

## MFCC 정의

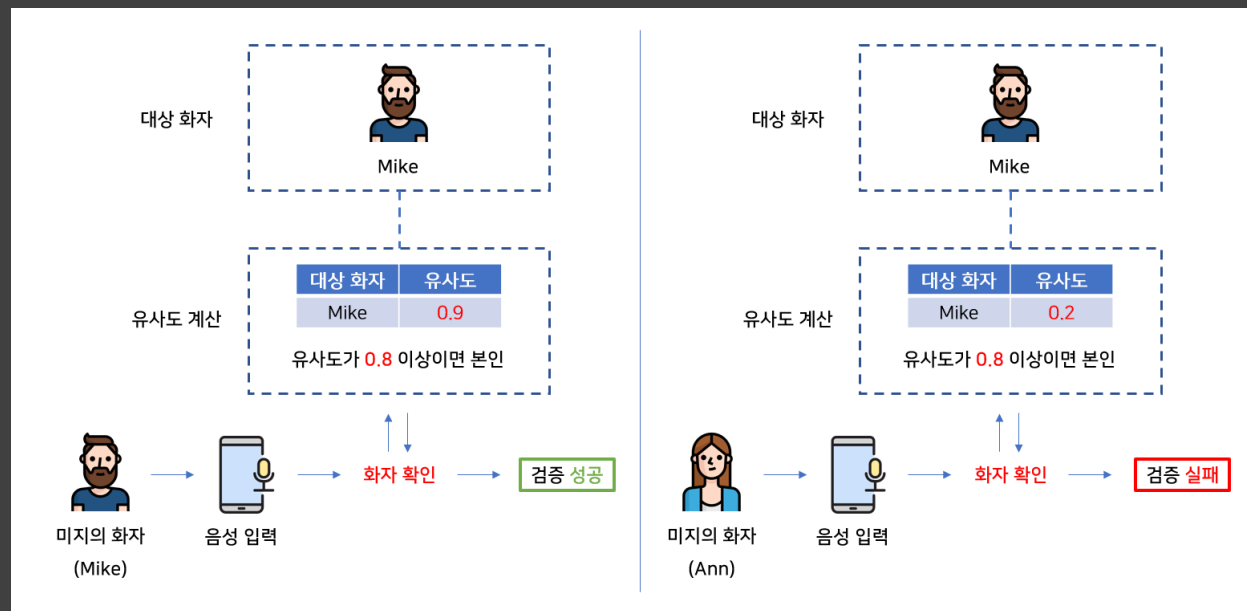
MFCC (Mel-Frequency Cepstral Coefficient)

