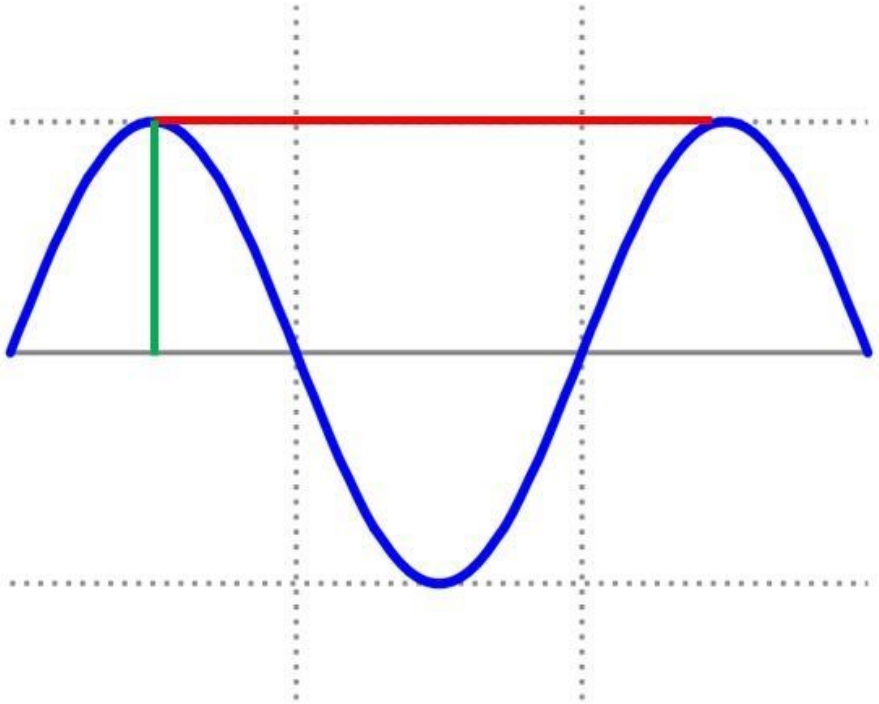
MFCC

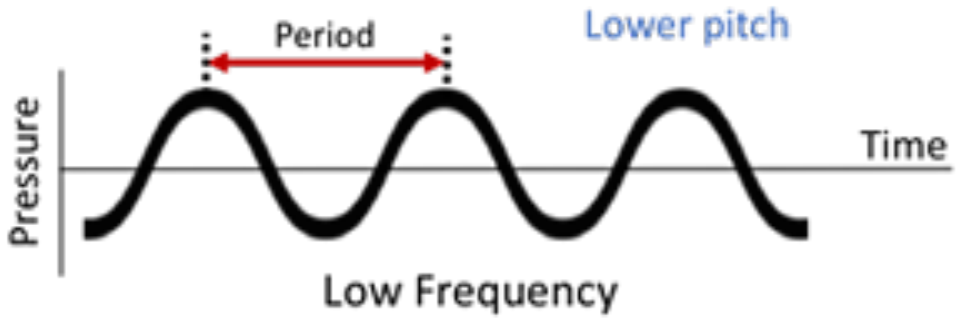
20200192S2 김나현

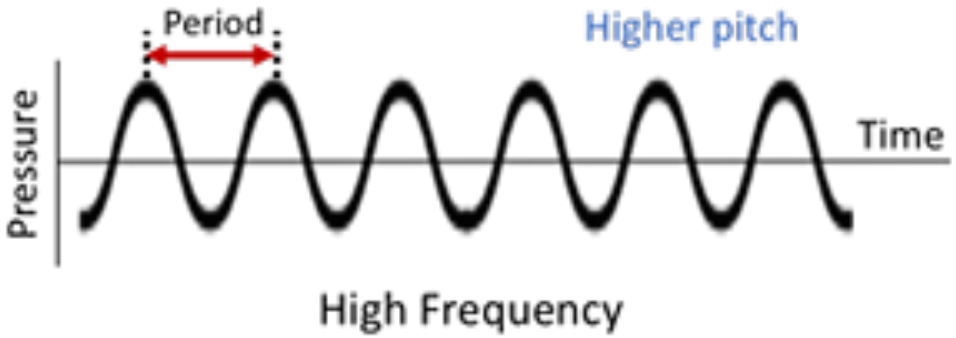
Defining Waves



- Phase(Degress of displacement) : 위상
- Amplitude(Intensity) : 진폭
- Frequency : 주파수

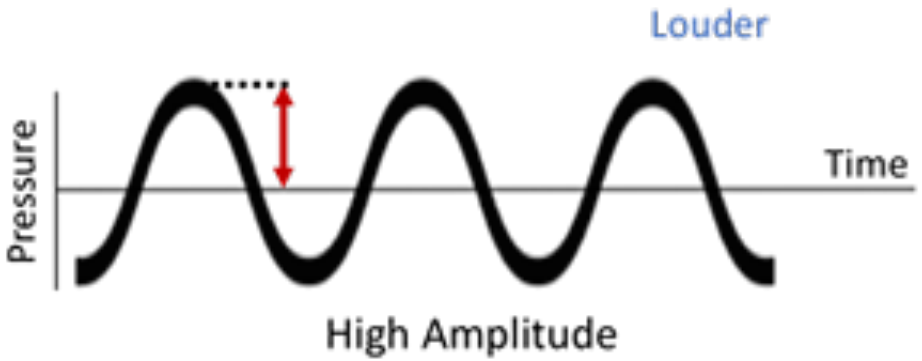
Frequency





•High frequency → High pitch

Amplitude



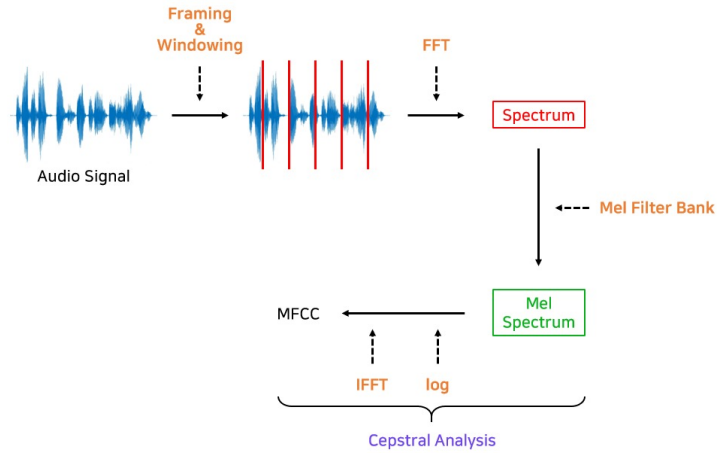
•High Amplitude → Loud sound

Energy

- waveform이 가지고 있는 에너지 값을 의미
- 즉 signal의 전체 amplitude에 대응되는 값
- signal의 각 amplitude 포인트를 x(n)이라고 할 때,

