩 취합하여 제공하는 2-D 유사도행렬과 k-mean 알고리즘을 이용하는 집단화 방법을 이용한 방법의 평가는 주관적인 평가인 MOS 테스트로 평가하였다. 다중 레벨 벡터양자화를 이용한 음악요약을 제공하는 알고리즘에서 고정된 길이의 음악요약을 제공하는 알고리즘은 사람이 직접 요약한 결과와 제안한 방법으로 구한 요약과의 중첩도(Overlapping Ratio)를 이용한 결과 기존의 방법들이 42.2%와 47.3%임에 비해 제안된 방법은 67.1%로 높은 성능을 보여주었고, 최적의 길이를 가지는 음악요약을 제공하는 알고리즘은 음악에 따라 다른 길이를 가지는 반복되는 부분의 포함 정도를 나타내는 최적 중첩비율(Optimal Overlapping Ratio)을 측정한 결과 고정된 길이를 가지는 음악요약 보다 최적의 길이로 음악마다 다른 길이의 반복되는 부분을 효과적으로 표현함을 알 수 있었다. 집단화 방법을 이용한 알고리즘은 두 가지 질문들(제공된 세그먼트들 중 특징이 비슷한 것의 개수, 제공된 세그먼트들 중 같은 구조에 속하는 것의 개수)을 이용한 MOS 테스트에서 우수한 결과를 보여주었다.

Abstract

Music summarization means a technique which automatically generates the most importantand representative a part or parts in music content. The techniques of music summarization have been studied with two categories according to summary characteristics. The first one is that the repeated part is provided as music summary and the second provides the combined segments which consist of segments with different characteristics as music summary in music content. In this paper, we propose and evaluate two kinds of music summarization techniques. The algorithm using multi-level vector quantization which provides a repeated part as music summary gives fixed-length music summary oroptimal length music summary. Fixed-length music summary is evaluated by overlapping ratio between hand-made repeated parts and automatically generated summary. As results, the overlapping ratios of conventional methods are 42.2% and 47.4%, but that of proposed method with fixed-length summary is 67.1%. Optimal length music summary is evaluated by the portion of overlapping between summary and repeated part which is different length according to music content and the result shows that automatically-generated summary expresses more effective part than fixed-length summary with optimal length. The cluster-based algorithm using 2-D similarity matrix and k-means algorithm provides the combined segments as music summary. In order to evaluate this algorithm, we use MOS test consisting of two questions (How many similar segments are there in the summarized music?, How many segments are included in same structure?) and the results show good performance.

Keyword: 음악요약, 다중레벨 벡터양자화, 2-D 유사도 행렬, k-means 알고리즘

a)한국정보통신대학교 공학부
School of Engineering, Information and Communications University
‡ 교신저자 : 김성탁(stkim@icu.ac.kr)

b)한국전자통신연구원 디지털방송연구단 방송미디어연구그룹
Digital Broadcasting Research Division, ETRI

Ⅰ. 서 론

인터넷과 저장기술의 발달로 방대한 양의 멀티미디어 데이터의 저장이 가능해짐에 따라 이와 함께 제공되는 방대한 디지털 멀티미디어 컨텐츠를 소비할 수 있는 PDA, 데스크탑 컴퓨터(Desktop PC), 노트북, 핸드폰등과 같은 다양한 종류의 단말기기가 사용되고 있으며, 더욱이 이러한 단말기기 간의 통신이 가능하게 되었다. 따라서 다양한 종류의 멀티미디어 단말기기를 통해 멀티미디어 컨텐츠를 자유롭게 소비할 수 있는 필요성이 대두되고 있다. 이런 필요성으로 인해 사용자에게 해당 컨텐츠의 특징을 간결하게 나타낼 수 있는 기술의 필요성이 요구되고, 최근 컨텐츠들을 효과적으로 요약하는 기술이 중요해지고 있다. 예를 들면 영화의 예고편이나 영화의 개관(Review), 책의 개관, 그리고 논문의 요약등은 전체 내용을 자세히 나타내지 않고 중요정보만 제공하는 것이다. 지금까지 텍스트나 동영상의 요약을 제공하는 기술들[1~4]은 지금까지 많은 연구가 있었다. 하지만, 텍스트나 동영상에 비해 음악 컨텐츠는 복잡한 구조를 가지는1.차원신호들의 집합이므로 음악 컨텐츠의 요약 기술 개발에는 많은 어려움이 있어왔다. 음악요약은 주어진 음악에서 가장 중요하고 특징적인 부분이나 부분들을 제공하는 것이다. 현재 음악 컨텐츠 제공자들은 구매자가 컨텐츠를 구매하기 전에 해당 컨텐츠의 미리듣기 서비스를 통해 음악 컨텐츠의 정보를 제공하고 있다. 하지만 대부분 음악 컨텐츠의 처음부분의 일정구간(30초~1.분)을 제공하는데 음악 컨텐츠의 내용을 고려하지 않은 이런 획일적인 미리듣기 제공으로는 구매자가 해당 컨텐츠의 특징을 알기가 어렵고, 또한 구매자의 구매욕구를 자극하기가 쉽지 않다. 하지만 음악 컨텐츠의 특징을 나타내는 음악요약 기술을 이용한 미리듣기를 제공한다면 구매자는 음악 컨텐츠의 특징을 쉽게 알 수 있고 구매자의 구매욕구를 높이는데 많은 기여를 할 것이다. 지금까지 연구되고 있는 음악요약 제공 기술은 크게 두 가지 방식으로 나누어 진다. 주어진 음악 컨텐츠내에서 자주 반복이 일어나는 부분을 제공하는 방식과 음악 컨텐츠내에서 다른특징(무드)을 가지는 부분들을 일정구간씩 모두 제공하는 방식이 있다. 음악내에서 반복되는 구간을 음악요약을 제공하는 기술은 일반적으로 길이가 짧은 음악에 유용하고, 긴 음악의 경우는 대부분 음악내에 여러 가지 특징(무드)이 존재할 가능성이 높으므로 다른 특징을 모두 제공하는 기술이 유용하다.