오디오 신호에서 추출할 수 있는 특징으로 소리의 고유한 특징을 나타내는 수치

## MFCC 예시

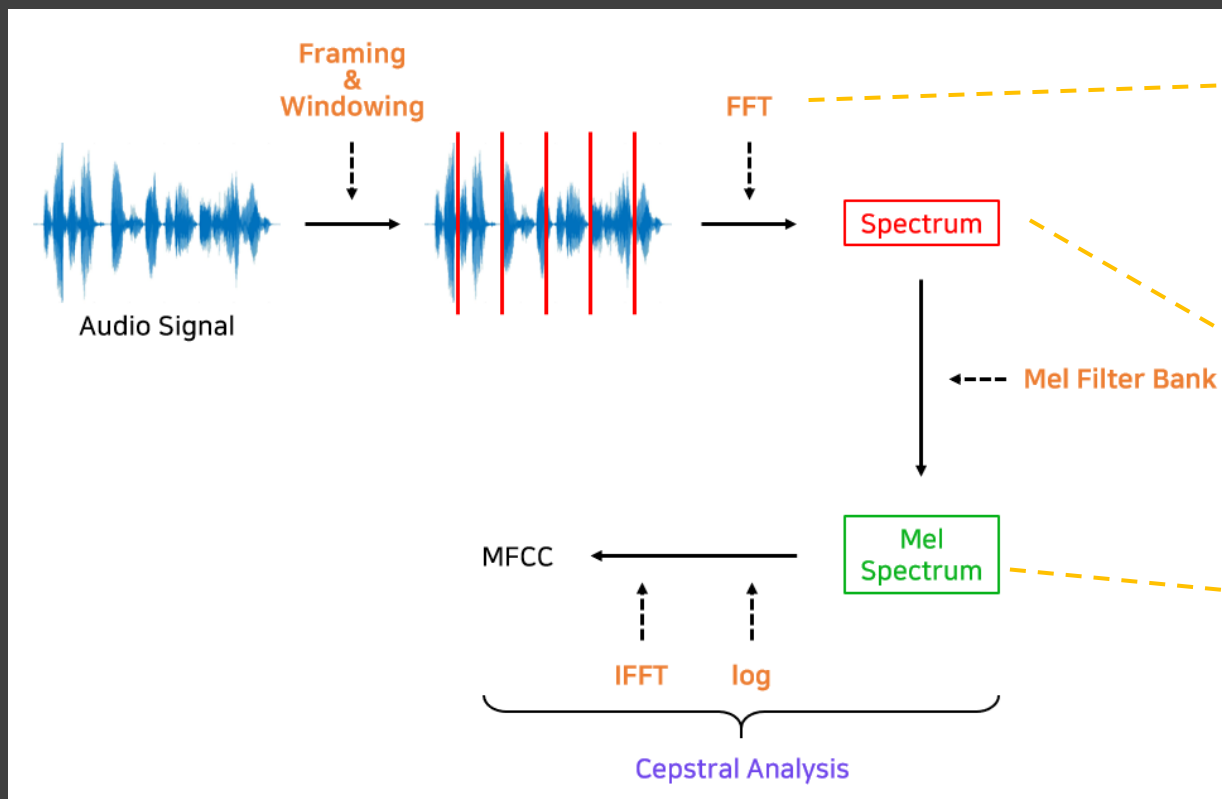
### 화자 검증

1. 말하는 사람이 그 사람이 맞는지 확인하는 기술  
ex) 시스템에 등록된 음성에만 반응하는 아이폰의 Siri
2. 등록된 음성과 현재 입력된 음성의 유사도를 판별하는  
근거의 일부로 사용



## MFCC 추출과정

MFCC는 Mel Spectrum(멜 스펙트럼)에서 Cepstral(켄스트럴)분석을 통해 추출된 값



1. 오디오 신호를 프레임별로 나눔 FFT를 적용해 Spectrum을 구함

[참고]

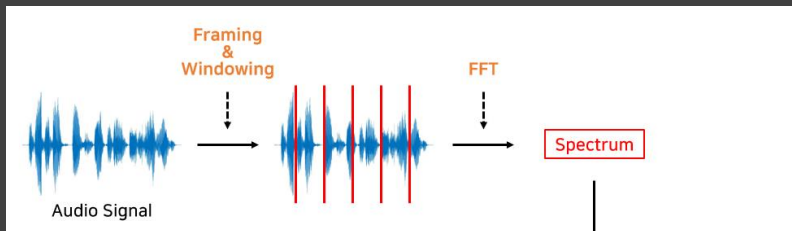
FFT(고속 푸리에 변환)

: 신호를 주파수 성분으로 변환하는 알고리즘

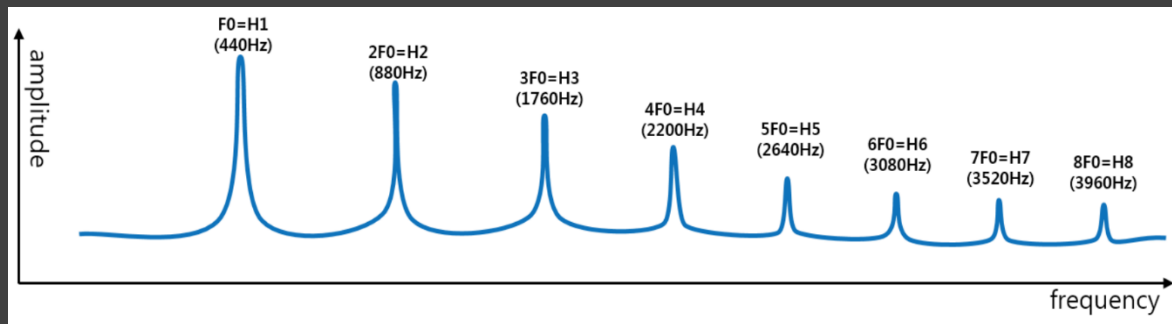
2. Spectrum에 Mel Filter Bank를 적용해 Mel Spectrum을 구함

3. Mel Spectrum에 Cepstral 분석을 적용해 MFCC를 구함

## Spectrum



1. 오디오 신호에 FFT를 수행하면 주파수 영역(frequency domain) 표현이 가능해지고 이를 Spectrum이라고 함
2. 각 주파수 대역별 세기를 알 수 있으므로 신호에서 어떤 주파수가 강하고 약한 지 알 수 있음 -> 소리의 고유한 특징을 추출



### ▶ 배음(harmonics) 구조

1. 소리 : 기본 주파수(fundamental frequency) + 기본 주파수의 정수배인 배음(harmonics)들  
ex) 4옥타브 '라' 음의 소리 = 기본 주파수(440Hz) + 정수배인 880Hz, 그 다음 배음들까지 포함
2. 배음 구조는 악기나 성대의 구조에 따라 달라지며 배음 구조의 차이가 음색의 차이를 생성
3. Spectrum에서 배음 구조를 유추해낼 수 있다면 소리의 고유한 특징을 찾아낼 수 있음  
-> 이를 가능하게 하는 것이 **Cepstral Analysis**



