

# 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 식별 알고리즘

류상현 \*김형국

광운대학교

imsp\_ljs@nate.com \*hkim@kw.ac.kr

## Background Music Identification in TV Broadcasting Program Algorithm using Audio Peak Detection

Jung-Sung Lee Hyoung-Gook Kim\*

Kwangwoon Univ.

### 요약

본 논문에서는 오디오 피크 검출을 적용한 TV 방송 프로그램내 배경음악 식별 알고리즘을 제안한다. 제안한 알고리즘은 음악 핑거프린트 추출 및 전송부, 음악구간 검출부, 음악 핑거프린트 고속 매칭 및 정보전송부 세 부분으로 구성되어 있다. 음악 핑거프린트 추출 및 전송부에서는 음악 원음 오디오 데이터를 푸리에 변환하여 스펙트럼 계수를 추출한다. 추출된 스펙트럼의 성분 중에서 일정한 문턱값 이상의 에너지를 가지는 값을 피크로 검출하고 검출된 피크를 이용하여 핑거프린트를 생성하고 데이터 베이스화한다. 음악구간 검출부에서는 입력된 방송 프로그램 오디오 데이터에 GMM(Gaussian Mixture Model)을 적용하여 음악과 음악 외 오디오 데이터로 분류한다. 음악 핑거프린트 고속 매칭 및 정보전송부에서는 음악구간이라고 인식된 쿼리 오디오 데이터를 음악 핑거프린트 추출 및 전송부와 동일한 과정을 통해 핑거프린트를 생성하고 데이터 베이스화된 음악 원음의 핑거프린트들과 비교하여 가장 유사한 음원의 정보를 TV의 화면에 자막으로 보여준다.

### 1. 서론

최근 방송의 다채널화, 초고속 인터넷의 발전, 저장장치의 대용량화는 사용자의 참여, 공유, 개방의 시대적 조류와 맞물려 디지털 정보를 기하급수적으로 증가시켰다. 이런 시대적 조류 속에서 많은 정보를 얻는 웹 2.0 생태계에서 사용자가 원하는 정보를 빠르고 정확하게 제공하는 웹 3.0생태계로 진화하고 있다.

최근 들어 단순한 텍스트 검색이 아닌 영상, 음성과 같은 다양한 멀티미디어 콘텐츠를 사용자의 요구에 맞춰 정확하게 검출하고 이를 빠르게 전달하는 기술에 대한 연구가 활발히 진행되고 있다.

본 논문은 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 식별 알고리즘을 제안한다. 제안된 알고리즘은 푸리에 변환된 음악 원음 오디오 데이터에서 스펙트럼 계수를 추출하고 추출된 스펙트럼의 성분 중에서 일정한 문턱값 이상의 에너지를 가지는 값을 피크로 검출한다. 검출된 피크를 이용하여 핑거프린트를 생성하고 데이터 베이스화한다. GMM(Gaussian Mixture Model)을 적용하여 입력된 방송 프로그램 오디오 데이터에서 음악구간을 분류한다. 음악구간으로 인식된 쿼리 오디오 데이터에서 핑거프린트를 생성하고 데이터 베이스화된 음악 원음의 핑거프린트들과 비교하여 가장 유사한 음원의 정보를 TV의 화면에 자막으로 보여준다.

본 논문의 구성은 다음과 같다. 오디오 피크 검출을 적용한 TV 방

송 프로그램내 배경음악 식별 알고리즘을 2장에서 설명하고, 3장에서 제안된 알고리즘의 검출 정확도를 측정하는 실험 및 결과를 설명한다. 마지막 4장에서 결론을 서술하며 마무리한다.

### 2. 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 검출 알고리즘

본 논문에서 제안하는 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 검출 알고리즘 구조는 그림 1.과 같다.

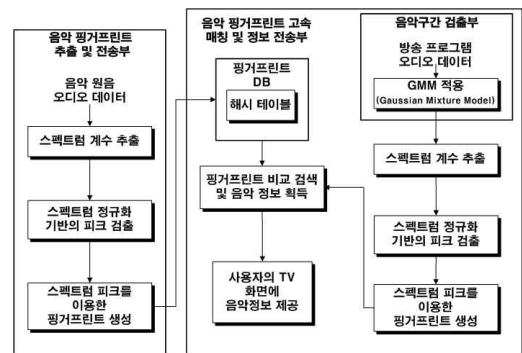


그림 1. 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 검출 알고리즘

음악 핑거프린트 추출 및 전송부에서는 음악 원음 오디오 데이터를 푸리에 변환하여 스펙트럼 계수를 추출한다. 추출된 스펙트럼 계수들 중에 증폭된 성분들과 종속된 성분들을 제거하고 입력된 오디오 신호의 고유한 특성을 나타내도록 로그도메인으로 변환한다. 로그도메인으로 변환된 스펙트럼에 zero-mean normalization을 적용하여 평균화 시켜줌으로써 음악 원음 오디오 데이터의 스펙트럼을 획득한다. 획득한 스펙트럼의 성분 중에서 일정 문턱값 이상의 에너지를 가지는 값을 피크로 검출하고 검출된 피크의 위치정보를 이용하여 핑거프린트를 생성한다. 생성된 핑거프린트를 음악 핑거프린트 고속 매칭 및 정보 전송부로 전달하여 데이터 베이스화 한다.