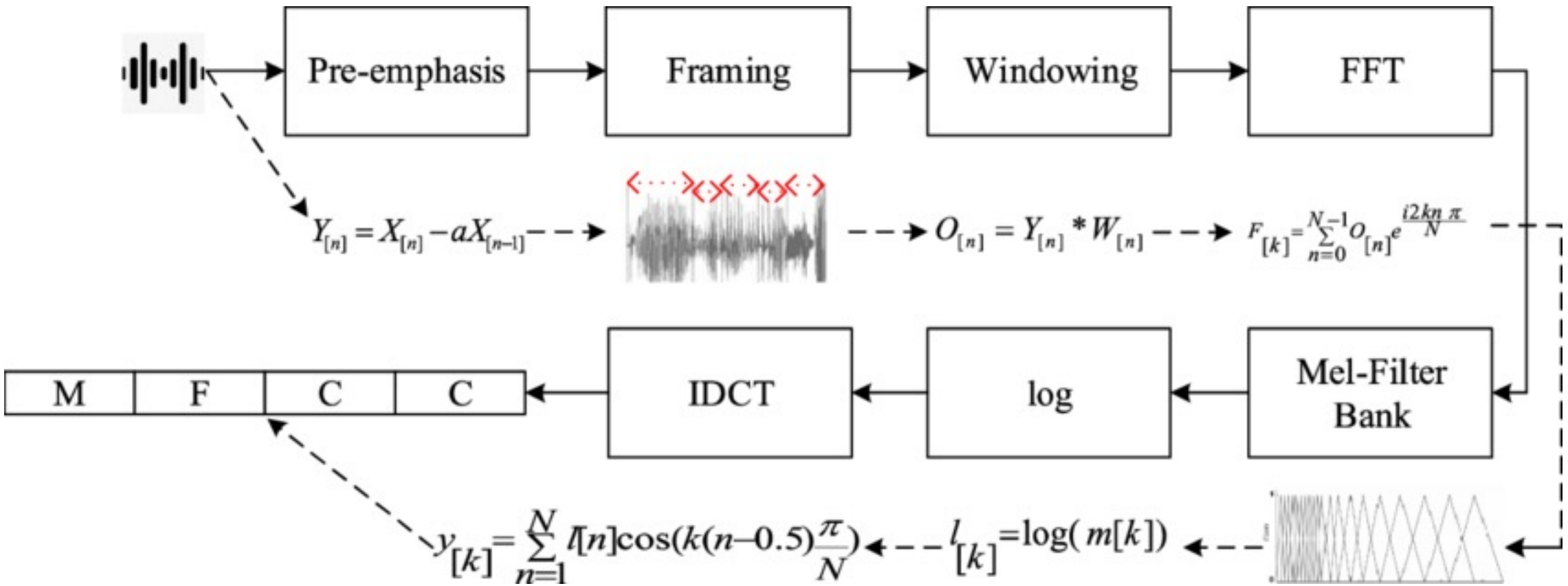
- signal의 energy는 $\sum_n |x(n)|^2$

root-mean-square energy(RMSE) : $\sqrt{\frac{1}{N}\sum_n |x(n)|^2}$

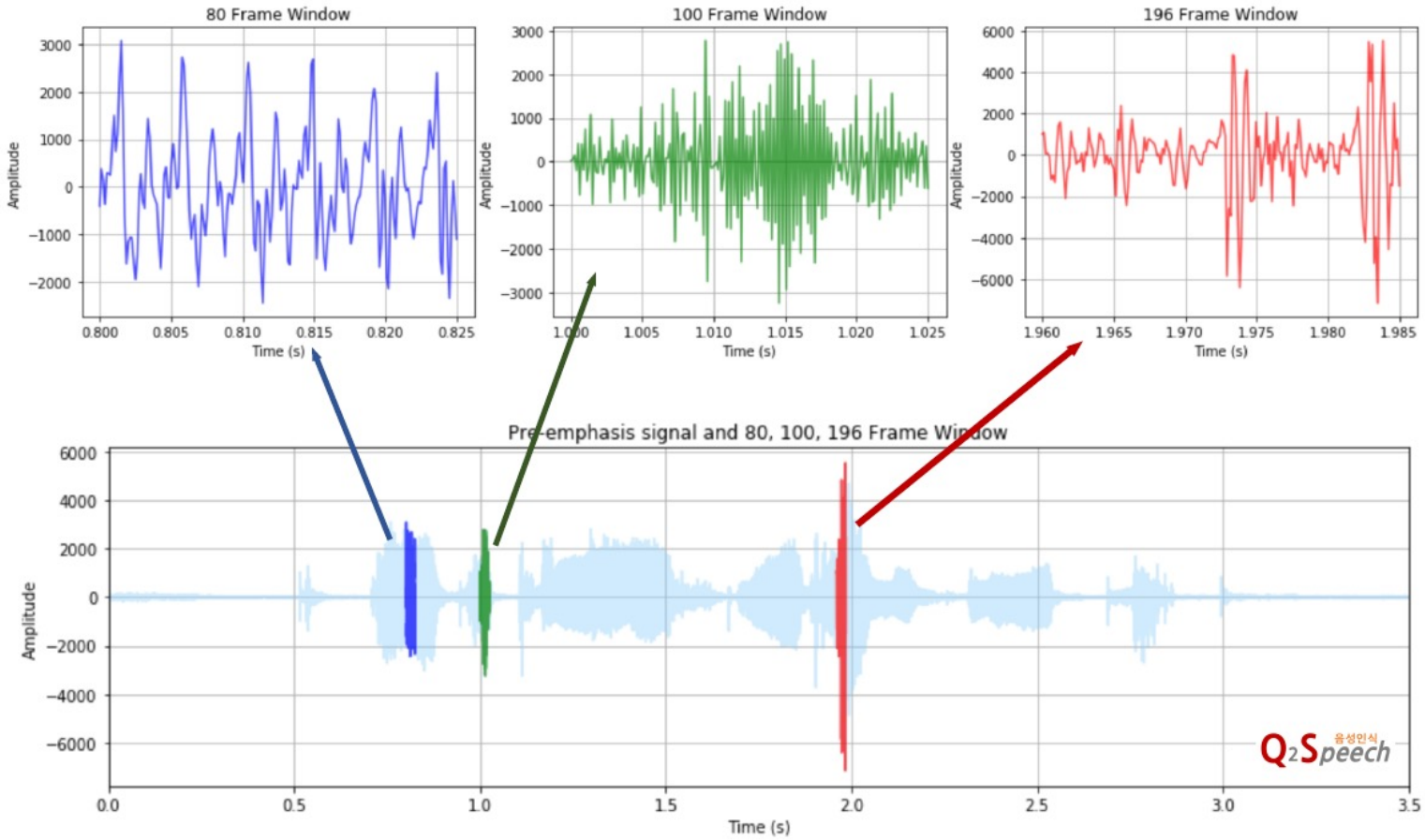


MFCC
(Mel Frequency Cepstral Coefficients)

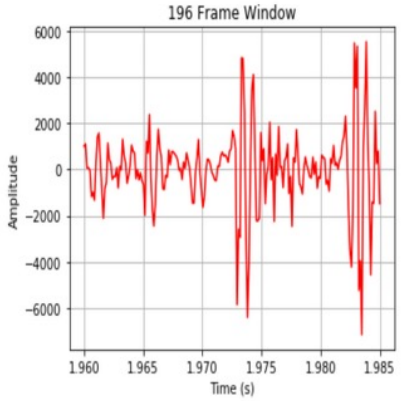
The diagram of MFCC algorithm



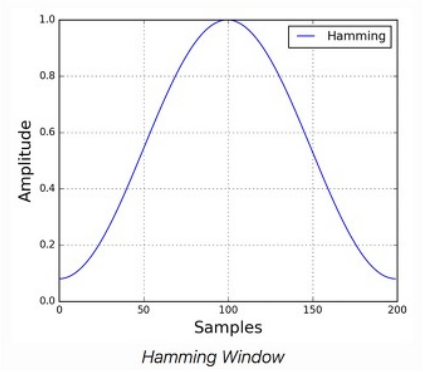
framing

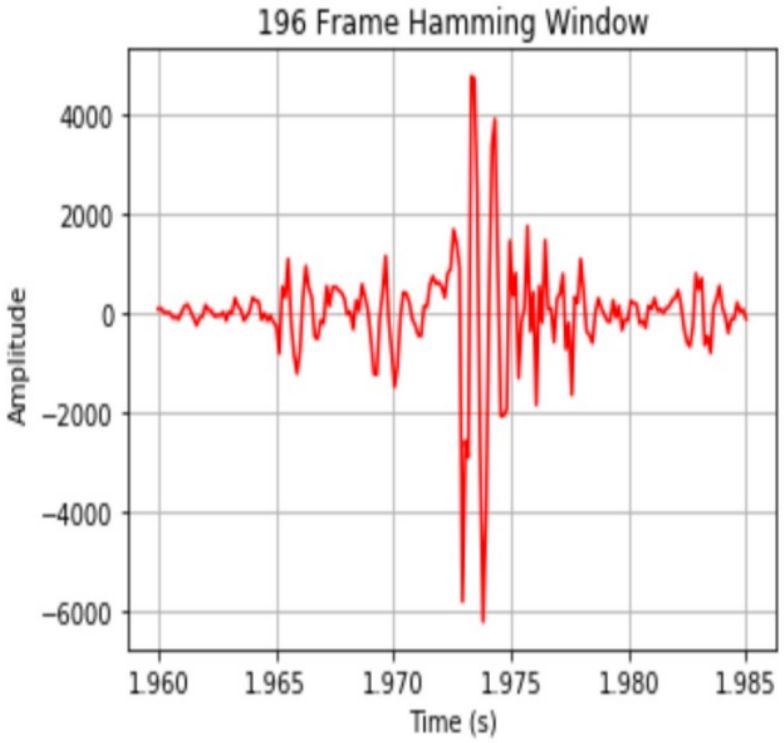

80 Frame Window
100 Frame Window
196 Frame Window
Amplitude
Time (s)
Pre-emphasis signal and 80, 100, 196 Frame Window
Q2Speech 음성인식

