

음악 장르 분류를 이용한 자동차 오디오 시스템에서의 이퀄라이저 자동 조절 방식

Automatic Equalizer Control Method Using Music Genre Classification in Automobile Audio System

김형국* 남상순**
(Hyoung-Gook Kim) (Sang-Soon Nam)

요 약

본 논문은 자동차 오디오 시스템에 내장된 라디오에서 실시간으로 재생되는 연속적인 오디오 신호로부터 음악 신호를 선별하고, 해당 음악에 대한 실시간 음악장르 분류를 통해 자동으로 이퀄라이저를 조절하는 방식을 제안한다. 제안된 방식에서는 음악분류 정확도를 높이고 실시간 신호처리를 실행하기 위해 연속적인 오디오 신호로부터 추출한 음색 특징 벡터와 리듬 특징 벡터를 GMM (Gaussian mixture model) 분류 방식에 적용하여 음악 분류를 수행한다. 제안된 방식은 카 오디오 시스템의 라디오로부터 출력된 오디오 신호로부터 분할된 다양한 오디오 구간을 5가지 음악장르로 분류하여 음악 장르 분류 성능을 측정하였다.

Abstract

This paper proposes an automatic equalizer control method in automobile audio system. The proposed method discriminates the music segment from the consecutive real-time audio stream of the radio and the equalizer is controlled automatically according to the classified genre of the music segment. For enhancing the accuracy of the music genre classification in real-time, timbre feature and rhythm feature extracted from the consecutive audio stream is applied to GMM(Gaussian mixture model) classifier. The proposed method evaluates the performance of the music genre classification, which classified various audio segments segmented from the audio signal of the radio broadcast in automobile audio system into one of five music genres.

Key words: Music genre classification, music/non-music classification, automatic equalizer control

* 주저자 : 광운대학교 전파공학과 부교수(교신저자)

** 공저자 : 광운대학교 전파공학과 석사과정

† 논문접수일 : 2009년 8월 3일

† 논문심사일 : 2009년 8월 25일

† 게재확정일 : 2009년 8월 26일

I. 서 론

오늘날 자동차 오디오 시스템은 내장된 라디오 방송과 MP3와 같은 저장된 음악 및 오디오 파일들을 운전자가 자동차 안에서 향상된 음질로 들을 수 있도록 하는 다양한 기능을 지원하고 있다. 그 중, 오디오 이퀄라이저는 음악 감상 시 재생되는 음악의 장르, 특성에 따라 특정 부분의 소리를 조절할 수 있는 장치이다. 이러한 이퀄라이저를 통해 반사음이 심한 자동차 안에서 위치에 따라 음의 크기가 다르게 들리는 음의 불균형 현상을 보완해줄 수 있을 뿐만 아니라, 음악을 듣는 운전자의 성향에 따라 음을 조절하여 들을 수 있다.

현재 많이 쓰이고 있는 자동차 오디오 시스템의 이퀄라이저들 대부분은 운전자가 직접 조절하는 방식으로 매번 재생되는 음악이 바뀔 때마다 이퀄라이저를 수동으로 조절해야하는 불편함이 있다. 또한 운전 중에 이퀄라이저 조절은 교통사고 발생의 위험이 존재하기 때문에 대부분의 운전자들이 사용을 기피하고 있는 상황이다.

본 논문은 자동차 오디오 시스템에서 이러한 수동 이퀄라이저 조절 방식의 문제점을 해결하기 위해 라디오에서 실시간으로 재생되는 연속적인 오디오 신호에서 음악 신호를 자동으로 구분한 후 해당 음악의 장르 분류를 통해 자동으로 이퀄라이저를 조절하는 방식을 제안한다.

