MFCC를 작업하기 위해 mp3 파일을 균일한 간격으로 20ms 프레임으로 가져온다. 각각의 프레임마다 edge를 부드럽게 하기 위해 hamming window로 곱한다. 그리고 frequency의 주요성분을 얻기 위해 fourier transform을 시행한다. 그러고는 frequency를 mel scale로 mapping한다. 이러한 mapping group은 mel scale을 기반한 triangle window coefficients로 계산하여 log를 취한다. 그러면 우리는 DCT(discrete cosine transform)을 얻는데 이는 대략적으로 karhunen-loeve transform (frequency 구성요소들을 decorrelate하게 하는 변환)이다. 마침내 우리는 파동을 cepstral features의 행렬 꼴로 가지게 된다. 이 때 행렬의 row는 각 20ms frame의 15 cepstral frequencies의 벡터가 된다.