MFCC를 구하는 방법을 간단히 설명한다. 음성신호는 anti-aliasing filter를 거친 다음, A/D변환을 거쳐서 디지털 신호 x(n)로 변환된다. 디지털 음성신호는 고대역 통과 특성을 갖는 디지털 프리엠퍼시스 필터를 거친다. 이 필터를 사용하는 이유는 첫째로 인간의 외이/중이의 주파수 특성을 모델링하기 위하여 고대역 필터링을 한다. 이는 입술에서의 방사에 의하여 20 dB/decade로 감쇄되는 것을 보상하게 되어 음성으로부터 성도 특성만을 얻게 된다. 둘째 청각시스템이 1 kHz이상의 스펙트럼 영역에 대하여 민감하다는 사실을 어느 정도 보상하게 된다. PLP 특징추출에서는 인간 청각기관의 주파수 특성인 equal-loudness curve를 직접 모델링에 사용한다. 프리엠퍼시스 필터의 특성 H(z)는 다음과 같으며, a는 0.95-0.98 범위의 값을 사용한다.

http://speech.cbnu.ac.kr/srhome/technology/images/feature_extraction_html_eqn1.gif

프리엠퍼시스된 신호는 해밍 윈도우를 씌워서 블록 단위의 프레임으로 나누어진다. 이후부터의 처리는 모두 프레임 단위로 이루어진다. 프레임의 크기는 보통 20-30 ms이며 프레임 이동은 10 ms가 흔히 사용된다. 한 프레임의 음성신호는 FFT를 이용하여 주파수 영역으로 변환된다. 주파수 대역을 여러개의 필터뱅크로 나누고 각 뱅크에서의 에너지를 구한다. 밴드 에너지에 로그를 취한 후 discrete cosine transform (DCT)를 하면 최종적인 MFCC가 얻어진다. 필터뱅크의 모양 및 중심주파수의 설정 방법은 귀의 청각적 특성(달팽이관에서의 주파수 특성)을 고려하여 결정된다. 아래 그림에서는 삼각형 모양의 필터를 사용하였으며 중심주파수는 1 kHz 까지는 선형적으로 위치하고 그 이상에서는 mel 척도로 분포하는 20개의 뱅크로 이루어져 있다. MFCC 계수는 c1~c12까지의 12개를 사용하며 이와는 별도로 구한 프레임 로그 에너지가 추가적으로 사용되어 음식인식의 입력으로 사용되는 특징벡터는 13차 벡터가 된다.