본 논문에서는 두 가지 방식의 음약요약 기술들을 모두 제안하고 각각을 평가하였다. 자주 반복이 일어나는 부분을 음악요약으로 제공하는 기술에서는 다중레벨 벡터양자화(Multi-level vector quantization)를 통해 음악 컨텐츠의 양자화 코드워드 정보를 이용해서 반복되는 부분을 찾아내고, 특징이 다른 부분들의 일정 구간을 모두 제공하는 방법에서는 2-D 유사행렬(2-D Similarity matrix)[5]와 k-mean 알고리즘을 이용해서 음악 컨텐츠내에서 특징이 바뀌는 부분을 찾아낸다.

Ⅱ. 반복되는 부분을 음악요약으로 제공하는 방법

그림 1은 반복되는 부분을 음악요약 제공하는 방법의 개념도이다. 그림 1을 보면 후렴(Chorus) 부분이 반복되고 반복되는 부분인 후렴부분을 음악요약으로 제공하는 방법이다.

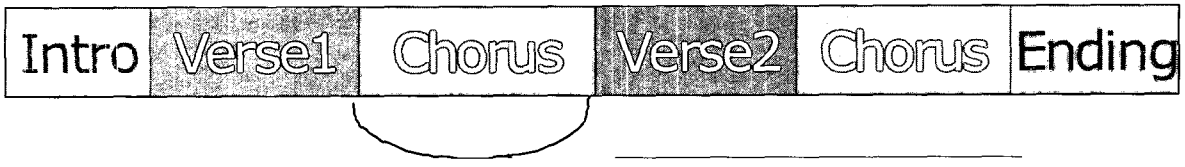


그림 1. 반복되는 부분을 음악요약으로 제공하는 방법
Fig. 1. Method of music summarization that is a repeated part

그림 2는 반복되는 부분을 음악요약을 제공하는 시스템의 블록도이다.

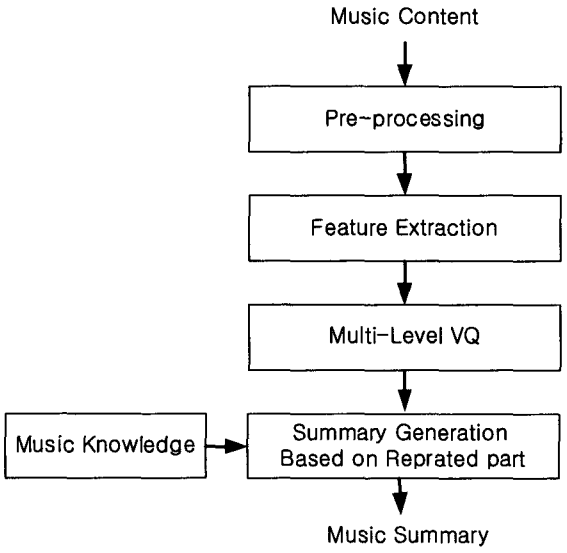


그림 2. 다중레벨 벡터 양자화를 이용한 음악요약 시스템 블록도
Fig. 2. Block diagram of music summarization using multi-level vector quantization

1. 전처리

음악 요약을 위한 특징벡터들을 추출하기 전에 고정된 길이를 갖고 50% 중첩이 되는 프레임(Frame)들로 분할하고 묵음 구간을 제거한다. 묵음을 제거하는 방법은 미리 정한 임계값과 프레임 에너지를 비교해서 작으면 제거한다.

2. 특징벡터 추출

본 논문에서 음악요약을 위해spectral power, amplitude envelope, 그리고 MFCC(Mel-Frequency Cepsral Coefficient)를 특징벡터로 사용하였다.

2.1 Spectral Power

주어진 음악 신호 $s(n)$에 대해 각 프레임에 아래의 Hanning 윈도우 $h(n)$를 적용한다.

$$h(n)=\frac{\sqrt{8/3}}{2}\left[1-\cos\left(2\pi\frac{n}{N}\right)\right] \quad (1)$$

여기서 N은 프레임내의 샘플 개수이다.
Spectral Power를 구하는 식은 아래와 같다.

$$S(k)=10\log_{10}\left[\frac{1}{N}\left|\sum_{n=0}^{N-1}s(n)h(n)\exp\left(-j2\pi\frac{nk}{N}\right)\right|^2\right] \quad (2)$$

2.2. Amplitude Envelope

Amplitude Envelope.은 시간영역에서 에너지의 변화를 나타낸다. 에너지의 변화는 음악에서 ADSK(Attack, Decay, Sustain, Release)와 같은 의미를 나타낸다. 음악 신호의 Amplitude Envelope.은 각 프레임마다 아래의 식을 이용해서 구한다.

$$RMS=\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x[n]^2} \quad (3)$$

여기서 $x[n]$는 차단주파수가 1200Hz인 저역통과 필터를 통과한 신호이다.

2.3 MFCC(Mel-Frequency Cepstral Coefficient)

음성인식의 가장 대표적인 특징 추출 방법인 MFCC 추출방법으로 사람의 귀가 주파수 변화에 반응하게 되는 양상이 선형적이지 않고 로그스케일과 비슷한 멜 스케일을 따르는 청각적 특성을 반영한 켑스트럼 계수 추출 방법이다. 멜 스케일에 따르면 낮은 주파수에서는 작은 변화에도 민감하게 반응하지만, 높은 주파수로 갈수록 민감도가 작아지므로 특징 추출시에 주파수 분석 빈도를 이와 같은 특성에 맞추는 방식이다.