음악구간 검출부에서는 입력된 방송 프로그램 오디오 데이터에 GMM(Gaussian Mixture Model)을 적용하여 음악 오디오 데이터와 음악 외 오디오 데이터로 분류하고 음악 오디오 데이터가 인식된 구간을 음악구간으로 정의한다.

음악 핑거프린트 고속 매칭 및 정보 전송부에서는 음악 구간이라 인식된 쿼리 오디오 데이터를 푸리에 변환하여 스펙트럼 계수를 추출한다. 추출된 스펙트럼 계수들 중에 증폭된 성분들과 종속된 성분들을 제거하고 입력된 오디오 신호의 고유한 특성을 나타내도록 로그도메인으로 변환한다. 로그도메인으로 변환된 스펙트럼에 zero-mean normalization을 적용하여 평균화 시켜줌으로써 쿼리 오디오 데이터의 스펙트럼을 획득한다. 획득한 스펙트럼의 성분 중에서 일정 문턱값 이상의 에너지를 가지는 값을 피크로 검출하고 검출된 피크의 위치정보를 이용하여 쿼리의 핑거프린트를 생성한다. 이 후 데이터 베이스화된 음악 음원의 핑거프린트들과 생성된 쿼리의 핑거프린트를 비교하여 가장 유사한 음원의 정보를 사용자 TV의 화면에 자막으로 보여준다.

### 3. 실험결과

본 논문에서 사용한 GMM,과 GMM-UBM(Gaussian Mixture Mode - Universal Background Model), HMM(Hidden Markov Model), SVM(Support Vector Machine)에 방송 프로그램 오디오 데이터를 입력하여 음악구간 검출 정확도를 비교하는 실험을 진행하였다. 실험결과는 다음 표 1과 같으며, GMM이 97%로 가장 높은 검출 정확도를 보였다.

	GMM	GMM-UBM	HMM	SVM
정확도(%)	97	92.85	90.38	90.45

표 1. 음악구간 검출 정확도

본 논문에서 설명한 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 검출 알고리즘의 검출 정확도 성능 측정은 TV 방송 프로그램에서 배경음악으로 사용한 음악음원 200개를 사용하여 핑거프린트 DB를 구성하였고 방송 프로그램 오디오 데이터에 GMM을 적용하여 음악구간으로 인식한 오디오 데이터 200개에 대해 2초, 3초, 5초, 7초의 길이로 쿼리를 구성하였다.

	2초	3초	5초	7초
정확도(%)	87	95	98	100

표 2. 배경음악 검출 알고리즘 검출 정확도

실험결과는 다음 표 2와 같으며, TV안에 장착된 오디오 피크검출 기술을 적용하여 잡음의 영향을 받지 않는 환경이기 때문에 7초 쿼리에서 100%, 5초 쿼리에서 98%, 3초 쿼리에서 95%, 2초 쿼리에서 87%의 검출 정확도를 기록하여 짧은 쿼리에서도 비교적 높은 성능을 나타냈다.

### 4. 결론

본 논문에서는 오디오 피크 검출을 적용한 TV 방송 프로그램 내 배경음악 검출 알고리즘을 제안하였다. 제안된 알고리즘은 음악 음원 오디오 데이터로 DB를 구성하였고 방송 프로그램 오디오 데이터에서 음악구간을 검출하기 위해 GMM을 사용하여 쿼리를 구성하였다. TV 안에 장착된 오디오 피크 검출 기술을 적용하여 잡음에 영향을 받지 않아 비교적 짧은 쿼리에서도 높은 성능을 보였다. 향후 사용자 원하는 정보를 빠르고 정확하게 제공하는 웹 3.0생태계의 스마트 TV에 적용하여 사용자에게 TV방송 프로그램 내 배경음악의 정보를 빠르고 정확하게 전달할 수 있을 것이라 판단된다.

### 감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2012-0001941).

### 참고 문헌

- [1] S. Baluja, "Audio Fingerprinting: Combining Computer Vision & Data Stream Processing," Acoustics, Speech and Signal Processing, 2007.
- [2] J. Haitsma and A. Kalker, "A Highly Robust Audio Fingerprinting System," International Symposium on Music Information Retrieval, 2002.
- [3] L. Wang, "An Industrial-Strength Audio Search Algorithm," International Symposium on Music Information Retrieval, 2003.
- [4] J. Bernard and G. Abby, "Searching for Multimedia: Analysis of Audio, Video and Image Web Queries", World Wide Web, 2000.
- [5] Y. Li and Y. Hou, "Search Audio Data with the Wavelet Pyramidal Algorithm", Information Processing Letters, 2004.
- [6] D. Povey, L. Burget, M. Agarwal, P. Akyazi, K. Feng, A. Ghoshal, O. Glembek, N. K. Goel, M. Karafiat, A. Rastrow, R. C. Rose, P. Schwarz and S. Thomas "Subspace Gaussian mixture models for speech recognition," in Proc. of IEEE ICASSP, pp. 4330-4333, Dallas, TX, USA, Mar. 2010.