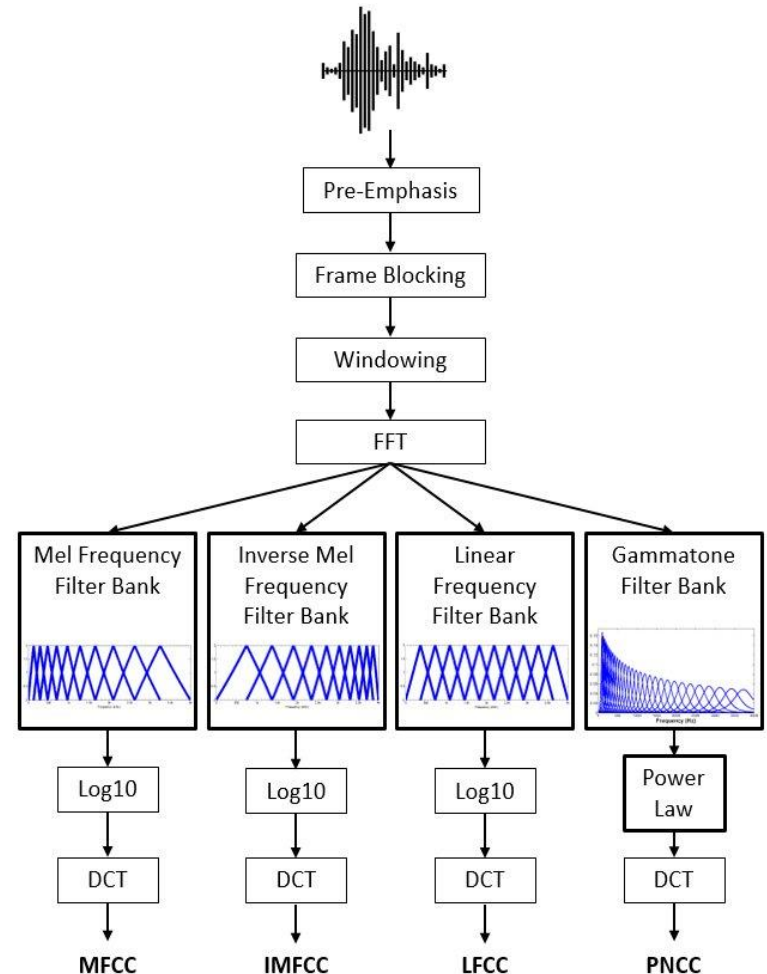
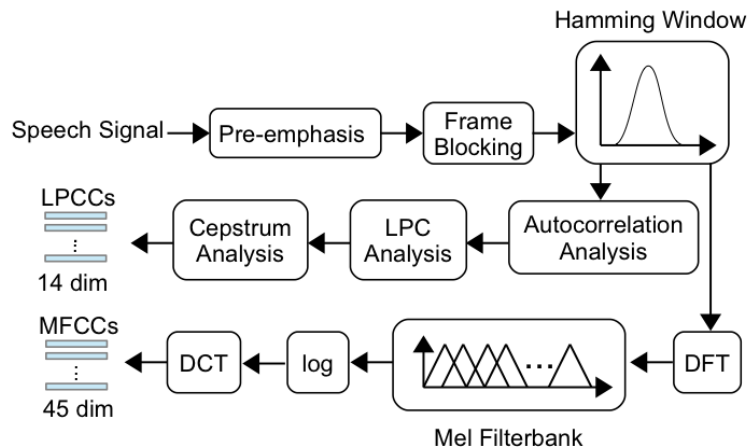


# Mel-Scale Frequency Cepstral Coefficients

# Features for speaker recognition

- 화자인식, 음성인식, 그밖에 음성을 이용한 다양한 정보 추출 및 패턴 인식을 위해 많이 사용되는 특징
  - MFCC: 1980 대 Davis와 Mermelstein 에 의해 처음 소개 되었으며 지금까지도 많이 사용되고 있음.
  - 그밖에 LPCC, LPC, LFCC, PNCC 등 다양한 특징벡터들이 존재함.