MFCC를 구하기 위해서는 우선 분석구간의 음성 신호에 푸리에 변환을 취하여 스펙트럼을 구한다. 구한 스펙트럼에 대해 멜 스케일에 맞춘 삼각 필터뱅크를 대응시켜 각 밴드에서의 크기의 합을 구하고, 필터뱅크 출력값에 로그를 취한다. 그리고, 로그를 취한 필터 뱅크 값에 이산 코사

인 변환을 하여 최종 MFCC를 구한다.

$$c_n = \sqrt{\frac{2}{K}\sum_{k=0}^{K}(\log S_k)\cos[n(k-0.5)\pi/K]} \quad (4)$$
$$n = 1,2,...,L$$

여기서 S_k는 필터팽크의 출력값을 나타낸다. 실험에서는 $K=38$ 그리고 $L=25$로 하였다.

3. 반복부분 기반 음악요약 생성

3.1. 길이가 고정된 음악요약 생성

만약 주어진 음악의 노트(Note)를 정확히 표현할 수 있다면 음악요약을 하는 것은 쉬운 일이다. 하지만 대부분의 음악들이 다성음악(Polyphony)의 형태를 가지고 있으므로 노트들을 표현하는 것은 어려운 일이다.

본 논문에서는 음악요약을 위해 음의 노트 대신 다중 레벨 벡터양자화 (Multi-Level VQ) 결과들을 이용하여 주어진 음악을 표현하였고, 벡터양자화 결과들을 이용하여 음악요약을 하였다. 음악요약을 위한 방법은 아래와 같다.

① 주어진 음악의 모든 프레임들($f_1, f_2, \ldots, f_N$)중에서 가장 큰 SIC(Same Index Count) 값을 가지는 프레임 f_i를 구한다. 여기서 N은 프레임의 수이다.

$$② \quad SIC_i = \sum_{M=M_1,M_2,...,M_l} w_{M_k}\left(\sum_{s=1}^{S} I_{M_k}(f_i+s, f_j+s)\right) \quad (5)$$

$$I_{M_k}(f_i,f_j) = \begin{cases} 1, & \text{if } C_{M_k}(f_i) = C_{M_k}(f_j) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$
$$i,j \in [1,N],\ i+S<j$$

식 (5)와 식 (6)에서 M은 양자화 레벨을 나타내고, $C_{M_k}(f_i)$는 M_k레벨 벡터양자화에서 프레임 f_i의 코드워드를 나타낸다. w_{M_k}는 양자화 레벨에 따른 가중치이고, S는 음악요약 길이에 해당하는 프레임의 수를 나타낸다. $[f_i, f_i+S]$를 음악요약으로 제공한다. 여기서 S는 음악요약 길이에 해당하는 프레임의 수이다.

다중 양자화 레벨에 따른 가중치 w_{M_k}는 양자화 에러를 구할 때 가장 일반적으로 쓰이는 MSE(Mean Square Error) 값들의 최대값으로 정규화된 역수값를 사용하였다.

3.2. 최적의 길이를 가지는 음악요약 생성

3.1절에서는 미리 제공할 음악요약의 길이를 정하고 반복되는 부분을 찾아 음악요약으로 제공한다. 하지만, 음악 컨텐츠에 따라 다른 길이를 가지는 반복되는 부분을 고정된 음악요약으로는 나타내기가 어렵다.

그림 3에서 보듯이 고정된 길이를 가지는 음악요약은 반복되는 부분인 후렴구(Chorus)부분을 모두 포함하지 못하

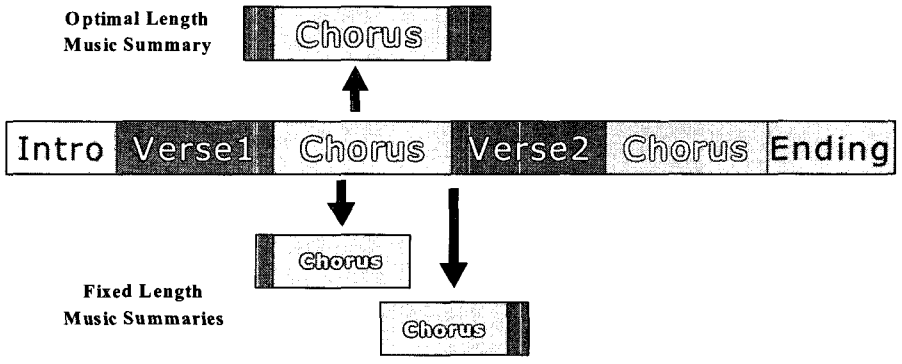


그림 3. 고정된 길이를 가지는 음악요약과 최적의 길이를 가지는 음악요약의예
Fig. 3. An Example of fixed length music summaries and optimal length music summary

