

추출된 특성 파라미터의 Gaussian Mixture Model 을 이용한 음성 및 음악 판별

정기훈, 이봉진, 김경태, 강홍구
연세대학교
{kihooni, lbjcom, kktae, hgkang}@dsp.yonsei.ac.kr

Discrimination of Speech and Music signals using the Gaussian Mixture Model of Feature Parameters

Ki-Hoon Jung, Bong-Jin Lee, Kyung-Tae Kim, Hong-Goo Kang
Yonsei University, Seoul, Korea

요 약

본 논문은 오디오 신호에서 음성 또는 음악을 분류하는 방법을 제안한다. 음성과 음악신호를 구분할 수 있는 특성 파라미터들의 분포를 GMM(Gaussian Mixture Model)으로 모델링하고, 이를 바탕으로 임의의 오디오 신호에서 추출한 특성 파라미터의 확률값을 계산하여 비교하면 음성인지 음악인지 판별할 수 있다. 이 때 한 개의 특성 파라미터만 이용할 경우보다 상관관계가 적은 두 개 또는 그 이상의 특성 파라미터를 이용한 다차원 GMM(Multivariate GMM)을 적용할 경우에 보다 나은 성능을 얻을 수 있다. 4 차원 GMM 을 적용하여 실험한 결과, closed-set test 에서 최고 95.87%, open set test 에서 최고 95.23%의 전체 판별 정확도를 얻었다.

I. 서 론

오디오 신호에서 음성과 음악을 분리하는 기술은 많은 부분에서 응용이 가능하다. 그 예로, 텔레비전 방송의 오디오 신호 중 음성신호만 분리해 내어 음성 인식기법을 통해 청각 장애인을 위한 자막 서비스(STT: Speech to Text)를 제공할 수 있고, 음성과 음악을 분리하여 드라마 등의 방송 편집에도 응용이 가능하다. 또한 오래된 음반의 클리닝 작업 등에도 응용될 수 있다. 이러한 응용분야에 적용하기 위해 음성과 음악 고유의 특성을 나타내는 여러 가지 특성 파라미터들을 추출하여 음성 및 음악을 판별하는 알고리즘에 대한 연구가 이루어 지고 있다[1].

오디오 신호를 음악과 음성으로 분류하거나 판별할 수 있는 이유는 음성의 경우 주기적 특성을 갖는 유성음과 비주기적 특성을 갖는 무성음의 조합으로 구성되고 발성 중간에 묵음 구간이 존재하지만, 음악의 경우는 음성의 주기적 특성인 피치보다 넓은 주기의 특성(리듬)을 갖기 때문이다. 이러한 특성은 다양한 오디오 샘플의 spectrogram 을 통해 확인할 수 있다[2].

본 논문에서는 음성과 음악신호의 특성 파라미터를 추출하고 이 추출된 특성 파라미터들의 분포를 GMM 으로 모델링하였다. 그리고 임의의 오디오 신호의 특성 파라미터 확률값과 비교하여 음성 또는 음악으로 판별해 보았다. 또한 음성과 음악의 판별 정확도 향상을 위해 두 개 또는 그 이상의 특성 파라미터를 사용하였다. 다차원 GMM 을 적용했을 경우 1 차원 GMM 을 적용한 경우 보다 성능이 향상되었음을 확인하였다.

나타내었다. 우선 임의의 오디오 신호를 음성 또는 음악으로 판별하기 위한 기준으로 통계 특성 모델링에 많이 사용되는 GMM 을 생성한다. 정확한 통계 모델 생성을 위해서 충분한 길이의 다양한 화자로 구성된 음성과 다양한 장르의 음악 데이터가 필요하다. 이 데이터를 시간 또는 주파수 영역에서 분석하여 음성과 음악을 판단할 수 있는 특성 파라미터를 추출하고, 추출된 특성 파라미터 분포를 이용해 GMM 을 생성한다.

한편, 음성/음악 판별 과정에서는 판별하고자 하는 오디오 신호에서 특성 파라미터를 추출하여 이미 생성된 GMM 을 통해 확률값을 계산하고 비교하여 음성과 음악, 두 개의 카테고리 중 하나를 선택한다.

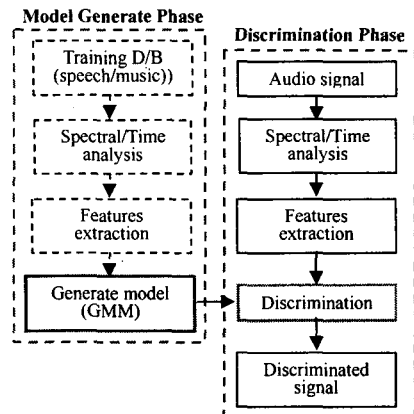


그림 1. 음성/음악 판별 과정

II. 음성/음악 판별과정 개요

음성과 음악 판별을 위한 전반적인 과정을 그림 1 에