자동 이퀄라이저 조절방식의 핵심요소인 음악 장르 분류에 대해서는 다양한 연구들이 수행되어 오고 있다. [1]은 음악 신호로부터 음색, 리듬, 그리고 피치 특징 벡터를 추출하고, Gaussian Classifier, GMM, K-Nearest Neighbor 분류 방식을 사용하여 장르 분류 결과를 비교하였다. [2]는 음악 템포 특징과 Mel-frequency Cepstral Coefficient를 결합한 특징 벡터를 활용하여 음악 장르 분류를 수행하였다. [3]은 전체 음악 데이터 집합에서 음색 특징 벡터를 추출한 뒤 음악 장르 간 유사도를 측정하여 장르를 분류하였다.

기존에 수행된 방식과 비교하여 본 논문에서는 실시간으로 라디오에서 출력되는 오디오 신호로부터 음악 신호를 구분하여 일정구간에 대한 음악장르 분

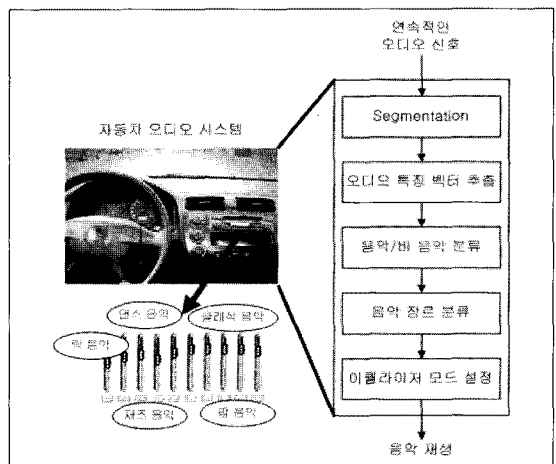
류 결과를 통해 실시간으로 이퀄라이저에 적용하는 방식을 사용한다. 그리고 음악 장르 분류의 정확도를 높이기 위해 음색 특징과 리듬 특징을 결합한 음악 특징 추출 방법을 적용한다.

본 논문의 구성은 다음과 같다. 2장에서 이퀄라이저 방식의 전체 흐름에 대해 살펴보고, 3장에서는 오디오 특징 추출 및 음악/비 음악 분류 방식에 대해 기술한다. 4장에서는 음악 장르 분류 방식, 이퀄라이저 조절 방식에 대해 설명하고, 5장에서는 실험 결과, 마지막으로 6장에서는 결론을 제시한다.

II. 자동 이퀄라이저 조절 방식

<그림 1>은 본 논문에서 제안한 자동차 오디오 시스템의 자동 이퀄라이저 방식 구성도를 나타낸다.

제안된 시스템의 입력은 라디오에 재생되는 연속적인 오디오 신호로써, 입력 신호에 대한 실시간 신호처리를 위해 초 단위로 분할된 오디오 구간에 대해 오디오 특징 벡터를 추출하고, 이퀄라이저 조절이 필요 없는 비음악 신호를 구분하기 위해 음악/비 음악 분류를 수행한다. 분류 결과가 음악 신호인 경우, 음악 장르 분류 과정을 통해 자동으로 이퀄라이저 모드가 설정되고, 설정된 이퀄라이저 모드는 현재 재생 중인 음악 신호에 적용된다.



<그림 1> 자동 이퀄라이저 시스템 구성도

<Fig. 1> Block diagram of automatic equalizer control method

III. 음악/비 음악 분류 과정

음악/비음악 분류 과정은 오디오 특징 추출 과정과, 추출된 오디오 특징 값을 사용하여 입력 오디오 신호에 대한 음악/비음악 분류로 구성되어 있다.

1. 오디오 특징 벡터 추출

실시간 입력에 대한 오디오 특징 벡터 추출을 위해 입력 오디오 신호를 중첩이 되지 않은 초 단위의 구간으로 분할하여 음색 특징 벡터와 리듬 특징 벡터를 추출한다.