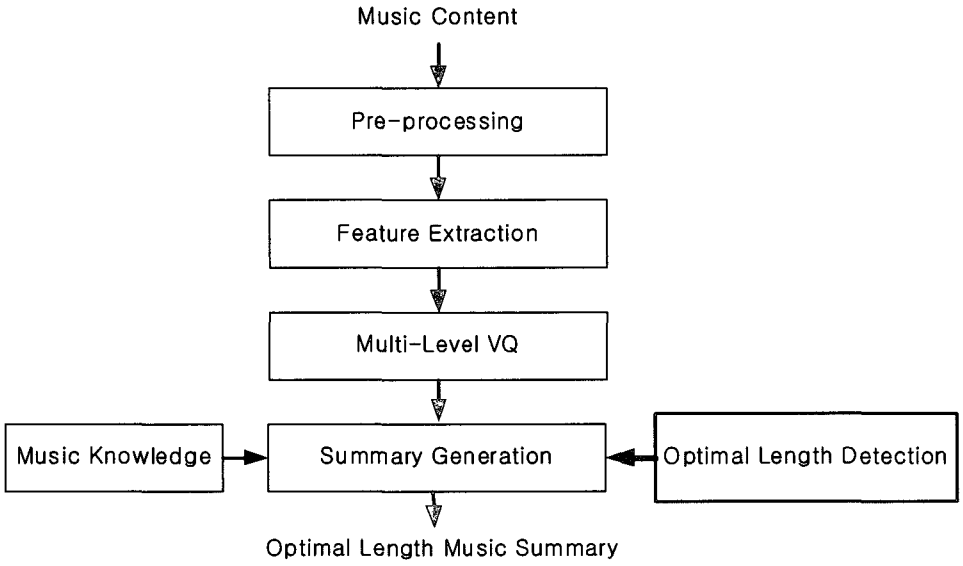


그림 4. 최적 길이를 가지는 음악요약 시스템 블록도
Fig. 4. Block diagram of optimal length music summarization

는 경우가 있지만, 최적의 길이를 가지는 음악요약(Optimal length music summary)의 경우는 비록 음악요약의 길이가 길어지는 단점은 있지만 사용자에게 반복이 일어나는 후렴구의 전체를 제공할 수 있다는 장점이 있다. 최적의 길이를 가지는 음악요약을 제공하는 시스템 블럭도는 아래 그림 4와 같다.

최적의 길이를 가지는 음악요약을 구하는 방법은 고정된 길이를 가지는 음악요약을 제공하는 방법에 최적의 길이를 찾는 방법만 추가하였다. 즉, 3.1장의 식 (5)대신 아래의 식 (7)을 사용하는 것이다. 그리고 제공할 음악요약의 길이를 S_1에서 S_2까지 변화시키면서 가장 큰 $SICM_i^*$을 찾아서 $[f_i, f_i + S^*]$를 음악요약으로 제공한다. 여기서 S^*는 SIC_i^*을 가질 때의 음악요약 길이에 해당하는 S값이다.

$$SIC_i^* = \underset{S_1 \le S \le S_2}{\operatorname{argmax}} SIC_i$$

$$SIC_i = \sum_{M=M_1,M_2,\ldots,M_l} w_{M_k} \left(\sum_{s=1}^{S} \cdot I_{M_k}(f_i + s, f_j + s) \right) \quad (7)$$

Ⅲ. 집단화(Clustering)방법을 이용한 음악요약 방법

집단화 방법을 이용한 음악요약 방법은 2장에서 설명한

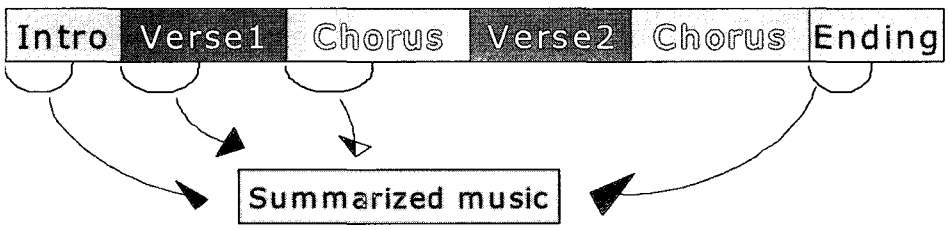


그림 5. 집단화(Clustering)기반 음악요약
Fig.5. Clustering-based music summarization

반복되는 부분을 음악요약으로 제공하는 방법과 달리 음악 컨텐츠내에서 특징이 다른 부분의 일정구간들을 모두 제공하는 음악요약 방법이다.

1. 특징벡터 추출 및 유사도

특징이 상이한 부분을 모두 제공하는 음악요약 방법에서는 반복되는 부분을 음악요약으로 제공하는 방법과는 다른 특징벡터를 사용한다. 특징 추출 과정과 유사도를 구하는 과정을 설명하면 다음과 같다. 그림 6에서 보듯이 특징 벡터로 Mel 필터 뱅크 에너지 45개와 음색과 관련된 3가지의 특징을 추가하여 총 48차의 특징 벡터를 추출하였다[6]. 그리고 이렇게 구해진 각 프레임의 특징 벡터들 간의 유사도는 아래 식 (8)과 같이 구할 수 있다. 뒤에서 사용될 세그먼트들 간의 유사도도 아래 식과 같이 구할 수 있다. 여기서 v_i와 v_j는 각각 i번째 j 번째 프레임의 특징벡터를 의미하고 $S(i,j)$는 i와 j 번째 프레임 간의 유사도를 의미한다.

$$S(i,j) = \frac{v_i \bullet v_j}{\| v_i \| \cdot \| v_j \|} \qquad (8)$$

2. 집단화 방법을 이용한 음악요약 기법

위 그림 5에서 알 수 있듯이 집단화 기반 요약방식은 음악의 다양한 부분을 추출해 취합하여 음악 요약으로 생성하는 것을 목적으로 한다. 주로 음악의 분위기가 다르거나 음악의 구조가 상이한 부분들을 모아 요약으로 생성하는 것이다.