framing

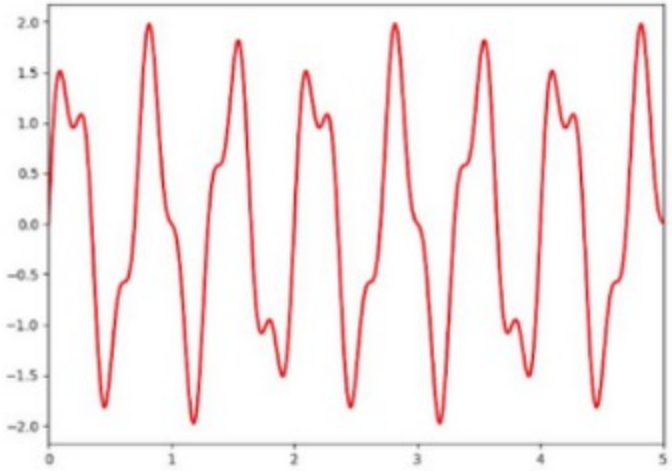


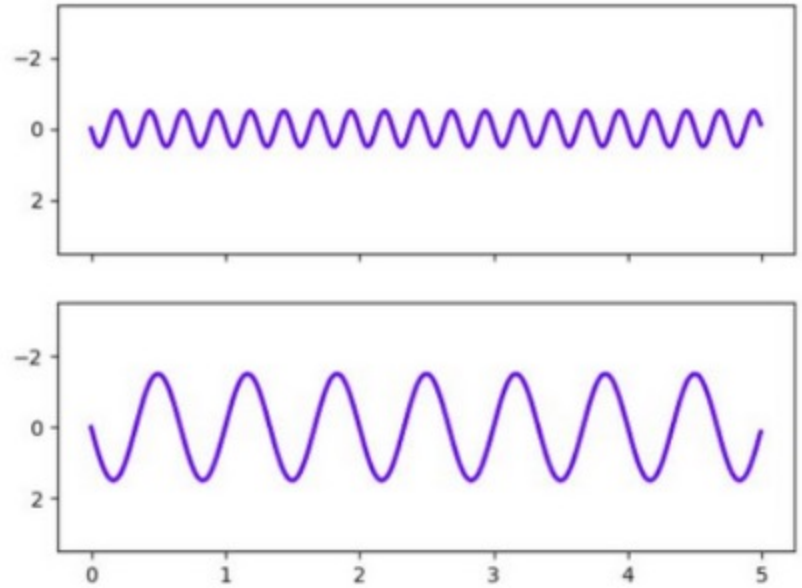
$$w[n] = 0.54 - 0.46cos(\frac{2\pi n}{N-1})$$





Fourier Transform





$A_1 = 0.5, f_1 = 4, \varphi_1 = 0$

$A_2 = 1.5, f_2 = 1.5, \varphi_2 = 0$

$$s = A_1 \sin(2\pi f_1 t + \varphi_1) + A_2 \sin(2\pi f_2 t + \varphi_2)$$

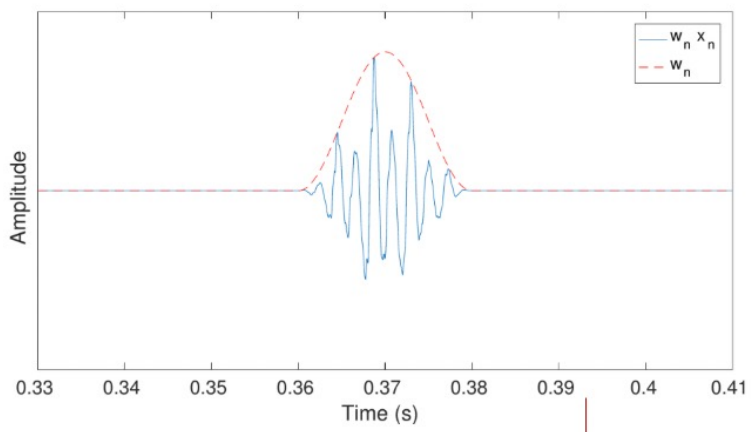
Fourier Transform

푸리에 변환의 단점

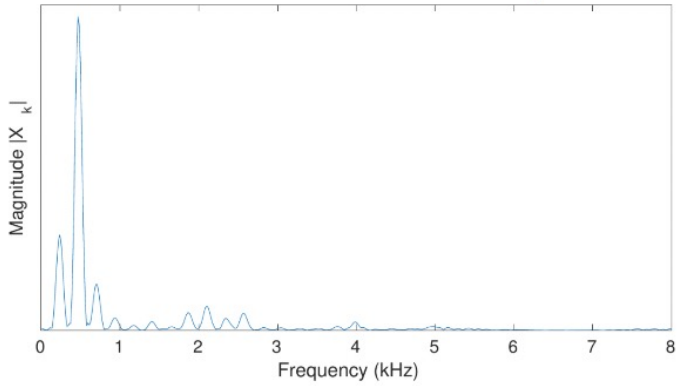
: 시간에 대한 연속성이 고려되지 않음으로써 많은 문제가 야기된다!

STFT
(Short-time Fourier transform)

