# MFCC

## Winter Vacation Capstone Study

TEAM Kai.Lib

발표자 : 원철황

2020.01.13 (MON)

# MFCC

## MFCC (Mel-Frequency Cepstral Coeffiicient)

음성/음악 등 Audio Signal Processing 분야에서 널리 쓰이는 **특징값(Feature)** 중 하나. 소리의 고유한 특징을 나타내는 수치이다. 주로 음성인식, 화자 인식, 음악 장르 분류 등 오디오 도메인 문제를 해결.

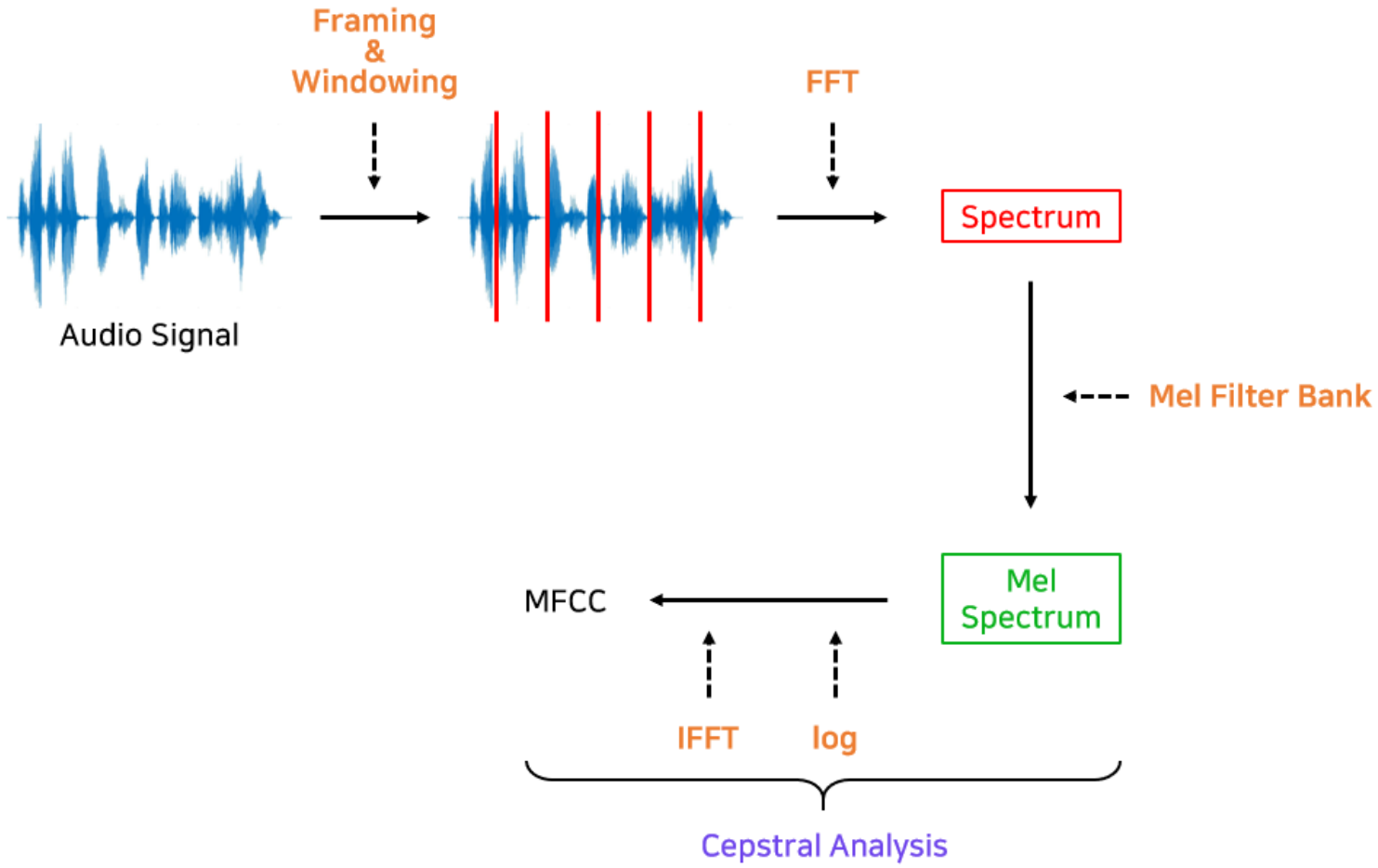
## MFCC의 기술적인 이해

아래는 MFCC를 구하는 과정을 설명한 글이다. 기술적으로 MFCC(Mel-Frequency Cepstral Coefficient)는 **Mel Spectrum**(멜 스펙트럼)에서 **Cepstral**(켑스트럴) 분석을 통해 추출된 값이라고 한다. 따라서 MFCC를 기술적으로 이해하려면 선행적으로 다음 개념들을 알아야 한다.