이 방식은 다음과 같은 순서로 한다. 먼저, 프레임들을 세그멘트하게 된다. 만약 음악 신호의 i번째 프레임과 (i+1)번째 프레임의 유사도가 임계값보다 크고 (i+1)번째 프레임과 (i+2)번째 프레임의 유사도의 임계값보다 작을 경우 또는 그 반대의 경우에 (i+1)번째 포인트를 세그먼트 포인트로 정한다. 최종적으로 N개의 세그먼트 포인트가 생성되면 결과적으로 음악 신호가 N+1개의 세그먼트로 나뉜다. 이 때 이 세그먼트 수가 미리 정의된 개수를 넘지 않으면 유사도의 임계값을 낮추어 다시 세그먼트를 하게 된다. 최종적으로 미리 정의된 세그먼트 개수를 초과할 때까지 유사도의 임계값을 계속 낮추어가며 세그멘트를 한다. 이렇게 함으로써 전체적으로 음악의 분위기나 특성의 변화에 관계없이 어떤 음악에 대해서도 충분히 많은 세그먼트를 얻을 수 있도록 한다. 이렇게 얻은 세그먼트를 각 세그먼트의 평균을 구하고 이를 바탕으로 세그먼트들 간의 유사도를 구한다. 이전 과정에서 세그먼트를 하는 것과 비슷하게 처리하는데, 유사한 세그먼트들을 하나의 세그먼트로 통합하는 과정에서 유사도의 임계값을 조절하여 세그먼트의 수를 줄인다. 그리고 최종적으로 얻어진 각 세그먼트의 평균값들을 구해 k-means 집단화 알고리즘의 초기 코드워드로 정한 후 k-means알고리즘을 수행하여 최종적으로 코드워드들을 구한다. 이 코드워드를 바탕으로 음악 신호의 각 프레임의 특징 벡터들과의 거리(Distance)를 구해 가장 적은거

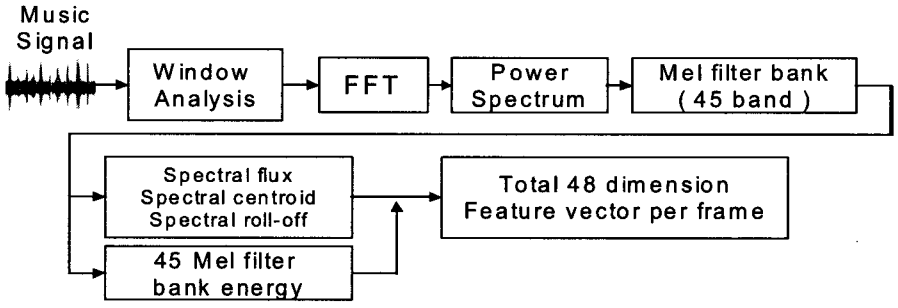


그림 6. 특징벡터 추출 과정
Fig. 6. Feature extraction process

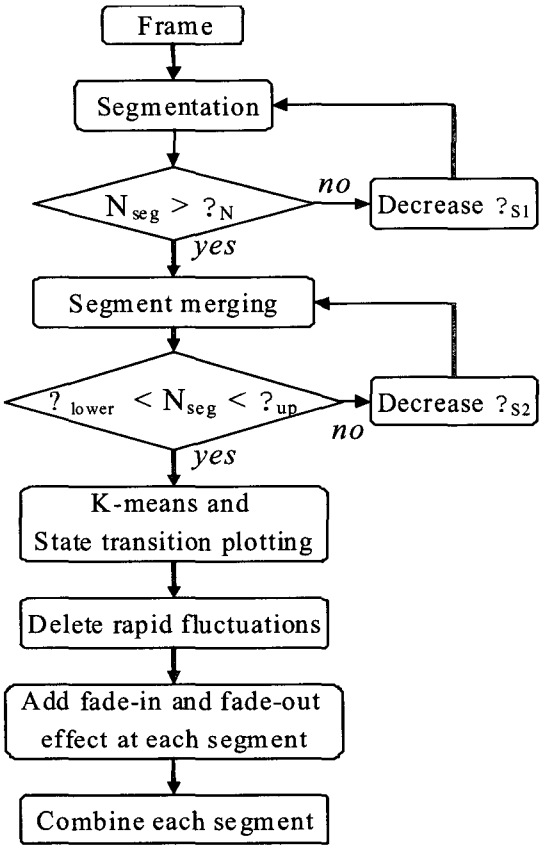


그림 7. 집단화 방법을 이용한 음악요약기법 순서도
Fig. 7. Flowchart of clustering-based music summarization

리를 가지는 코드워드의 인덱스(state)를 해당 프레임의 y축에 도시한다. 프레임 인덱스가 증가함에 따라 똑같은 방식으로 각 프레임의 해당state를 y축에 도시한 후 만약 i+1번째 프레임의 state와 i-1번째 프레임의 state가 동일할 경우 i번째 state를 i-1번째 프레임의 state로 정하여 state 천이의 고주파 부분을 제거한다. 이렇게 함으로써 음악이 동일한 구조에 속하거나 인지적으로 동일한 분위기에 속하는 부분임에도 불구하고 음악의 짧은 전자효과음 등으로 인해 어떤 프레임의 state가 근접 프레임의 state와 상이하게 되는 효과를 보정할 수 있게 된다. 그리고 그 state의 천이도를 바탕으로 각 state에서 일부분씩 추출하여 각 세그먼트의 시작과 끝에 fade in, fade out 효과를 준 후에 이를 취합하여 최종적으로 음악 요약을 생성하게 된다. 전체적인 순서도는 그림 7에 제시되어 있고 각 프레임의 state의 변화 과정을 도시한 그림이 그림 8에 제시되어 있다. 그림 5에서 N_{seg}은 생성된 세그먼트의 수, θ_N은 세그먼트 과정에서 미리 정의된 세그먼트의 수, θ_{s1}은 세그먼트 과정에서 유사도의 임계값, θ_{lower}와 θ_{up}은 세그먼트 통합 과정에서의 미리 정의된 세그먼트 수의 범위, θ_{s2}는 세그먼트 통합 과정에서의 유사도 임계값이다.