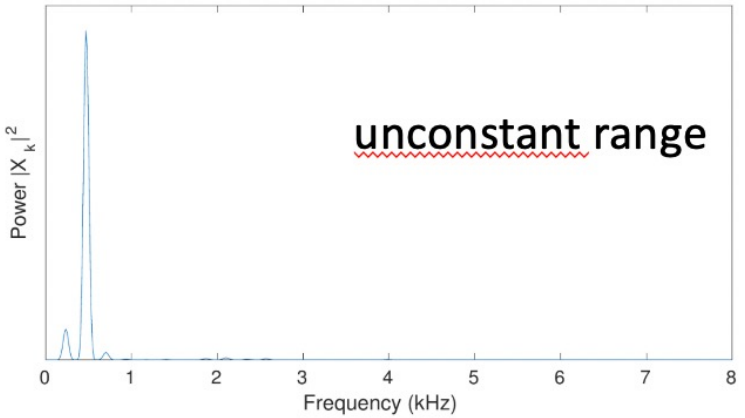
Fourier transform of a frame



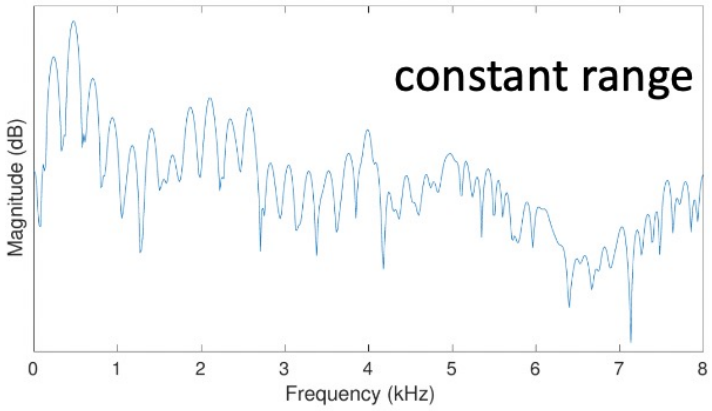
single frame



spectrum



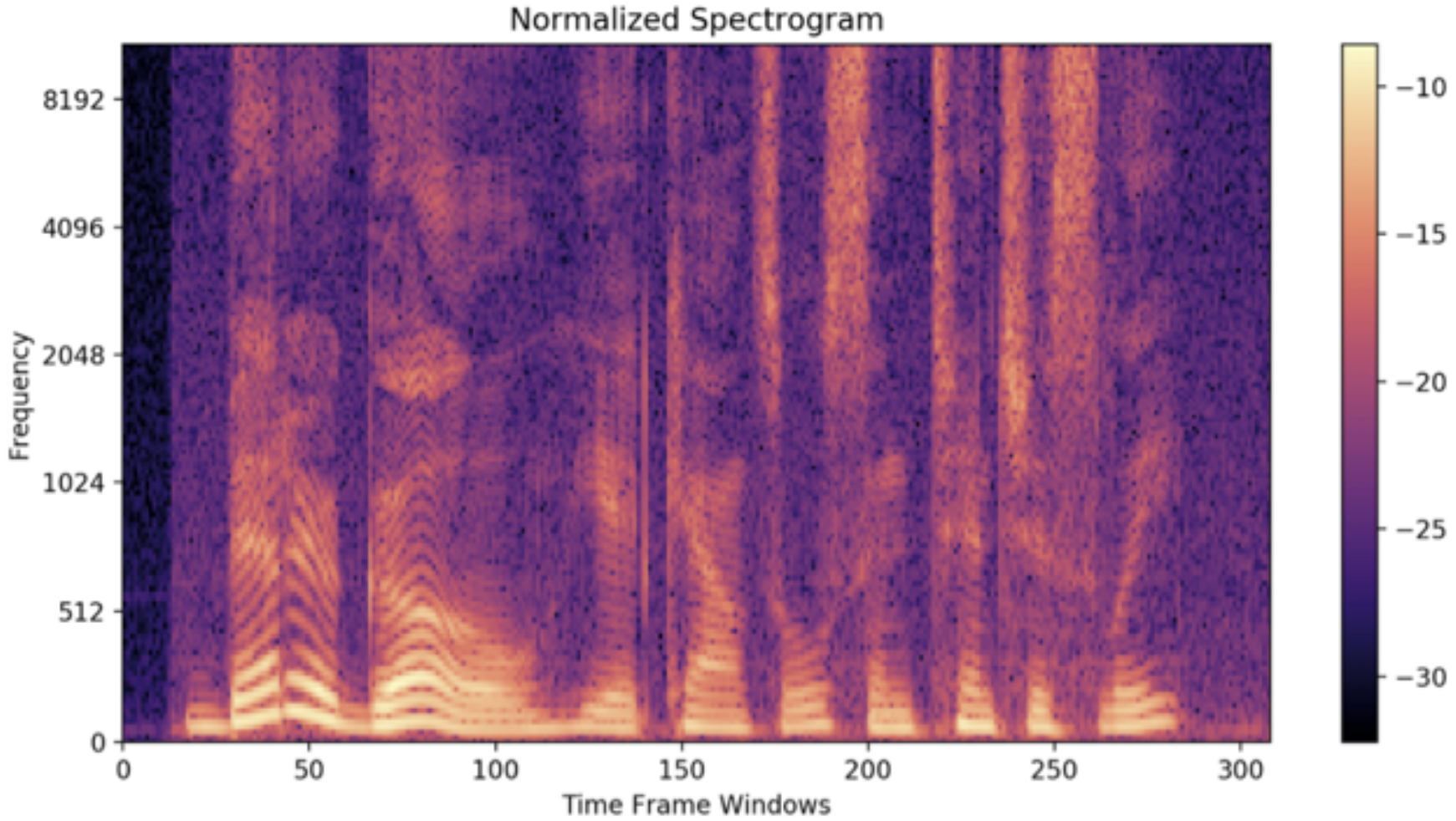
power spectrum



Log-spectrum

STFT