1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
3. Take the logs of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

- **Spectrum (스펙트럼)**
- **Cepstrum (켑스트럼)**
- **Mel Spectrum (멜 스펙트럼)**

# MFCC 추출 과정




## 각 순서에 대한 설명

1. 오디오 신호를 프레임별로 나누어 **FFT**를 적용해 **Spectrum**을 구한다.
2. **Spectrum**에 **Mel Filter Bank**를 적용해 **Mel Spectrum**을 구한다.
3. **Mel Spectrum**에 **Cepstral 분석**을 적용해 **MFCC**를 구한다.

# MFCC 추출 과정

오디오 신호는 **시간**(가로축)에 따른 **음압**(세로축)의 표현이다. 즉, **Time – Amplitude** 표현이다. 여기에 **FFT**를 수행하면 **주파수**(가로축)에 따른 **값**(세로축)의 표현이 가능해진다. 이것이 **Spectrum**이다.

※ **FFT**(Fast Fourier Transform : 고속 푸리에 변환)
신호를 주파수 성분으로 변환하는 알고리즘으로, 기존의 이산 푸리에 변환(DFT)을 더욱 빠르게 수행할 수 있도록 최적화한 알고리즘.

**Spectrum**을 사용하면 **각 주파수의 대역별 세기**를 알 수 있다. 이는 다시 해석하면 해당 음성신호에 주파수 성분이 얼마나 존재하는 지에 관한 정도를 확인할 수 있다. 즉, 어떤 주파수가 강하고 약한지를 알 수 있다.

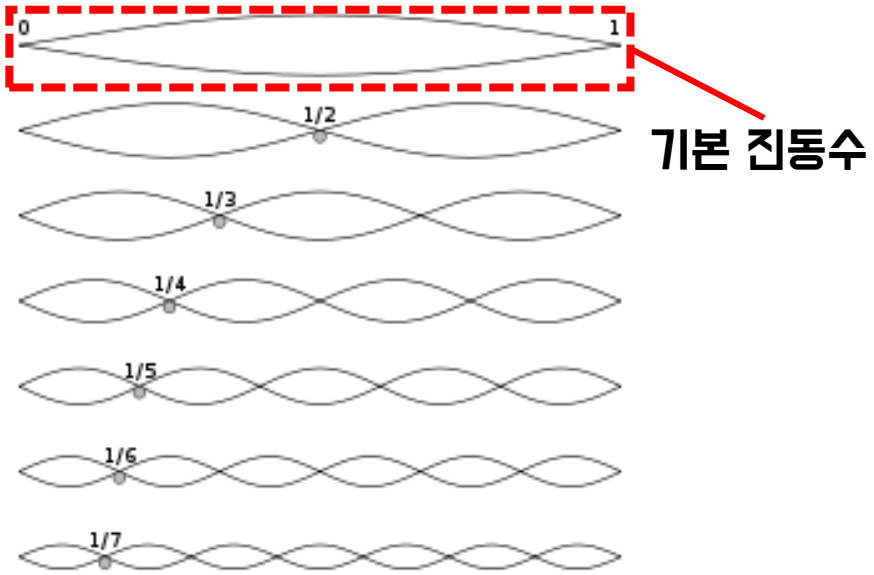
이 주파수 정보를 가진 **Spectrum**에서 소리의 고유한 특징을 추출할 수 있다. 그리고 그 정보를 추출 할 때 사용하는 방법이 **Cepstral Analysis**이다.

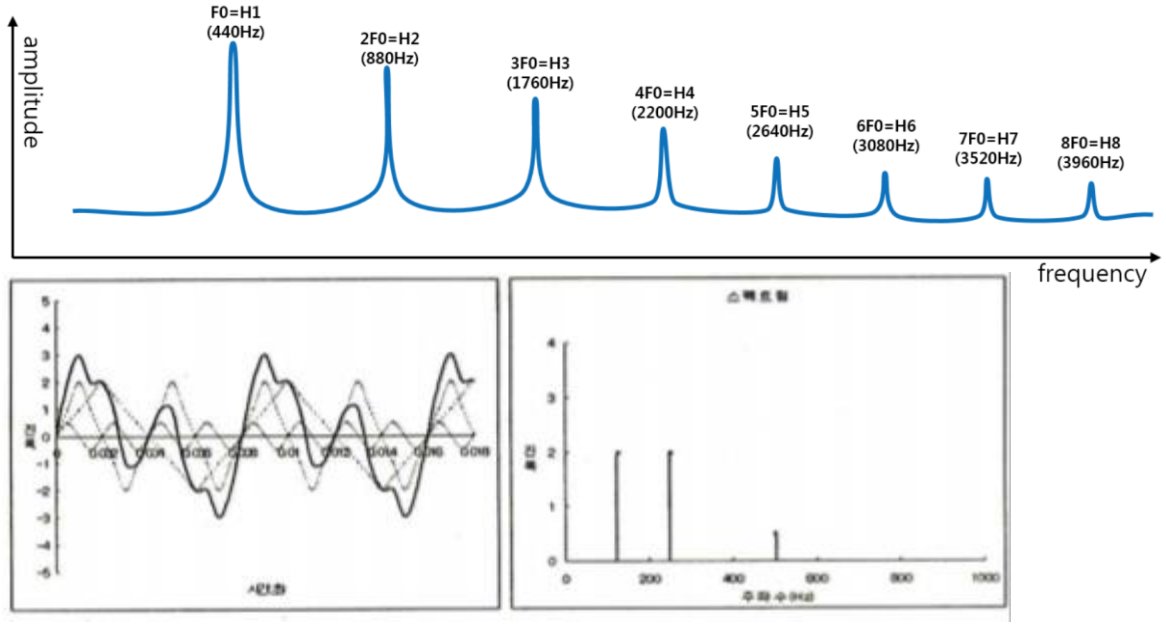
다만 MFCC는 일반적인 **Spectrum**이 아니라 특수한 필터(**Mel Filter Bank**)를 통과시킨 **Mel Spectrum**에 **Cepstral** 분석을 적용해 추출한다.