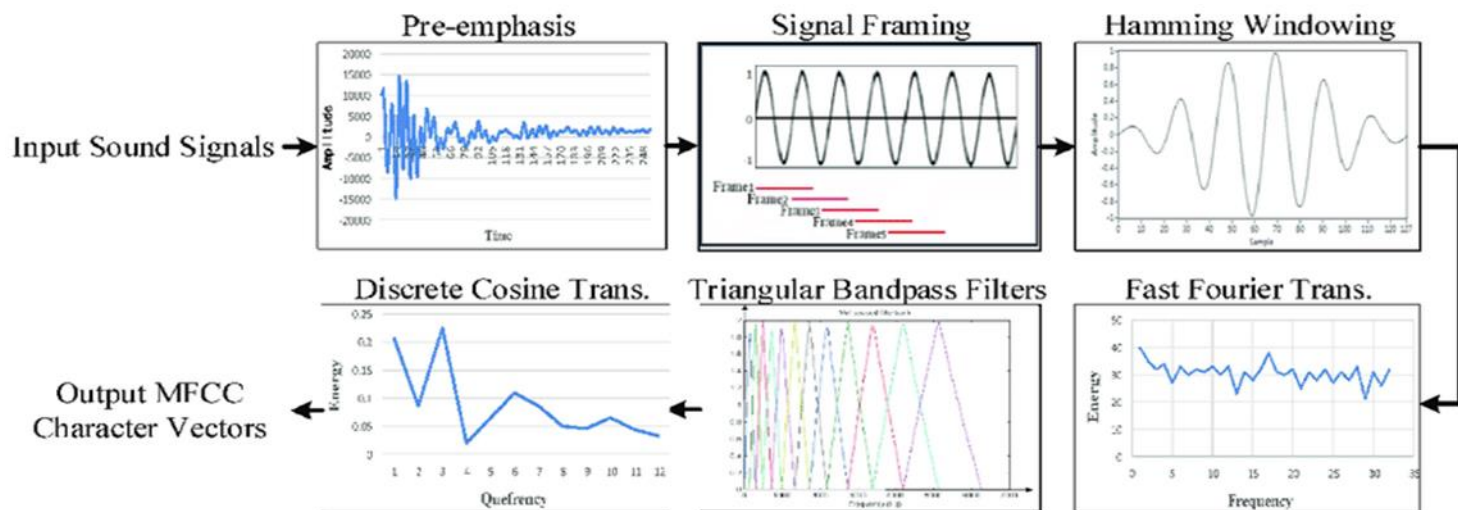


# Mel-Scale Frequency Cepstral Coefficients

## ➤ MFCC

- 사람의 청각 시스템을 모방한 Mel scale frequency에 기반한 방법임.
- 아래와 같이 6가지 단계로 나눌 수 있다.

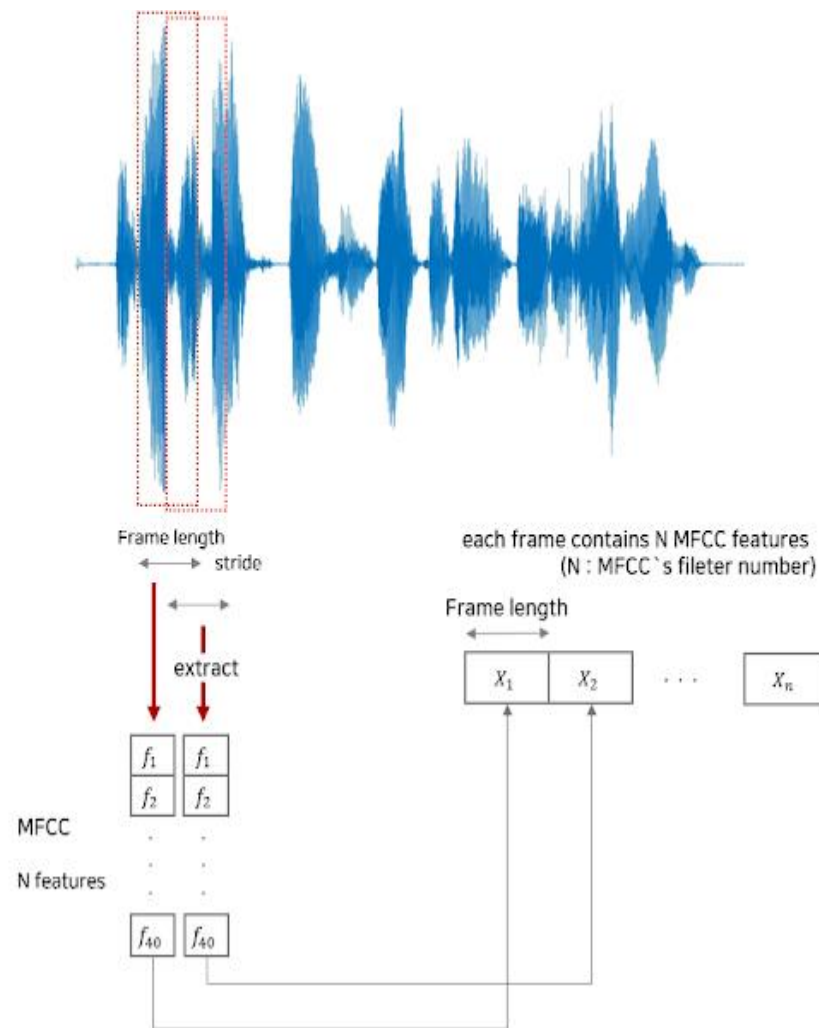
1. 입력 시간 도메인의 소리 신호를 작은 크기 프레임으로 자름.
2. 각 프레임에 대하여 푸리에 변환을 계산함.
3. 2번에서 구한 결과에 Mel Filter bank를 적용하고, 각 필터에 에너지를 합함.
4. 3번에서 구한 모든 필터 뱅크 에너지의 Log를 취함.
5. 4번 값에 DCT를 취함.