Each Fourier transformed frame is stretched vertically in chronological order.



spectrum에서 소리의 고유한 특징을 추출하기 (cepstral 분석)

How?

fundamenial frequency

Voice = fundamental frequency(F0) + Harmonics

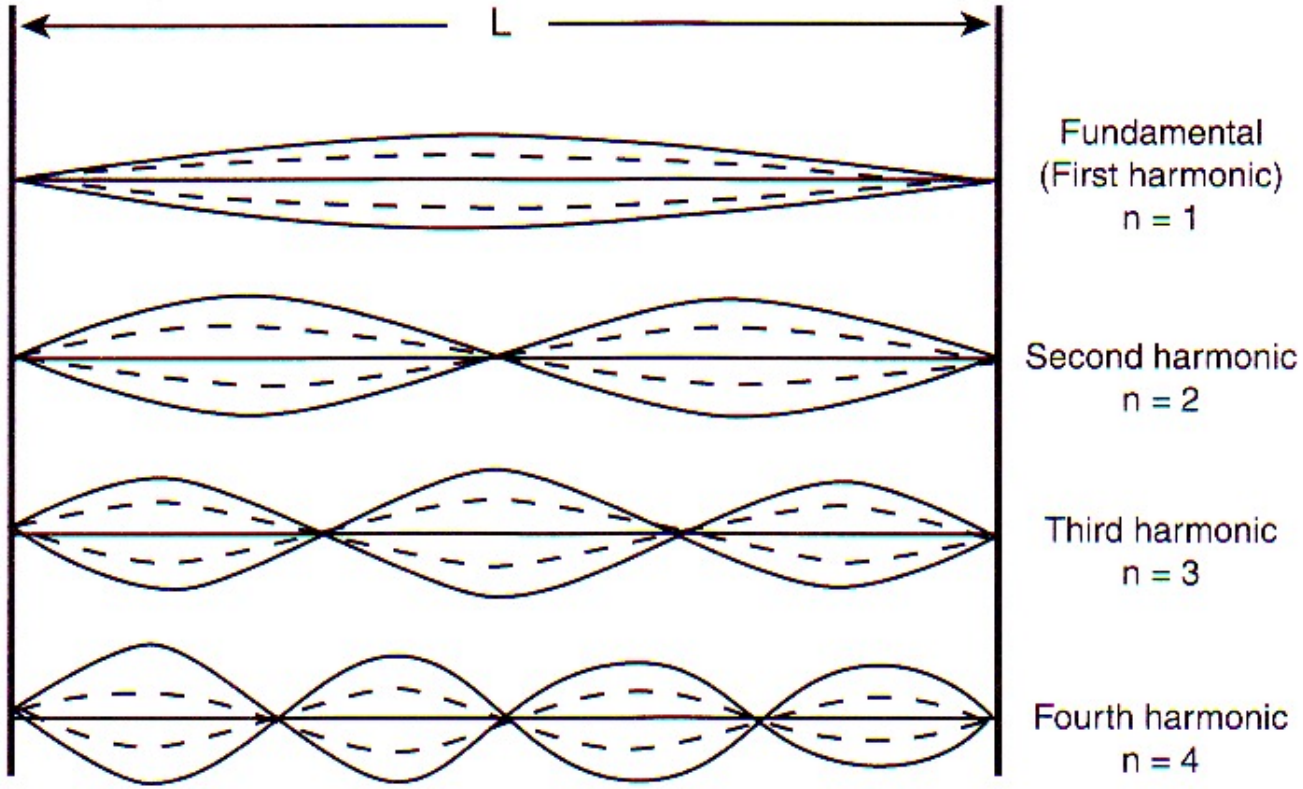
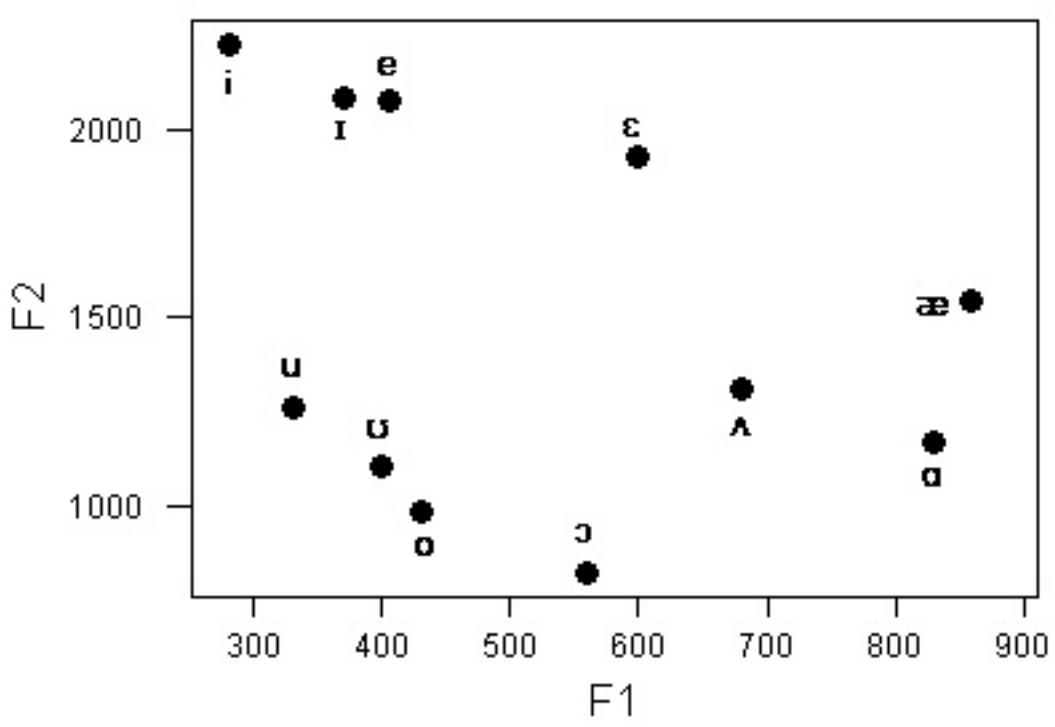