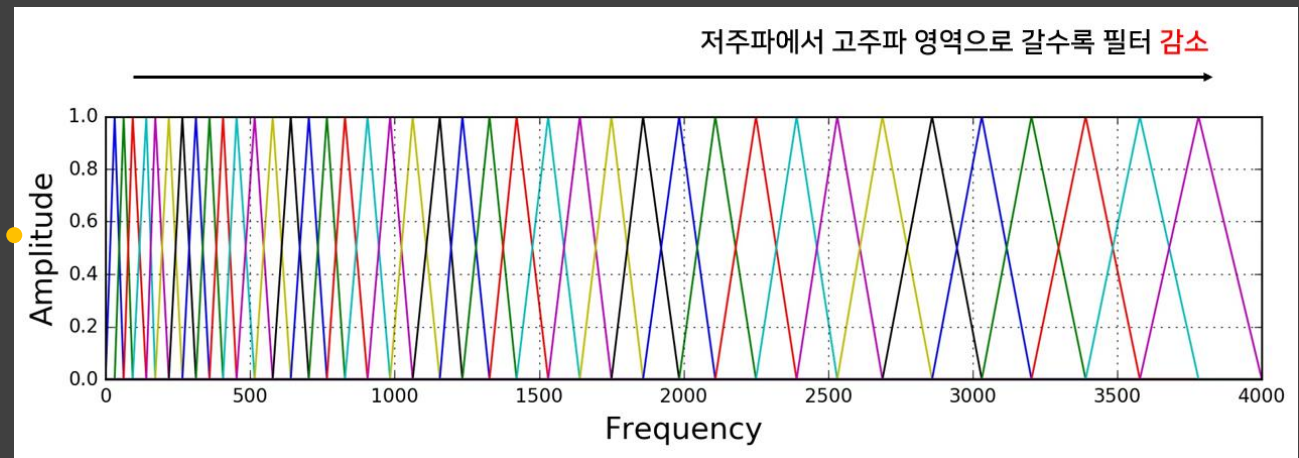
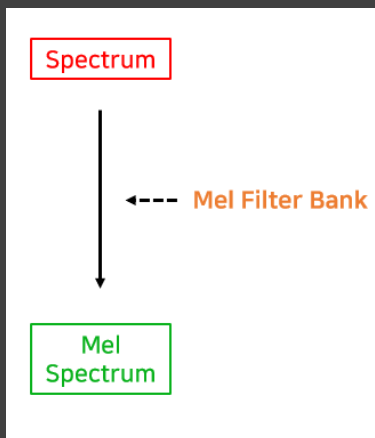
## Mel Spectrum

### 1. Mel Scale

: 사람의 청각 기관이 고주파수보다 저주파수 대역에서 더 민감하다는 특성을 반영해 물리적 주파수와 실제 사람이 인식하는 주파수의 관계를 표현한 것  
Mel Filter Bank를 나눌 때 어떤 간격으로 나눌지 알려주는 역할

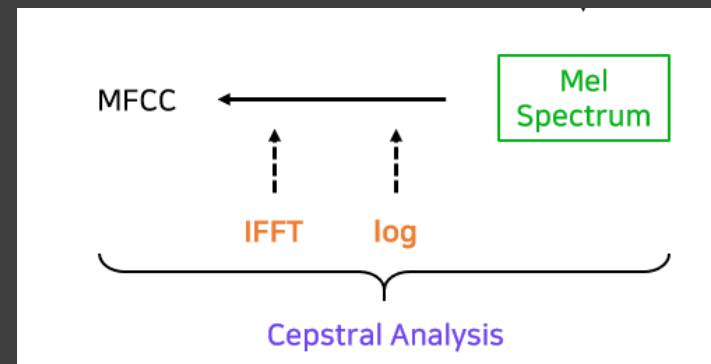
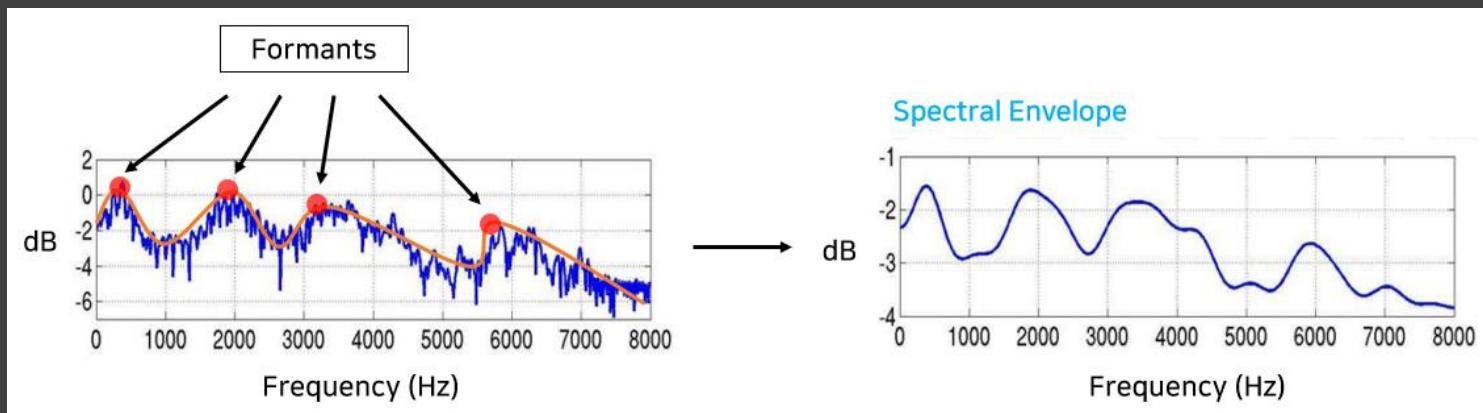
### 2. Mel Spectrum