음색 특징 추출 과정은 7개의 옥타브로 분할된 각 서브 밴드의 주파수 영역에서 식 (1)~(5)을 통해 식 spectral centroid C_t , spectral bandwidth B_t , spectral flux F_t , spectral flatness L_t 등의 특징[4]을 추출한다.

$$C_t = \frac{\sum_{i=N_0}^N (|S_t(i)|^2 \times i)}{\sum_{i=N_0}^N |S_t(i)|^2} \quad (1)$$

$$B_t = \sqrt{\frac{\sum_{i=N_0}^N (|S_t(i)|^2 \times i - C_t)^2}{\sum_{i=N_0}^N |S_t(i)|^2}} \quad (2)$$

$$\sum_{i=N_0}^{R_t} |S_t(i)|^2 = 0.95 \cdot \sum_{i=N_0}^N |S_t(i)|^2 \quad (3)$$

$$F_t = \sum_{i=N_0}^N (|S_t(i)|^2 - |S_{t-1}(i)|^2) \quad (4)$$

$$L_t = 10 \cdot \log \left(\frac{(N - N_0 + 1) \sqrt{\prod_{i=N_0}^N |S_t(i)|^2}}{\left(\sum_{i=N_0}^N |S_t(i)|^2 \right) / (N - N_0 + 1)} \right) \quad (5)$$

여기서, N 은 서브 밴드 최대 인덱스, t 는 프레임 인덱스, S_t 은 DFT (discrete fourier transform) 계수를

나타낸다.

$$P_t(k) = \max_{i=0 \dots N_k} (|S_t(k, i)|^2) \quad (6)$$

$$V_t(k) = \min_{i=0 \dots N_k} (|S_t(k, i)|^2) \quad (7)$$

$$M_t(k) = \frac{1}{N_k} \left(\sum_{i=1}^{N_k} |S_t(k, i)|^2 \right) \quad (8)$$

식 (6), (7), (8)은 7개의 옥타브 스케일의 서브 프레임의 peak, valley, mean을 나타내며, k 는 각 서브 밴드 인덱스, N_k 는 각 서브 밴드의 최대 인덱스를 나타낸다. 추출된 스펙트럼 특징 벡터를 가산하고 각 프레임 별 특징 벡터의 평균값과 분산값을 추출하여 총 52차원의 음색 특징 벡터가 추출된다.

리듬 특징 추출 과정은 낮은 주파수 영역에 주로 분포하는 음악 악기음의 리듬 특징 값을 추출하기 위해 서브 밴드의 대역 주파수만 통과시키는 저주파 필터링을 수행한다. 필터링 된 오디오 신호에 대해 고주파 신호를 제거하기 위해 다운 샘플링 과정을 거치고, 주파수 영역으로 전환한다. 주파수 영역에서 추출된 파워 스펙트럼은 각 프레임 별로 대역폭이 대수적으로 분포된 12개의 triangular 윈도우를 사용하여 프레임 당 12차원의 리듬 특징 벡터를 추출한다.

2. 음악/비 음악 분류

라디오에서 재생되는 오디오 신호는 음악과 비음악 신호로 구성되어 있다. 음악 신호의 경우 음악의 장르나 사용되는 음악에 따라 주파수 특성이 다르게 나타나기 때문에 이퀄라이저 조절을 통해 변화를 줄 수 있는 반면에, 뉴스, 안내 방송, 라디오 토크쇼와 같은 비음악 신호의 경우에 대해서는 이퀄라이저 조절이 불필요하다.

본 논문에서는 입력되는 연속적인 오디오 신호로부터 음악구간을 구별하기 위해 GMM 분류방식을 기반으로 음악/비음악 분류를 수행한다.

분류 과정은 분류 기준이 되는 각각의 음악/비 음악 모델 생성을 위한 EM기반의 학습 과정과, 생성된

모델과 입력 오디오 신호의 유사도를 측정하여 음악/비음악을 판단하는 과정으로 구성된다. 학습 과정은 앞서 설명한 52차원의 음색 특징과 12차원의 리듬 특징을 사용하여 수행된다.

$$p(C_i|x_i) = p(x_i|C_i) \cdot p(C_i) \quad (9)$$

식 (9)는 생성된 모델과 입력 오디오 신호의 유사도 비교를 나타내고, 여기서 C_i 는 생성된 음악/비음악 모델, x_i 는 입력 오디오 신호에서 추출된 특징 벡터를 나타낸다.