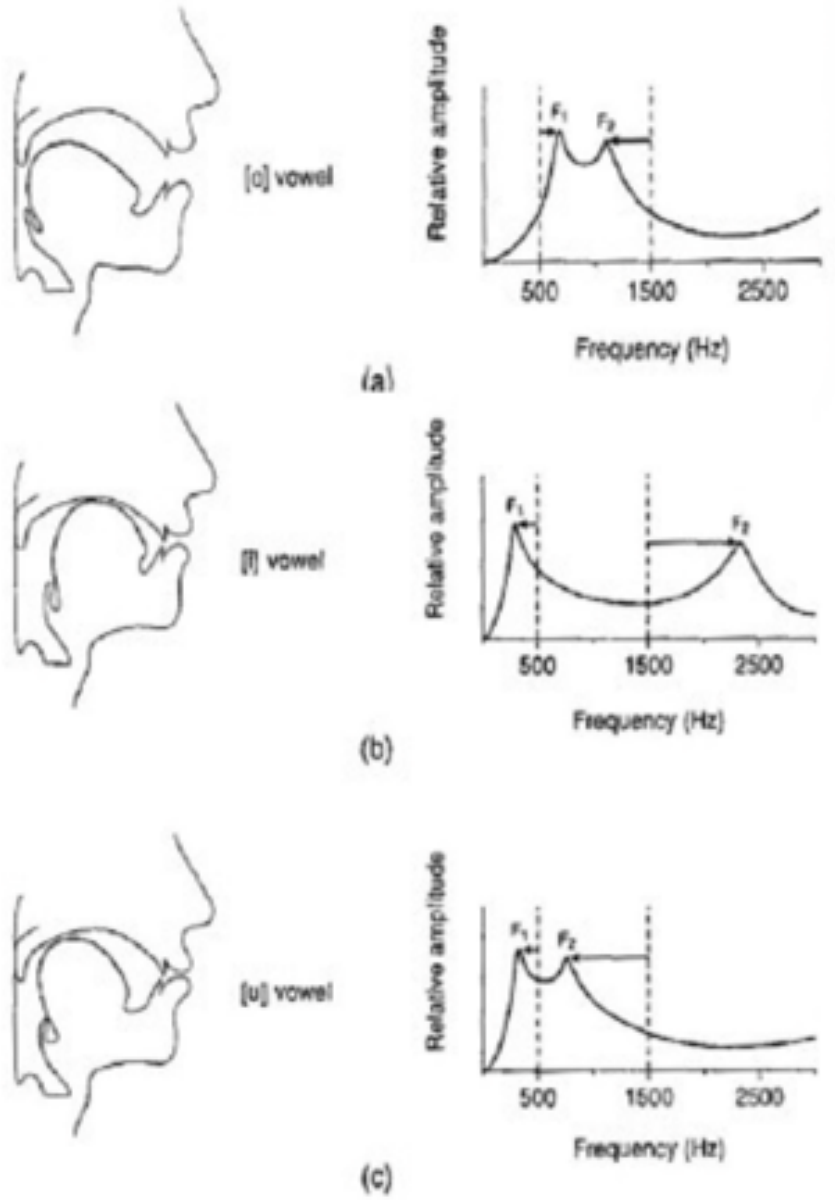
Figure 5

formants

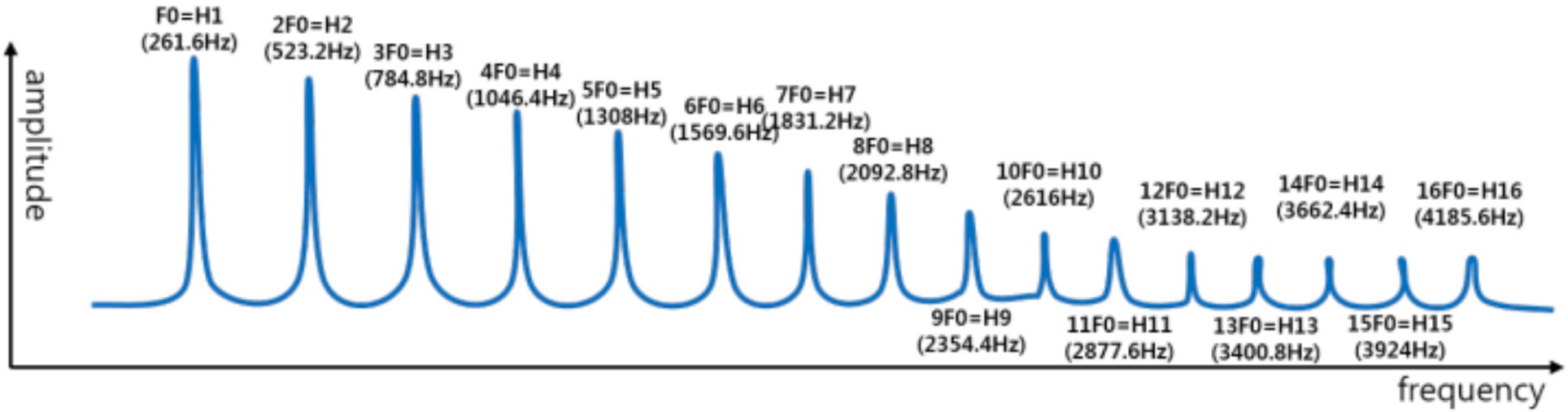
소리가 공명되는 특정 주파수 대역



articulation

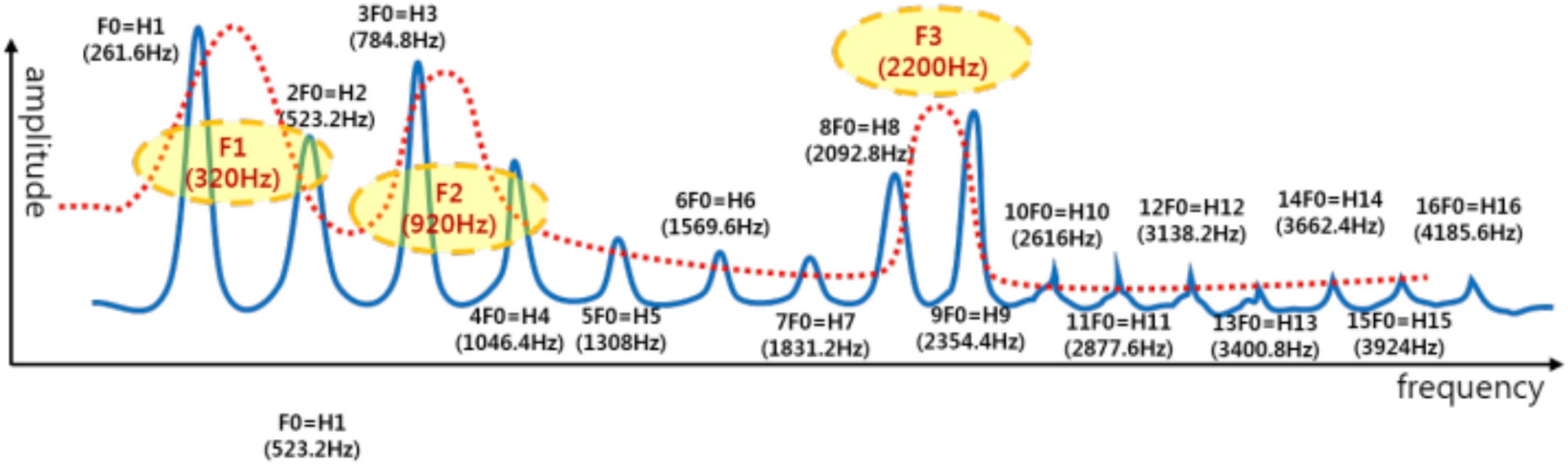


On C4



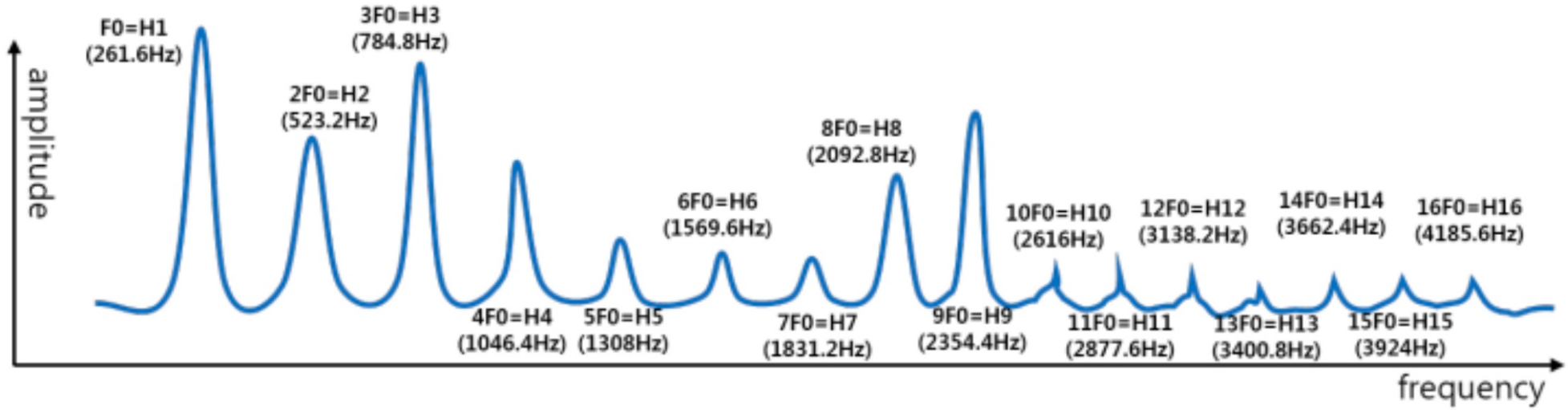
Vowel	F1(Hz)	F2(Hz)	F3(Hz)
iː	280	2620	3380
ɪ	360	2220	2960
e	600	2060	2840
æ	800	1760	2500
ʌ	760	1320	2500
ɑː	740	1180	2640
ɒ	560	920	2560
ɔː	480	760	2620
ʊ	380	940	2300
uː	320	920	2200
ɜː	560	1480	2520

/u/
On C4



F0=H1
(523.2Hz)