Ⅳ. 실험결과

본 논문에서 제안한 두 가지 방식의 음악요약 알고리즘의 성능을 평가하기 위해 서로 다른 평가 방법을 사용

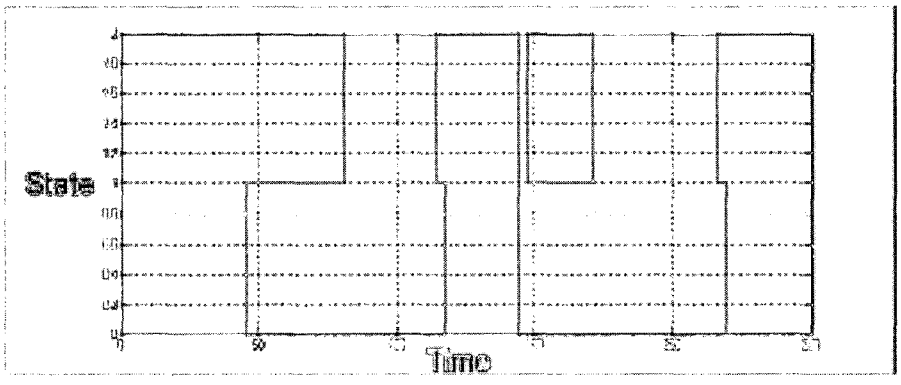


그림 8. 상태 천이도(서태지의 "난 알아요")
Fig. 8. State transition ("Blind love" by Seo Tai Ji)

하였다. 객관적인 방법과 주관적인 방법(MOS 테스트)를 사용하였는데, 음악내에서 반복되는 부분을 찾는 것은 악보를 참고하거나 음악을 여러 번 들으면 쉽게 찾아낼 수 있다. 하지만 음악내에서 특징이 다른 부분을 찾거나 음악의 구조 (전주, 코러스, 간주, 후주)를 이해하고 객관적인 평가를 하는 것은 어려운 작업이다. 그래서 반복되는 부분을 음악요약으로 제공하는 알고리즘은 객관적인 방법으로 집단화 방법을 이용한 음악요약 알고리즘은 주관적인 방법인 MOS 테스트를 실시하여 성능을 평가하였다.

1. 반복되는 부분을 음악요약으로 제공하는 알고리즘 성능 평가

1.1 고정된 길이의 음악요약 성능 평가

반복되는 부분을 음악요약으로 제공하는 알고리즘의 성능을 평가하기 위해 사람이 직접 만든 요약과 자동으로 생성된 요약과의 중첩비율(Overlapping Ratio)을 사용하였다.

$$\text{Overlapping Ratio} = \frac{P(S_{hand} \cap S_{VQ})}{P_{VQ}} \qquad (9)$$

여기서 P_{VQ}는 벡터양자화를 이용해서 자동으로 생성된 요약의 길이이고, $P(S_{hand} \cap S_{VQ})$는 사람이 직접 만든 요약과 자동으로 생성된 요약과 겹치는 길이를 나타낸다. 본 논문에서는 P_{VQ}를 20초로 하였다.

제안한 고정된 길이의 음악요약의 성능을 평가하기 위해 사용한 음악은 2000년 발표된 비틀즈(Beatles)의 "1" 앨범의 27곡을 사용하였다. 그리고 모노, 16bit 양자화, 그리고 22.05kHz.로 샘플링 하였다. 200ms의 프레임을 100ms씩 이동시키면서 특징벡터를 추출하였다.

제안된 음악요약 방법의 성능을 2000년과 2002년에 발표된 기존의 음악요약 방법들[5][6]과 비교하였다. 표 1은 한 개의 양자화 레벨만 사용한 경우의 실험 결과이다.

표 1. 단일 양자화 레벨에 따른 음악요약 성능
Table 1. Performance of music summarization using single-level vector quantization

양자화 레벨(M)	Overlapping Ratio(%)
8	62.2
16	63.9
32	64.9
64	64.3
128	64.9
256	64.1

표.2는 제안한 다중 레벨 벡터양자화를 이용한 음악요약 방법에서 다중 레벨 양자화 수에 따른 성능을 보여준다

표 2. 다중 레벨 양자화수에 따른 음악요약 성능
Table 2. Performance of music summarization according to the number of quantization level

다중 양자화 레벨 (M)	Overlapping Ratio(%)
128,64,32,16,8	66.9
128,64,32,16	67.1
128,64,32	66.6
128,64	65.8

실험결과 다중 양자화 레벨을 M=128,64,32,16.으로 하였을 경우가 가장 성능이 높았다. 아래 표 3은 기존의 방법과의 성능을 비교하였다.

표 3. 기존 음악요약 방법과의 성능비교
Table 3. Comparison between Performances of conventional methods and proposed method

	Overlapping Ratio(%)
Beth Logan's method[7]	42.2
Chansheng Xu's method[8]	47.3
Proposed method	67.1

기존에 제안된 음악요약 방법들에서 사용한 음악의 종류는 Beth Logan's method[7]에서는 비틀즈의 음악들을 이용하였고, Chanshen Xu's method[8]의 경우는 4 가지 장르 (Pop, Classic, Rock, Jazz) 장르의 음악들을 사용하였다.

이들 방법들은 주관적인 방법인 MOS 테스트를 이용하여 성능을 평가하였지만, 본 연구에서는 객관적인 방법으로 성능을 평가하였고 기존의 음악요약방법들이나 제안된 방법의 목적은 음악내에서 반복되는 부분을 찾는 것이고, 그 결과를 객관적으로 평가한 것이므로 음악의 종류와는 무관하다고 볼 수 있고, 실험결과 제안한 방법으로 음악요약을 생성한 결과가 기존의 방법들 보다 성능이 우수함을 알 수 있었다.