IV. 음악 장르 분류 과정

1. 음악 장르 분류

<그림 2>는 본 논문에서 사용한 음악 분류 과정을 나타낸다.

클래식, 팝, 재즈, 댄스, 락 음악의 5개 음악 장르에 대해 각 장르별로 선택된 곡들로부터 음악 특징을 추출하고, 추출된 음악특징을 GMM 학습방식에 적용하여 각 장르별 모델을 생성한다. 라디오에서 출력되는 오디오 신호로부터 음악신호로 구분된 음

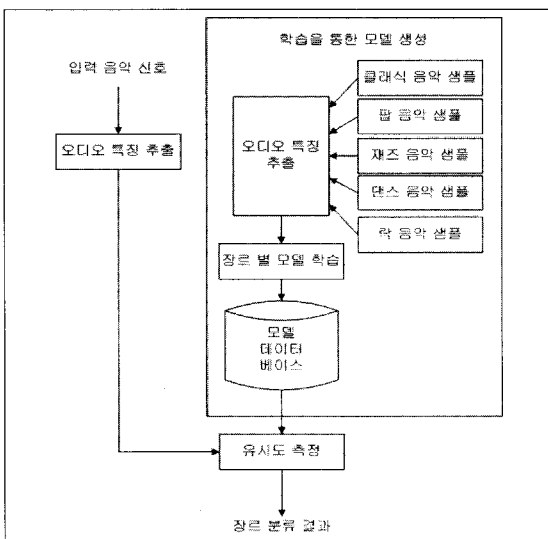
악구간 신호는 학습에 의해 생성된 각 장르별 모델들과의 유사도 비교를 통해 5가지 장르 중 하나로 판별된다. 본 논문에서 GMM 분류방식을 사용한 이유는 GMM 방식은 다른 통계적인 모델기반 분류방식과 비교하여 구조가 단순하고 분류정확도가 높기 때문이다.

일반적인 음악은 도입부, 연결부, 절정부, 그리고 종결부의 반복으로 구성되어 있는데 음악의 시작 부분인 도입부는 비교적 변화가 없고 단조롭기 때문에 장르 분류를 위한 정보를 충분히 포함하고 있지 않다. 그렇기 때문에 제안된 방식에서는 분류 결과의 신뢰성을 높이기 위해 시작 구간에서 일정 구간이 지난 후의 구간부터 음악 장르 분류를 수행하며 음악 구간의 분류 결과가 가장 많은 빈도수를 나타내는 음악 장르를 해당 음악의 장르로 결정한다. 이때, 병렬적으로 수행되는 음악/비음악 분류 결과가 음악에서 비 음악으로 상태 변화가 발생할 때까지 음악 장르 분류는 따로 수행되지 않는다. 음악에서 비음악으로 분류가 전환될 때에 선택된 이퀄라이저의 모드는 정지된다.