/u/
On C4

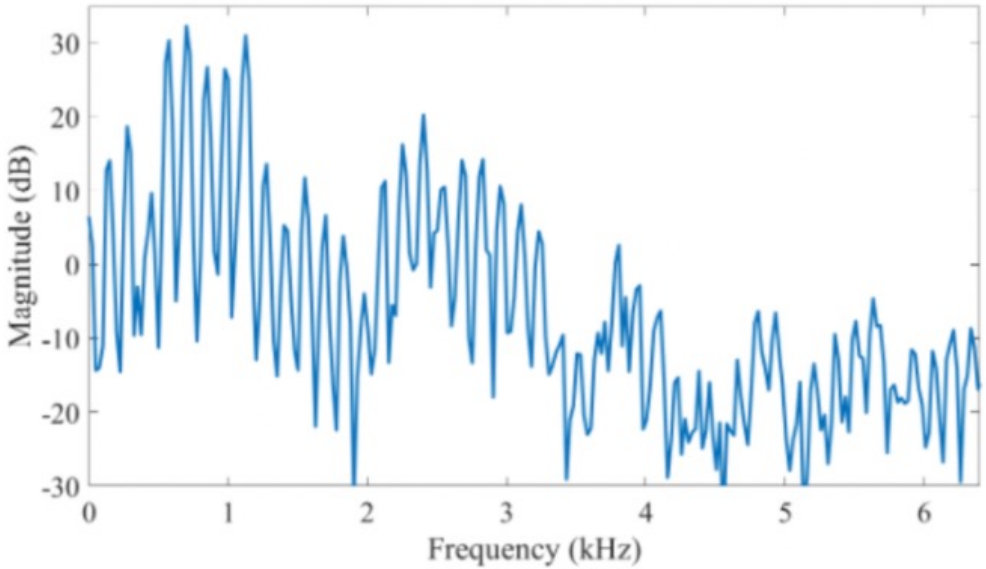


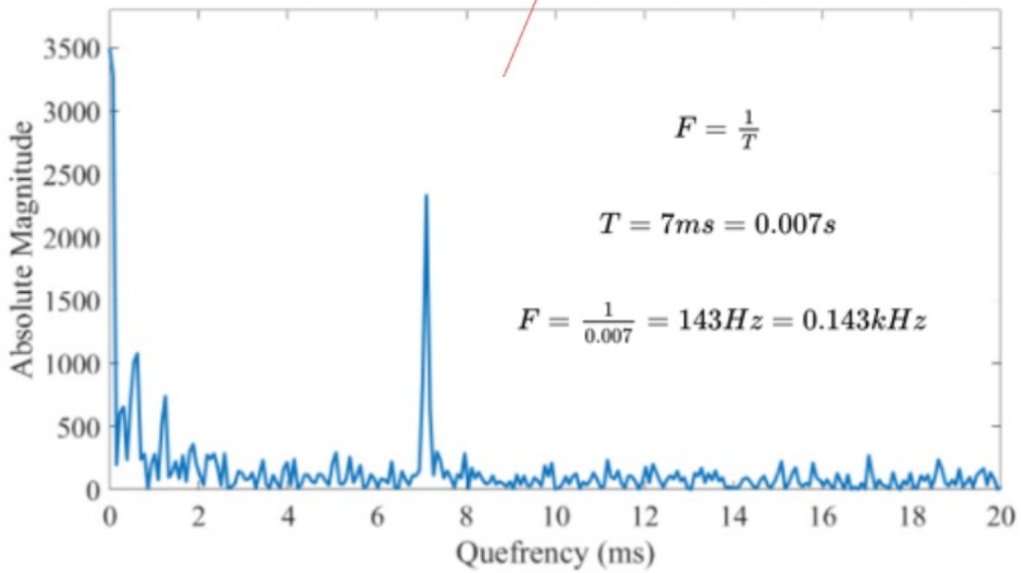
Cepstrum

Cepstrum을 통해 F0 구하기

IFFT

F0 = 0.143kHz





FFT 거친 spectrum

IFFT 적용

그래프에서 삐죽 튀어나온 부분이 F0

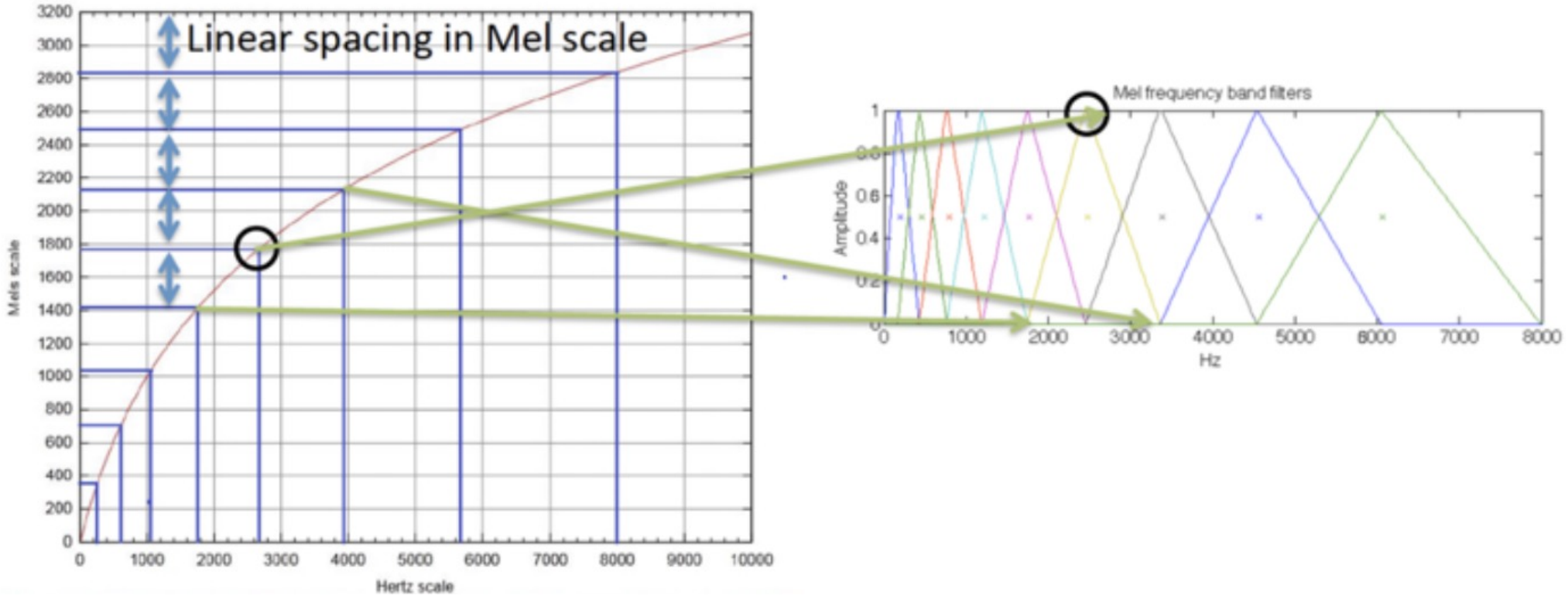
Cepstrum

Various Ways to Find a Cepstrum

- (1) Fourier transform -> Complex log -> Fourier transform
- (2) Fourier transform -> Complex log -> Inverse Fourier transform
- (3) Fourier transform -> power spectrum -> Mel-filter bank
  -> Real log -> Discrete cosine transform(DCT)

Mel Filter Bank

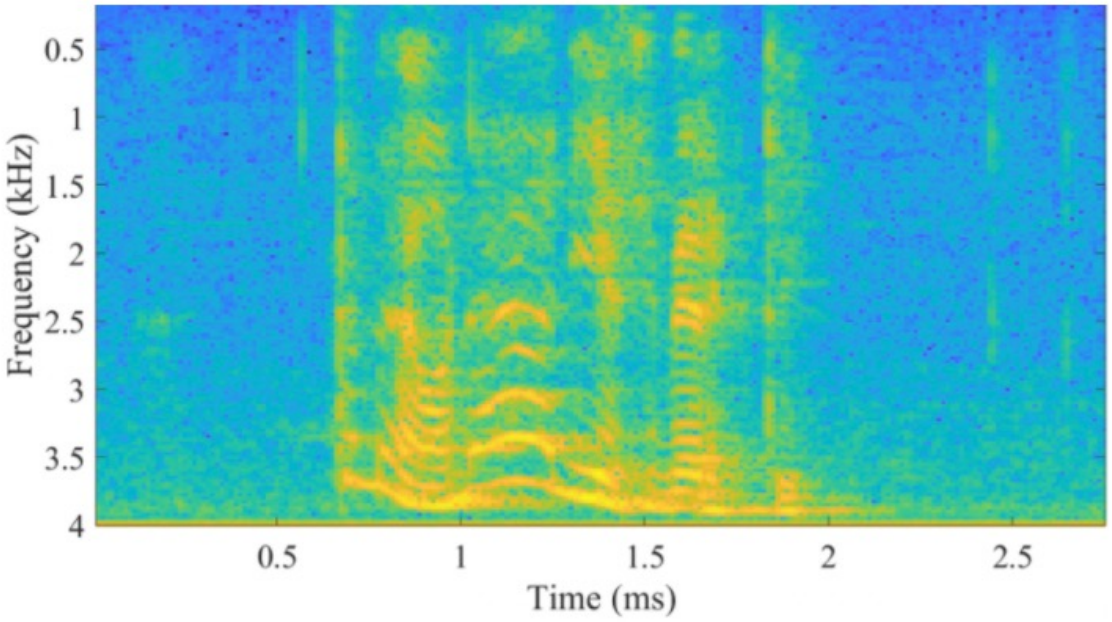
사람의 청력은 1000Hz 이상의 frequency에 대해서는 덜 민감하므로
1000Hz까지는 Linear하게 그 이상은 Log scale로 변환해줘야 함