# Framing

## ➤ Frame

- 사람의 시간영역에서 입력된 소리 신호는 지속적으로 변화함.
- 이런 변화하는 소리는 간단히 하기 위하여 짧은 시간 내에서는 소리 신호가 많이 변하지 않는다고 가정함.
- 실제로 신호의 변화가 없다는 것이 아니라, 짧은 구간에서 통계적으로 변화가 거의 없다는 것임.
- 일반적으로 프레임의 길이는 20~40ms 정도로 함.
- Sampling Rate에 따라 샘플의 수는 달라지겠지만, 44.1KHz 기준으로 프레임 당 약 천 개의 데이터 포함.
- 샘플의 수가 너무 적으면 주파수 분석에 신뢰도가 떨어지고, 너무 많으면 신호의 변화가 한 프레임 내에서 너무 크게 되기 때문에 분석이 어려움.



# Cepstral Analysis

## ➤ Cepstrum

### • 스펙트럼

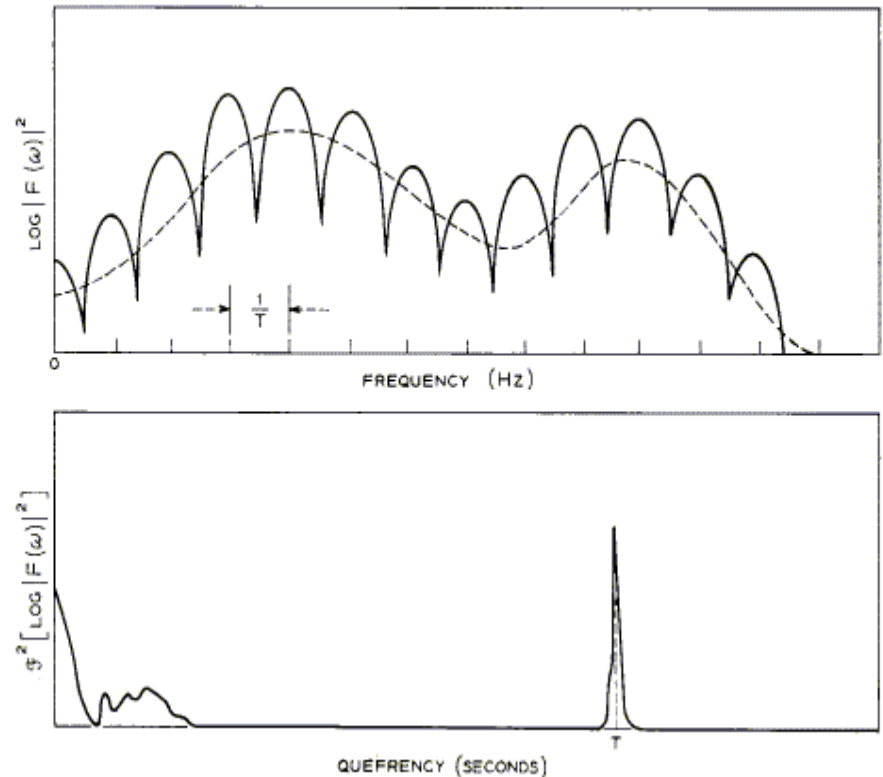
-시간의 영역에서 표현된 신호가 푸리에 변환을 거쳐 얻어진 결과인 주파수 영역이고, 이 주파수 영역을 데이터를 그래프로 표현한 것임.