2. 이퀄라이저 조절 방식

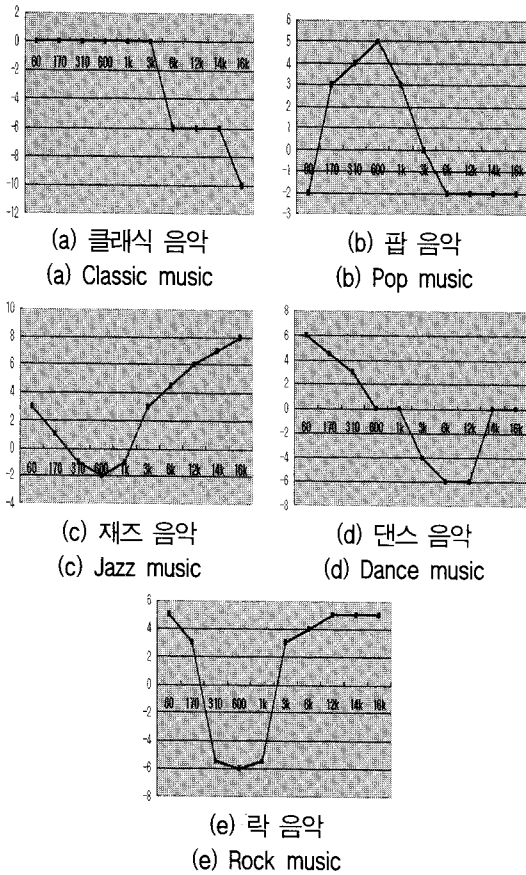
음악 장르 분류 결과에 따른 장르 별 주파수 특성은 <그림 3>과 같이 이퀄라이저 모드 설정부에 저장되어 있으며, 분류된 음악장르 결과에 따라 자동으로 이퀄라이저에 연결되어 각 음악에 맞게 음악을 재생시킨다.

<그림 3(a)>는 클래식 음악의 경우로 낮은 주파수 대역에서는 일정하지만, 높은 주파수 대역에서 감소하는 형태를 나타낸다. <그림 3(b)>는 팝 음악의 경우로 낮은 주파수 대역에서 주파수 특성이 증가하다가 감소하기 시작하여 6kHz 구간부터 일정한 상태를 나타낸다. <그림 3(c)>는 재즈 음악인 경우로 600Hz 까지 감소하다 그 이후로 증가하는 것을 알 수 있다. <그림 3(d)>의 댄스 음악은 주파수 특성이 감소하다가 12kHz부터 증가하여 일정해지는 형태를 나타낸다. <그림 3(e)>는 락 음악의 경우로 600Hz까지 감소하다 그 이상부터는 증가하는 형태를 나타낸다.



<그림 2> 음악 장르 분류 과정

<Fig. 2> Block diagram of music genre classification



<그림 3> 음악 장르에 따른 주피수 특성

<Fig. 3> Frequency responses according to music genre

V. 실험 결과

실험을 위해서 사용된 오디오 DB는 국내 FM 라디오 방송으로부터 10시간 분량의 오디오 신호를 22050Hz, 16bits, mono WAV 형식의 파일로 녹음하여 제안된 방식에 적용하였다. 그리고 이퀄라이저 조절 방식의 성능에 영향을 주는 음악/비음악 분류와 음악 장르 분류를 다양하게 분할된 오디오 구간과 그 구간에 대한 연속적인 장르분류 빈도수에 따라 음악분류 정확도를 비교하였다.

<표 1>은 오디오 신호 입력 시 분할하는 구간 간격을 1초, 2초, 3초, 4초의 다양한 구간으로 나누어 음악/비 음악 분류 성능을 측정한 실험결과를 나타내었다.

<표 1> 분할 구간 별 음악/비 음악 분류 결과
<Table 1> Accuracy of music/non-music classification based on the interval of segmentation

	1초	2초	3초	4초
Accuracy	54.45%	78.07%	86.57%	87.53%

<표 1>에 나타난 바와 같이 오디오 세크먼트 길이가 1초에서 3초까지 1초 단위로 증가시킨 경우에는 음악/비음악 분류결과가 향상되었으나, 3초에서 4초로 오디오 구간간격을 늘렸을 경우에는 음악/비음악 분류결과향상이 미소함을 알 수 있었다.