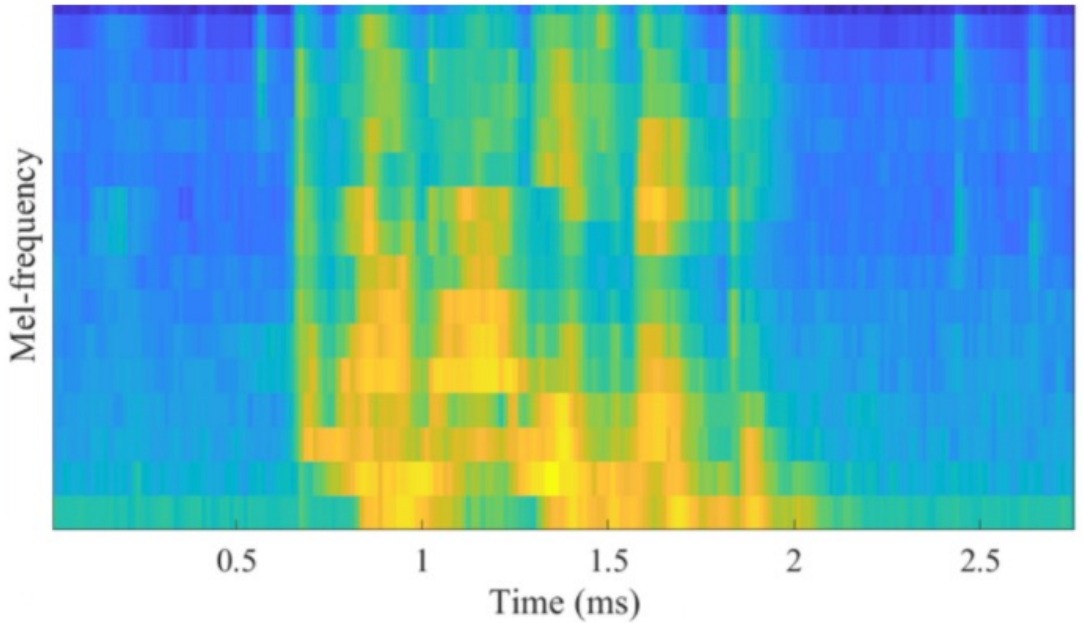
Mel Spectogram

주파수 단위를 다음 공식에 따라 멜 단위로 바꾼 것

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$



spectrogram

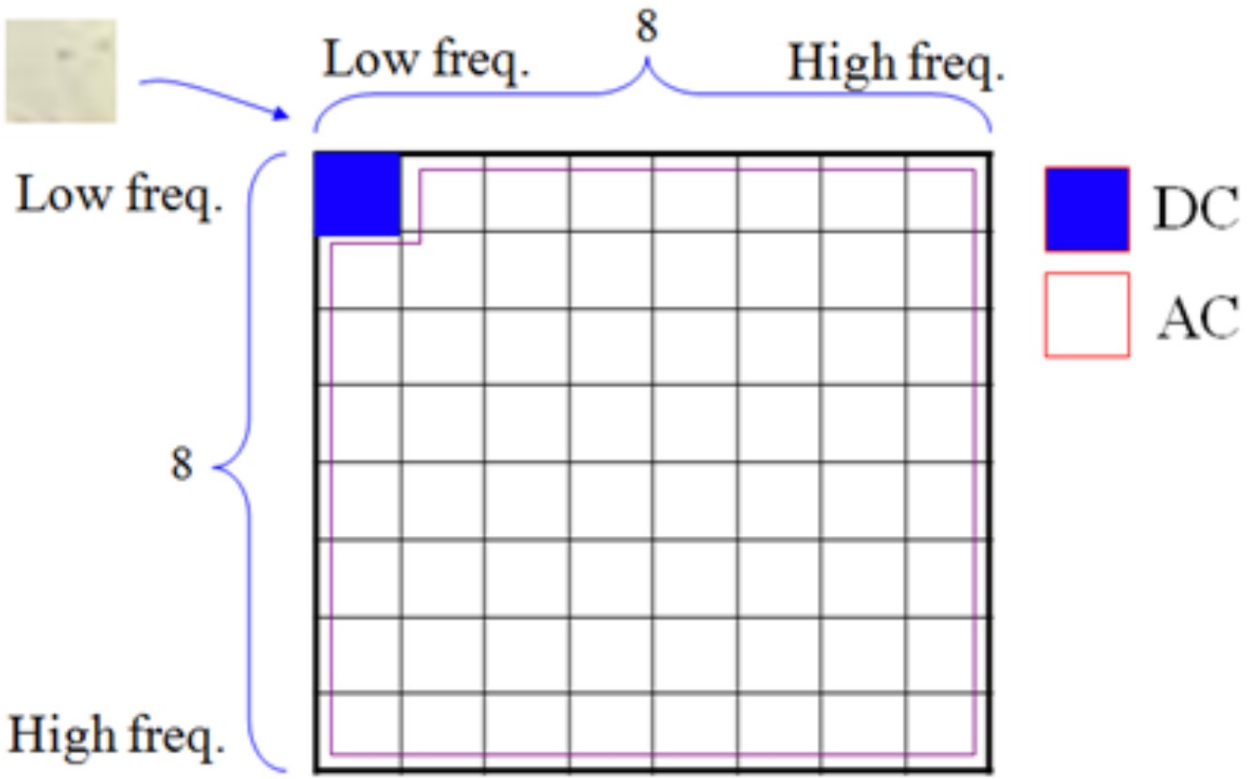


Mel-spectrogram

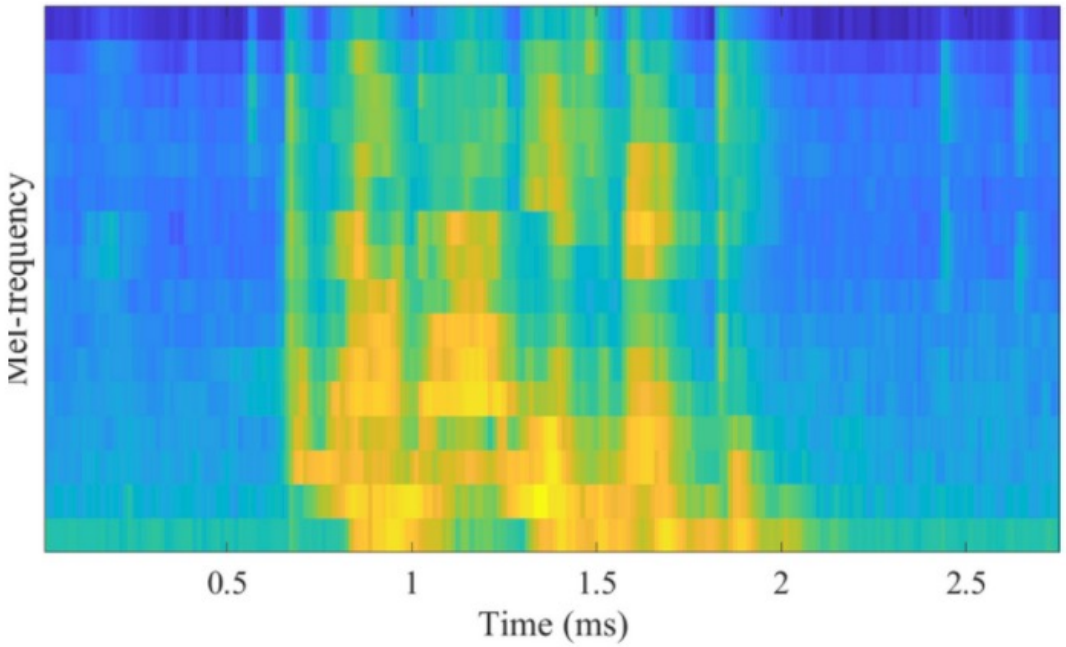
Discrete Cosine Transform(DCT)

Cepstrm coefficient

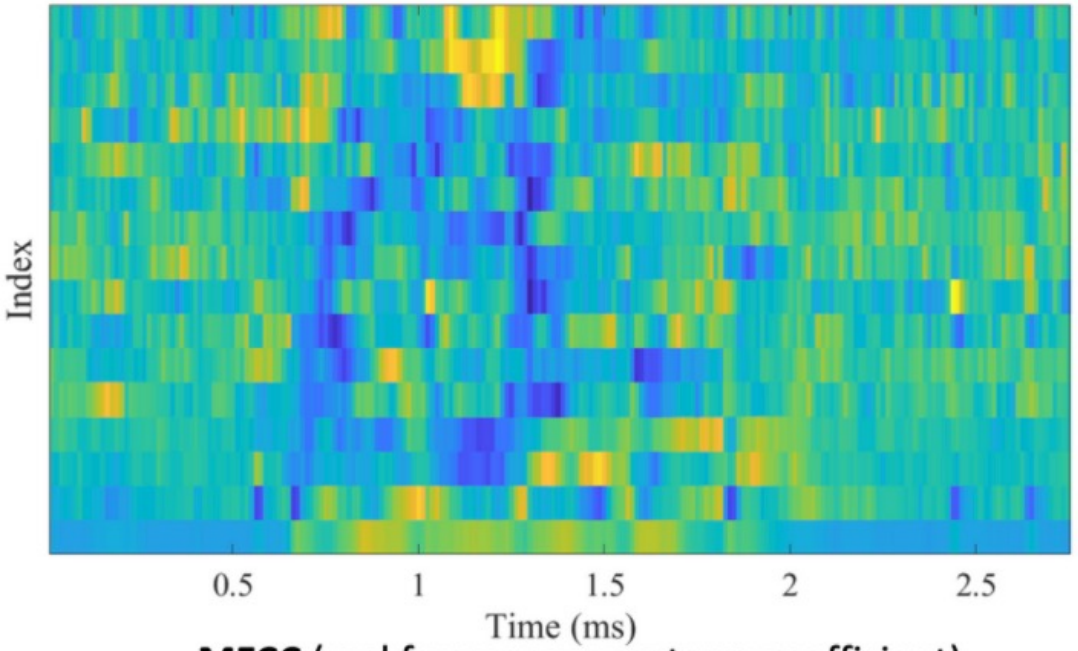
$$C_n = \sqrt{\frac{2.0}{\mathrm{fN}}} \times \sum_{j=1}^{\mathrm{fN}} \log_{10} \left(\sum_f W_j(f) \, |X(f)|^2 \right) \cos \left(\frac{\pi}{\mathrm{fN}} n \, (j - 0.5) \right)$$



MFCC



Mel-Spectrogram



MFCC (mel frequency cepstrum coefficient)

Q&A