### • 캡스트럼

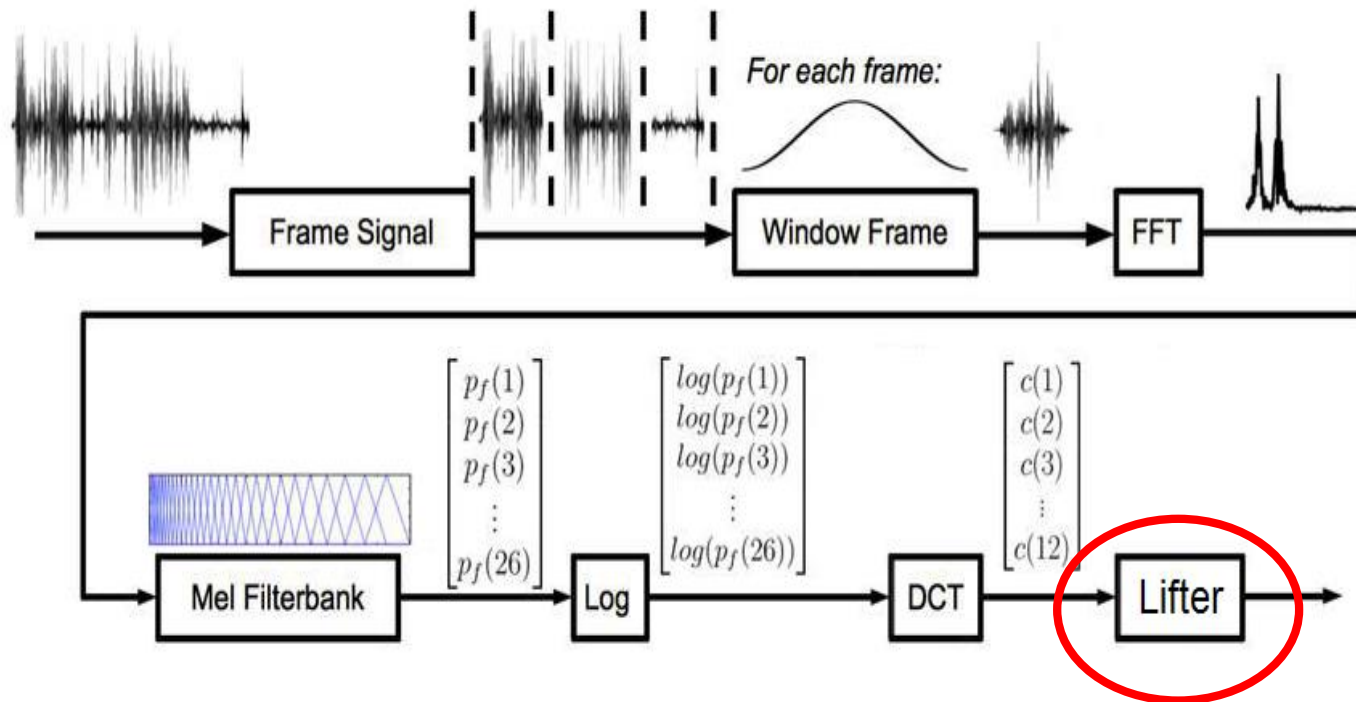
-스펙트럼 데이터에 로그를 취하고 이를 역푸리에 변환하면 얻어짐.

-Log의 역할은 두 개의 곱으로 형성된 함수를 합의 형태로 변화한 후 분리해 낼 수 있게 해줌

-특이한 것은 스펙트럼에서의 Frequency 대신 Quefrequency라는 용어를 사용함.