1.2. 최적의 길이를 가지는 음악요약 성능 평가

최적의 길이를 가지는 음악요약 알고리즘의 성능을 평가하기 위해서 고정된 길이의 음악요약 알고리즘의 성능을 평가하는 방법과는 다른 방법으로 성능을 평가하였다. 성능평가를 위해 사용한 음악 데이터는 고정된 길이를 가지는 음악요약 성능을 평가할 때 사용한 데이터와 동일한 데이터를 사용하였다. 성능평가를 위해 식 (10)의 최적 중첩비율(Optimal Overlapping Ratio)를 사용하였다.

$$\text{Optimal Overapping Ratio} = \frac{L(S^{*}_{hand} \cap S_{VQ})}{L(S^{*}_{hand})} \qquad (10)$$

$$S^{*}_{hand} = \arg\max_{k=1\ldots K} L(S_{hand,k} \cap S_{VQ})$$

여기서 K는 음악 컨텐츠내에서 반복되는 구간의 수이다. 최적 중첩비율을 이용한 성능은 아래 표들과 같다.

표 4. 고정된 음악요약 성능 (다중 양자화 레벨 종류: 128,64)
Table 4. Performance of fixed length music summarization (the kind of multi-level quantization level: 128,64)

Summary Length	Optimal Overlapping Ratio(%)
20s	83.8
30s	91.0
40s	91.4
50s	92.9

표 5. 최적의 길이를 가지는 음악요약 성능
Table 5. Performance of optimal length music summarization

Range of Summary Length (변화간격)	Average Summary Length	Optimal Overlapping Ratio(%)
20s ~ 40s (0.5s)	36.1s	93.6
20s ~ 50s (1.0s)	40.4s	93.2

위의 결과들로 알 수 있듯이 최적의 길이를 제공하는 음악요약 알고리즘이 고정된 길이의 음악요약을 제공하는 알고리즘에 비해 곡내에서 반복되는 부분을 최적의 길이를 사용하면서 더 많이 포함한다는 것을 알 수 있다.

2. 집단화 방법을 이용한 음악요약 알고리즘 성능 평가

집단화 방법을 이용한 알고리즘의 성능을 평가하기 위해서 MOS test를 사용하였다. 2가지의 질문을 하였는데 이는 아래와 같다.

질문1) 각각의 요약된 음악에서 몇 개의 비슷한 무드를 가지는 세그먼트가 있는가?
질문2) 각각의 요약된 음악에서 같은 구조에 속하는 세그먼트가 몇 개나 되는가?

첫 번째 질문에 대한 답은 Number of Similar Part (NSP) 라고 정의할 수 있고 두 번째 질문에 대한 답은 Number of Same Structure (NSS)라고 정의할 수 있다. 여기서 NSS, 즉 같은 구조는 대중 음악의 전형적인 구성 형식을 말하는 것으로써 전주, 코러스, 간주, 후주 등으로 나뉠 수 있다. 평가자가 자동 생성된 음악 요약의 세그먼트들 중에 몇 개의 세그먼트가 분위기가 비슷한지(NSP) 또는 몇 개의 세그먼트가 같은 구조에 속하는지(NSS)를 평가하는 것이다. 예를 들면, 평가자가 음악 요약을 듣고 첫 번째 음악의 요약의 NSP를 0, 두 번째 음악의 NSP를 2로 체크했다고 하고, 첫 번째 음악의 NSS를 2, 두 번째 음악의 NSS를 3으로 체크했다고 하면 NSP Ratio (NSPR) 와 NSS Ratio (NSSR)는 아래 <표6>과 같이 구할 수 있게 된다.

표 6. 평가방법예제
Table 6. An example of evaluation

Song	Segments	NSP	NSS	NSPR	NSSR
1	5	0	2	0/5	2/5

여기서, NSPR의 평균 0.2이고 NSSR의 평균은 0.5가 된다. 만일, 이 두 비율이 아주 작다면 자동적으로 생성된 여러 세그먼트가 서로 비슷하지 않고 각각의 세그먼트가 음악의 다른 구조에 속할 확률이 높다는 것을 의미한다. 이것이 바로 우리가 원하는 방향이고 이 기법이 추구하는 방향이다. 평가를 위해 가요 50곡, 락 30곡, 프로그레시브 락과 하드락 중에서 비교적 음악 길이가 긴 곡들로 이루어진 임의로 정의 된 장르인 Long-length Rock 20곡으로 분류하여 실험하였다. 평가자는 일반 음악이론과 평가 대상 음악에 대해 충분한 지식과 경험을 가진 전문가 1명이 하였고 자동 요약된 곡을 평가자가 원하는 만큼 여러 번 들어가며 수행하였다. 실험 결과는 표 아래와 같다.

표 7. 집단화 방법을 이용한 음악요약 성능
Table 7. Evaluation of the clustering-based music summarization method

Genre	NSPR	NSSR
Gayo	0.11	0.19
Long-length Rock	0.04	0.14
Rock	0.04	0.12

표 7의 결과로부터 알 수 있듯이 NSPR이 NSSR보다 작음을 알 수 있다. 즉, 이 기법은 다른 음악 구조에 속하는 부분을 추출하는 것보다 분위기나 신호 특성이 다른 부분을 추출하는 목적에 더 부합한다고 볼 수 있다. 이는 spectral energy와 같은 비교적 로우 레벨의 특징 벡터를 사용하였고 인간이 음악 구조를 인지하는 과정을 특별히 모사하여 알고리즘에 넣지 않았기 때문이다. 프레임 사이즈를 수 초 대로 하여 처리하여도 성능 차이가 나지 않아 처리 시간이 곡 당 3초 이내로 적어 Real-time 어플리케이션에 적용 가능하다는 점도 장점으로 생각할 수 있다. 하지만 이미 말했듯이 음악 구조에 더 가중치를 두어 세그먼트 추출을 하기 위해 하이 레벨의 특징을 추출하고 인간의 음악 구조 인지 과정을 별다른 성능 저하 없이 알고리즘에 추가한다면 더 좋은 성능을 보일 것으로 기대한다.

