GMM 모델링에 기초한 음색의 특징 연구

이 은, 공은배(충남대학교 컴퓨터공학과)

최근 인간과 기계의 man-machine interface로서 음성 기술이 광범위하게 응용되고 있다. 주식시세, 날씨 정보 등과 같은 정보의 전달분야에서 정보를 음성을 통해서 전달하는 것에서 확장하여 음성인식 컴퓨터, 음성인식을 통한 접근제어 시스템이나 음성인식 전화기 등에 활용이 확대되고 있다. 사람의 음성은 각각의 고유한 특징을 가지고 있어서 그 사람을 구별할 수 있게 해준다. 이것은 앞으로의 man-machine interface에 아주 편리하게 이용될 수 있으며, 음성합성 기술이나 어떤 화자의 목소리를 다른 화자의 목소리로 변환하는 음성변환 기술을 위한 기초 연구가 될 수 있으며 활용범위가 매우 크다고 할 수 있다.

요 약

기존의 음성기술 방법에서는 특정인이 발성한 다양한 음소를 포괄적으로 녹음, 수집하여 corpus를 만든 후, 이런 음소를 자연스럽게 연결하여 필요한 음성을 만들어내는 방식을 취하였다(corpus-based). 사람의 음성은 기분이나 컨디션에 따라 같은 목소리라도 약간의 다른 특징을 갖고 있다. 따라서 corpus-based 방식을 쓴다면, 녹음이 되어있는 음성만을 인식해 낼 수가 있으며, 음성 합성을 위해서도 녹음을 위한 막대한 비용과 시간과 노력이 소모된다. 그러나 각각 화자의 고유 특징, 즉 피치, 강세, 억양, 속도 등에 대한 데이터를 가지고 화자의 음색의 특징을 학습해 낼 수 있다면 비용과 효율성을 매우 높일 수 있게 된다. 학습된 데이터를 이용해서 목소리를 만들어 낼 수 있으며, 데이터베이스에 없는 음소들도 만들어 이용할 수 있다.

이러한 학습을 위해서 확률 모델인 Gaussian Mixture Modeling(GMM)을 이용할 것이다. 이러한 모델링을 위해 음성에 대한 신호처리가 이루어져야 할 것이고, 신호처리된 데이터를 이용하여 모델링을 할 것이다.

각각의 음소들이 어떻게 시그널로 표현될 수 있는지에 대해 살펴보고, 시그널 모델링을 통해서 음색의 특징을 추출해 내어 개인 고유의 음색의 특징을 알아낼 것이다.

제 2절에서는 음소의 모델링 방법인 GMM에 대해서 살펴볼 것이다.

Gaussian Mixture Model은 local model로써 각 components의 model 로 Gaussian Normal distribution을 사용한다. Normal distribution은 mean 와 covariance 로 완벽하게 specify된다. 어떤 복잡한 현상을 모델링할 때 하나의 모델로 전체 현상을 설명하려는 방법 (Global model) 과, 부분을 더 잘 설명할 수 있는 부분 모델들을 먼저 만들고 이 부분 모델들을 잘 결합하여 전체 현상을 설명하려는 방법(Local model) 이 있다. 음성 현상은 매우 복잡한 현상으로 global model로 설명하기에는 좀 어려운 점이 있다. Phonetic events들에 따른 local models을 결합하여 좀 더 좋은 모델을 만들 수 있다.

Gaussian Mixture Model은 local model로써 각 components의 model 로 Gaussian Normal distribution을 사용한다.

제 3절에서는 GMM을 이용하여 각각의 음소들을 학습시키고, 학습된 결과를 가지고 음색의 특징을 추출해낼 것이다.

제 4절에서는 이렇게 추출된 음색의 특징들을 GMM의 컴포넌트들을 변화시켜가면서 얻은 최적의 음색 정보를 얻어낸다.