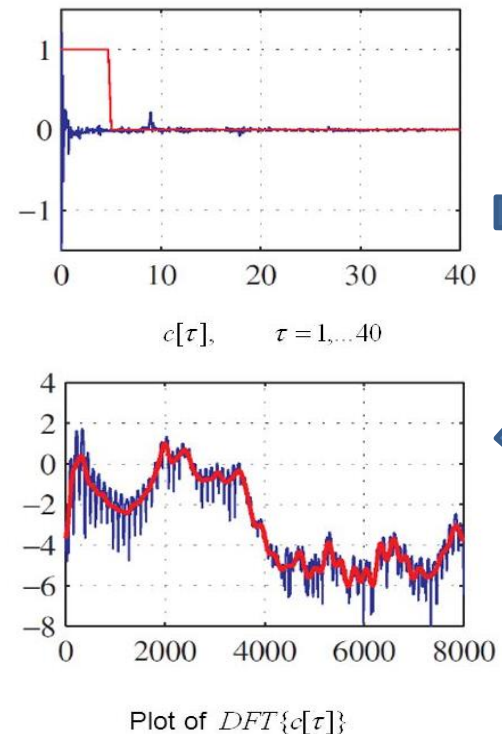
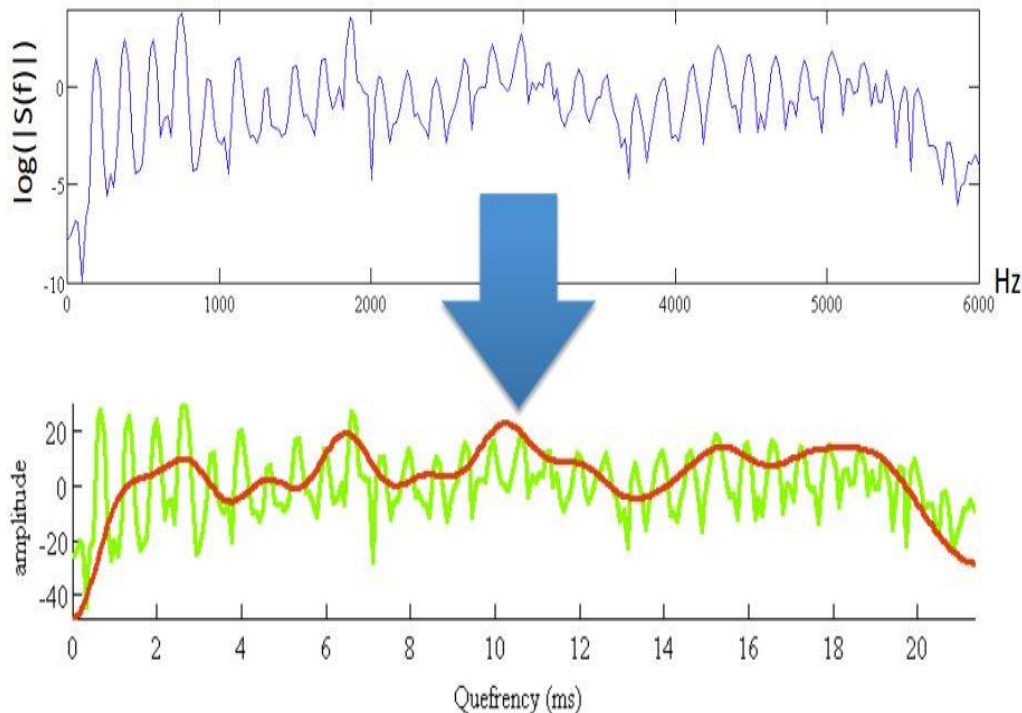
# Cepstrum & Lifter



# Liftering

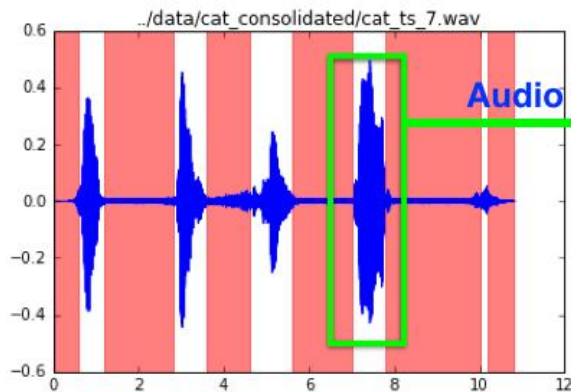
## ➤ Spectral Envelope Extraction

- **Liftering** operation is similar to filtering operation in the frequency domain where a desired quefrency region for analysis is selected
- By removing high quefrency components (to remove glottal excitation), it takes a DFT results in spectral envelope of the signal.