음악 장르 분류의 성능 측정을 위해 Support Vector Machines(SVM), Multi-Layer Percptron(MLP), GMM 등의 세 가지 방식을 200개의 곡들(각 장르별 40곡)에 대해 음악장르 분류 실험을 수행한 결과, MLP 방식을 사용한 음악분류 결과가 88.56%로 GMM방식의 87.75%, SVM 방식의 87.34%에 비해 미소하게 우수한 성능을 나타내었다. 그러나 GMM 방식이 연산량이 작기 때문에 제안된 방식에서는 GMM 방식을 이용하여 음악장르 분류를 수행하였다. 즉, 5가지 음악 장르에 대해 장르 별로 200개의 곡을 사용하여 각 장르별 모델을 생성하고, 표1에서 실험한 오디오 구간을 기반으로 분류된 음악구간에 대해 장르별 분류를 수행하였다. 이 경우, 음악이 시작된 후 5초 구간인 음악 도입부 구간을 제외한 음악구간에 대해 음악장르 분류를 수행하였고, 음악 장르 결정을 위한 판단 횟수를 비교 파라미터로 하여 음악장르 분류 성능을 <그림 2>와 같이 측정하였다.

실험 결과 도입부 이후 4초 구간의 5번에 대한 장

<표 2> 분할 구간 별 및 분류 횟수 별 음악 장르 분류 결과 비교

<Table 2> Comparing the accuracy of music genre classification based on the interval of segmentation and the number of classification

	1초	2초	3초	4초
1회	65.4%	73.4%	83.4%	87.5%
3회	72.5%	78.5%	85.2%	89.3%
5회	76.3%	85.4%	89.3%	89.7%

르 분류 결과를 적용하였을 때 이퀄라이저 모드 선택을 위한 정확도가 가장 높은 것을 알 수 있다. 하지만 4초 구간의 5번의 분류 수행은 전체 음악에서 많은 시간을 소요하게 되어 실시간으로 재생되는 음악에 이퀄라이저를 적용하기에는 이퀄라이저 지연이 많이 발생하게 된다. 따라서 본 논문에서는 실시간 음악에 적용하기 위한 최적의 이퀄라이저 모드 조건으로 2초 구간의 5번의 분류 인증결과를 사용한다.

VI. 결 론

본 논문에서는 자동차 오디오 시스템의 라디오에서 실시간으로 재생되는 오디오 신호에서 음악 신호를 선별하여, 음악 장르 분류를 통한 자동 이퀄라이저 조절 방식을 제안하였다. 제안된 방식은 실시간 오디오 신호로부터 음악/비음악 분류방식에 의해 구분된 음악신호에 대한 음악 장르 분류 결과의 분석을 통해 사용자에게 실시간으로 재생되는 음악에 대해 빠른 시간 안에 선택된 이퀄라이저 모드가 적용할 수 있게 되었다. 이러한 자동 이퀄라이저 조절 방식은 현재 상용화 되어 있는 카오디오 시스템에 적용되어 기존 시스템에 장착된 이퀄라이저 사용의 불편함을 해결할 수 있을 뿐만 아니라 운전자들에게

새로운 음악 감상의 방식을 제공할 수 있을 것이다.

참 고 문 헌

- [1] G. Tzanetakis, and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech and Processing*, vol. 10, no. 5, pp. 293-302, Sep. 2002.
- [2] A. Meng, P. Ahrendt, J. Larsen, and L. Hansen, "Temporal feature integration for music genre classification," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1654-1664, July 2007.
- [3] K. West and S. Cox, "Features and classifiers for the automatic classification of musical audio signals," *Proc. Int. Conf. Music Information Retrieval*, pp. 531-536, Oct. 2004.
- [4] H. G. Kim and K. W. Eom, "Automatic emotion classification of music signals using MDCT-Driven timbre and tempo feature," *J. Acoustic Society of Korea*, vol. 25, no. 2, pp. 74-78, June 2006.

저자소개



김 형 국 (Kim, Hyoung-Gook)

2007년 3월 ~ 현재 : 광운대학교 전파공학과 부교수



남 상 순 (Nam, Sang-Soon)

2009년 3월 ~ 현재 : 광운대학교 전파공학과 석사과정

2009년 : 광운대학교 전파공학과 공학사