V. 결 론

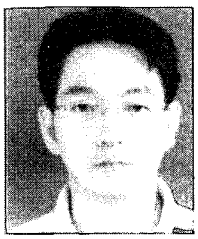
본 논문에서는 두 가지 방식의 음악요약을 제공하는 알고리즘들을 제안하였다. 주어진 음악 컨텐츠내에서 반복되는 부분을 음악요약으로 제공하는 방법에서는 다중 레벨 양자화를 이용해서 주어진 음악의 구조를 표현하였고, 같은 코드워드를 가지는 프레임 수의 가중합을 이용해서 반복되는 구간을 찾아서 제공하였다. 성능을 평가하기 위해 중첩비율과 최적중첩비율을 이용하여 성능을 평가하였다. 집단화 방법을 이용한 음악요약을 제공하는 방법에서는 두 가지 질문을 통해 NSPR과 NSSR을 이용한 MOS 테스트를 통해서 분위기나 신호 특성이 다른 부분을 추출하는데 좋은 성능을 보여줌을 알 수 있었다. 그리고 비교적 단순한 특징 벡터와 일반적인 기법을 사용했지만 임계값에 피드백을 주어 음악 신호의 각 프레임 특징의 유사도 변화 정도와 관계없이 다양한 곡에서 적용 가능하다는 장점이 있다.

참 고 문 헌

[1] L. Mani and M. T. Mabury, Advances in Automatic Text Summarization, Cambridge, Massashusetts: MIT Press, 1999.
[2] F. Ren and Y. Sadanaga, An Automatic Extraction of Important Sentences Using Statistical Information and Structure Feature,NL 98-125, pp. 71-78, 1998.
[3] Y. Gong and X. Liu, "Summarizing Video by Minimizing Visual Content Redundancies," IEEE International Conference on Multimedia and Expo, pp. 788-791, Tokyo, Japan, 2001.
[4] I. Yahiaoui, B. Merialdo and B. Huet, "Generating Summaries of Muli-Episode Video," IEEE International Conference on Multimedia and Expo, pp. 792-795, Tokyo, Japan, 2001.
[5] Matthew Cooper and Jonathan Foote, "Automatic Music Summarization via Similarity Analysis," Proc. IRCAM, pp. 81-85, 2002.
[6] Jonathan Foote, "Visualizing Music and Audio using Self-Similarity," Proc. ACM Multimedia Conference, pp. 77~80,

Orlando, Florida, November 1999.
[7] Beth Logan and Stephen Chu, "Music Summarization Using Key Phrases," IEEE International Conference on Audio, Speech and Signal Processing, pp. 749-752, 2000.
[8] Chansheng Xu, Yongwei Zhu, and Qi Tian, "Automatic Music Summarization based on Temporal, Spectral and Cepstral Feature," IEEE International Conference on Multimedia and Expo, pp. 117-120, 2002.

저 자 소 개



김 성 탁

· 2000년 2월 : 울산대학교 전자공학과 (학사)
· 2003년 8월 : 한국정보통신대학교 공학부 (석사)
· 2003년 9월~현재 : 한국정보통신대학교 공학부 (박사과정)
· 주관심분야 : 음성인식, 음향정보 인덱싱

김 상 호

· 2002년 2월 : 세종대학교 전자공학과 (학사)
· 20002년 12월~2004년 6월 :이트로닉스 연구원
· 2005년 3월~현재 : 한국정보통신대학교 공학부 (석사과정)
· 주관심분야 : 음향신호 처리, 음향정보 인덱싱

김 회 린

· 1984년 2월 : 한양대학교 전자공학과 (학사)
· 1987년 2월 : 한국과학기술연구원 전자공학과 (석사)
· 1992년 2월 : 한국과학기술연구원 전자공학과 (박사)
· 1987년 10월~1999년 12월 : ETRI 선임연구원
· 1994년 6월~1995년 5월 : 일본 ATR-ITL 방문연구원
· 2001년 1월~현재 : 한국정보통신대학교 공학부 부교수
· 주관심분야 : 음성인식, 화자인식, 음향코딩, 음향정보 인덱싱

최 지 훈

· 1999년 2월 : 경희대학교 전자공학과 (학사)
· 2001년 2월 : 경희대학교 대학원 전자공학과 (석사)
· 2001년 1월~현재 : 한국전자통신연구원 연구원
· 주관심분야 : 데이터방송, 멀티미디어통신, TV-Anytime

저 자 소 개

이 한 규

- 1994년 2월 : 경북대학교 전자공학과 (학사)
- 1996년 2월 : 경북대학교 전자공학과 (석사)
- 2003년 2월~현재 : 한국정보통신대학교 공학부 (박사과정)
- 1996년 2월~현재 : 한국전자통신연구원 방송미디어연구그룹 맞춤형방송연구팀 팀장
- 주관심분야 : 멀티미디어시스템, 멀티미디어 통신, TV-Anytime

홍 진 우

- 1982년 2월 : 광운대학교 응용전자공학과 졸업 (학사)
- 1984년 2월 : 광운대학교 대학원 전자공학과 졸업 (석사)
- 1993년 8월 : 광운대학교 대학원 전자계산기광학과 졸업 (박사)
- 1998년~1999년 : 독일 프라운호퍼연구소 (교환연구원)
- 1984년 3월~현재 : 한국전자통신연구원 방송미디어연구그룹장, 책임연구원
- 2000년 1월~현재 : 한국방송공학회 학술위원 및 편집위원
- 1993년 1월~현재 : 정보통신표준화연구단 방송기술위원회 위원
- 주관심분야 : 오디오 신호처리 및 부호화, 디지털 컨텐츠 보호 및 관리, 디지털 오디오 방송