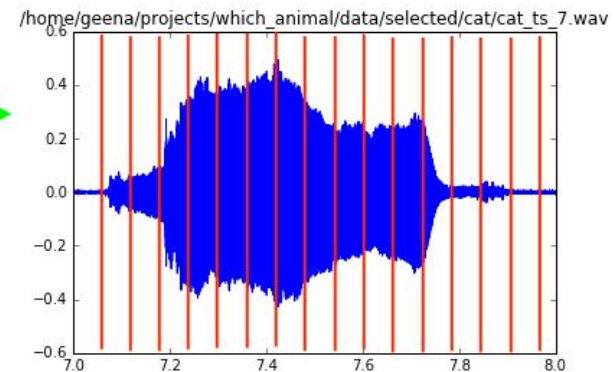
# MFCC

**Peak detection & slicing**



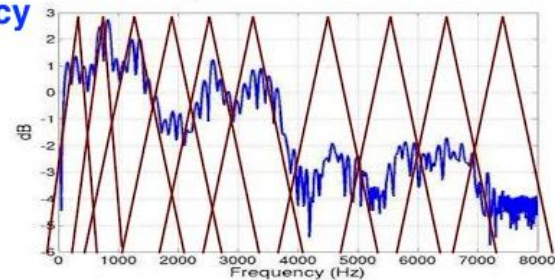
Audio signal

**Fast Fourier Transform**



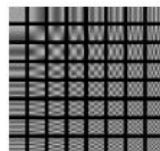
Spectrum

**Mel-Frequency  
Spectrum**



**Mel Scale Filtering**

**Mel-Frequency  
Cepstral  
Coefficients  
(MFCC)**



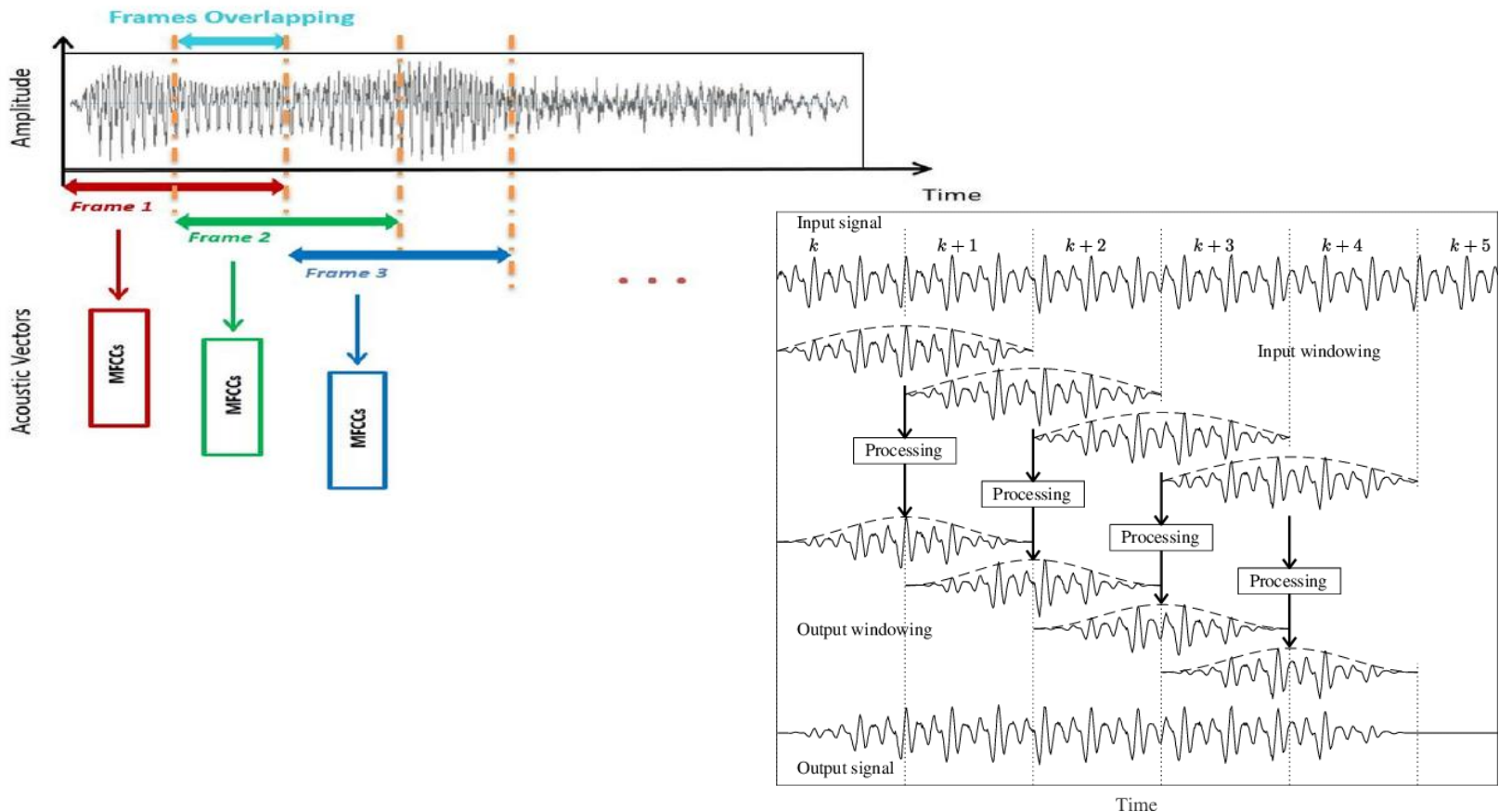