: Mel Scale에 기반한 Filter Bank를 Spectrum에 적용하여 도출해낸 것





## Cepstral Analysis



## 포먼트

스펙트럼의 피크, 소리가 공명되는 특정 주파수 대역

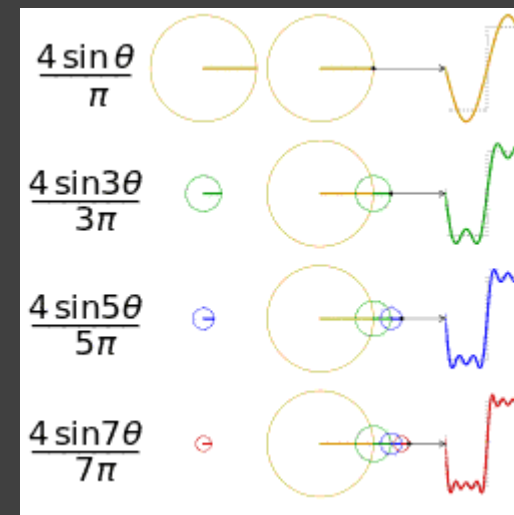
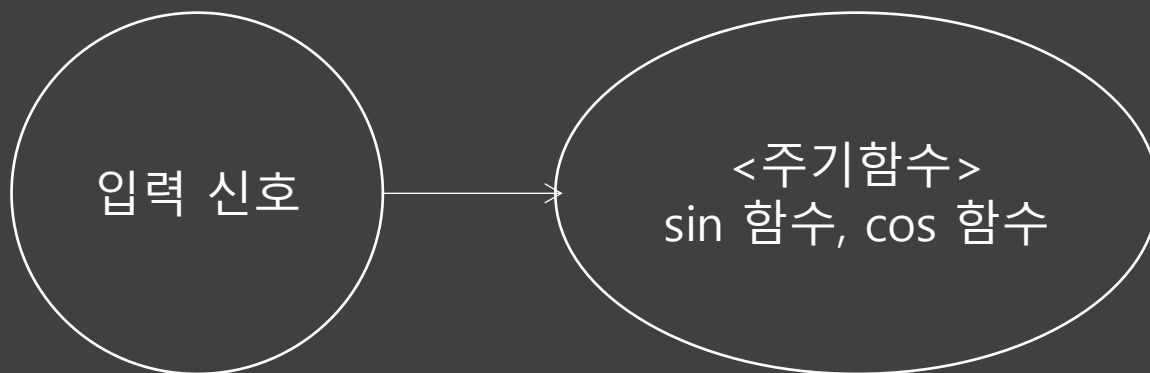
사람의 음성은 성대에서 형성되어 성도(vocal track)를 거치며 변형되는데, 소리는 성도를 지나면서 포먼트를 만나 증폭되거나 감쇠됨  
포먼트는 배음과 만나 소리를 풍성하게 혹은 선명하게 만드는 필터 역할 -> 소리의 특징을 유추할 수 있는 중요한 단서

포먼트들을 연결한 곡선과 Spectrum을 분리한 곡선을 Spectral Envelope(스펙트럼 포락선)이라 하고 이를 분리하는 과정에서 MFCC가 도출됨  
두 곡선을 분리하는 과정에서 Mel Spectrum에 log와 IFFT(역 고속 푸리에 변환)이 적용됨

## 푸리에 변환

### 푸리에 변환

임의의 입력 신호를 다양한 주파수를 갖는 주기 함수들의 합으로 분해하여 표현하는 것



음성 데이터

특정 시간 간격으로 분해

푸리에 변환 적용

## 푸리에 변환의 직관적 이해

임의의 신호