**Discrete  
Cosine  
Transform**



# MFCC 구현 시 고려사항 1

## ➤ Frame

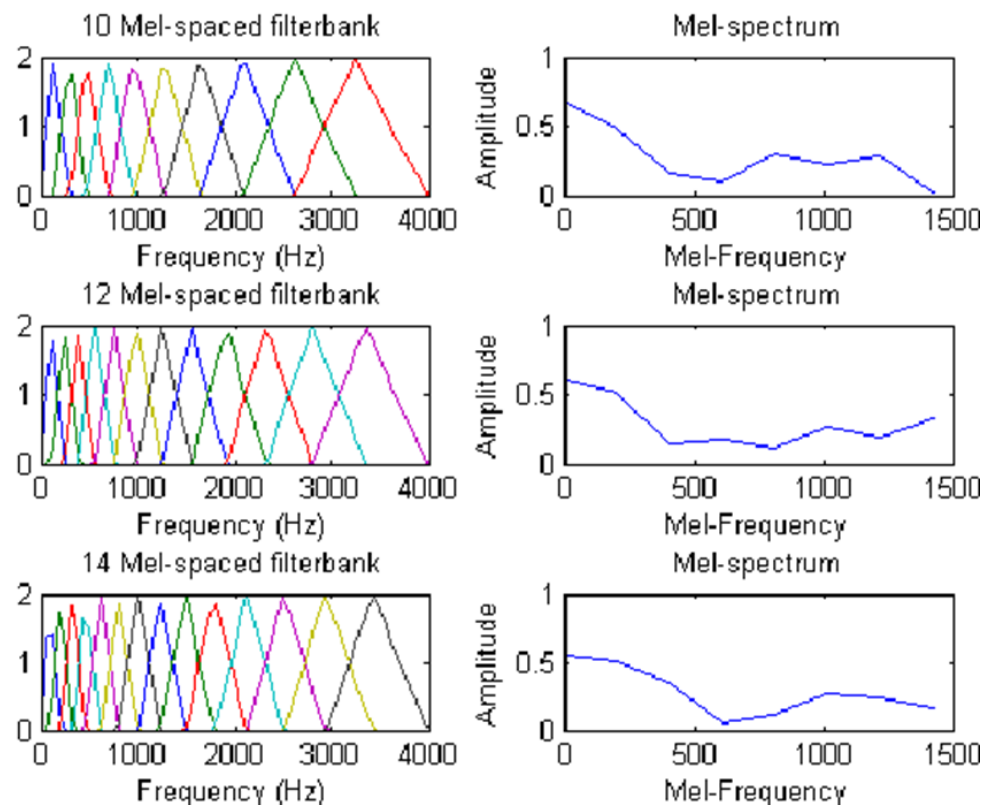
- Frame의 길이와 이웃한 frame 간의 overlap 길이를 정해야 함. 일반적으로는 frame length 20-30msec, overlap 10msec 정도임.



# MFCC 구현 시 고려사항 2

## ➤ Filterbank

- Filter의 개수, 또는 channel의 개수를 정해야 하는데, 정답은 없는 게 현실...
- 보통은 12, 19, 24, 그 이상의 개수도 사용하고 연구개발 대상과 개발자의 직감(?)으로 결정함.



# MFCC in detail