# Spectrum

## Spectrum이 갖고 있는 정보

악기 소리나 사람의 음성은 일반적으로 Harmonics(배음) 구조를 가짐.
**배음**은 하나의 음을 구성하는 부분음들 중 기본음보다 높은 정수배의 진동수(Frequency)를 갖는 모든 음들을 가리키는 말.



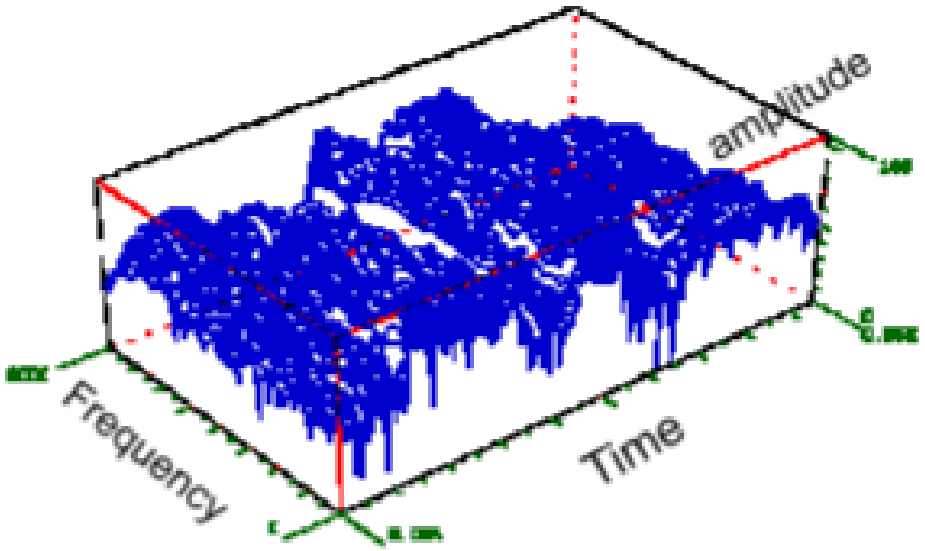

자료에서 아래로 내려갈 수록 진동수는 **2배, 3배, 4배.. N배** 로 커지는 것을 확인할 수 있다.
배음 구조는 악기나 성대의 구조에 따라 달라지며 **배음 구조의 차이**가 **음색의 차이**를 만든다.
즉, **Spectrum**에서 **배음 구조**를 유추해낼 수 있다면 **소리의 고유한 특징**을 찾아낼 수 있는 것이다.
이것을 가능케 하는 것이 **Cepstral Analysis**이다.

# Formant (포먼트)

## ※ Spectrogram (스펙트로그램)

소리나 파동의 시각화로 특징을 파악하기 위한 도구로, 파형(Waveform)과 스펙트럼(Spectrum)의 특징이 조합되어 있다.

- 파형에서는 시간축의 변화에 따른 진폭 축의 변화를 볼 수 있다.
- 스펙트럼에서는 주파수 축의 변화에 따른 진폭 축의 변화를 볼 수 있다.
- 스펙트로그램에서는 시간축과 주파수 축의 변화에 따라 진폭의 차이를 인쇄 농도/표시 색상의 차이로 나타낸다.




## 포먼트 정리

**Formant**는 **Spectrogram**에서 음향 에너지가 집중된 주파수 대역이 까만 띠 모양으로 나타나는 것을 뜻한다. 포먼트가 있는 주파수 대역은 **음향 에너지**가 **비교적 높은 강도**를 가지고 있다는 것을 의미한다. 모음은 제 1포먼트, 제 2포먼트, 제 3포먼트까지 나타나며 발성 형태에 따라 제 4포먼트, 제 5포먼트까지 나타날 수 있다.

# Formant (포먼트)

## 남성의 포먼트 예시

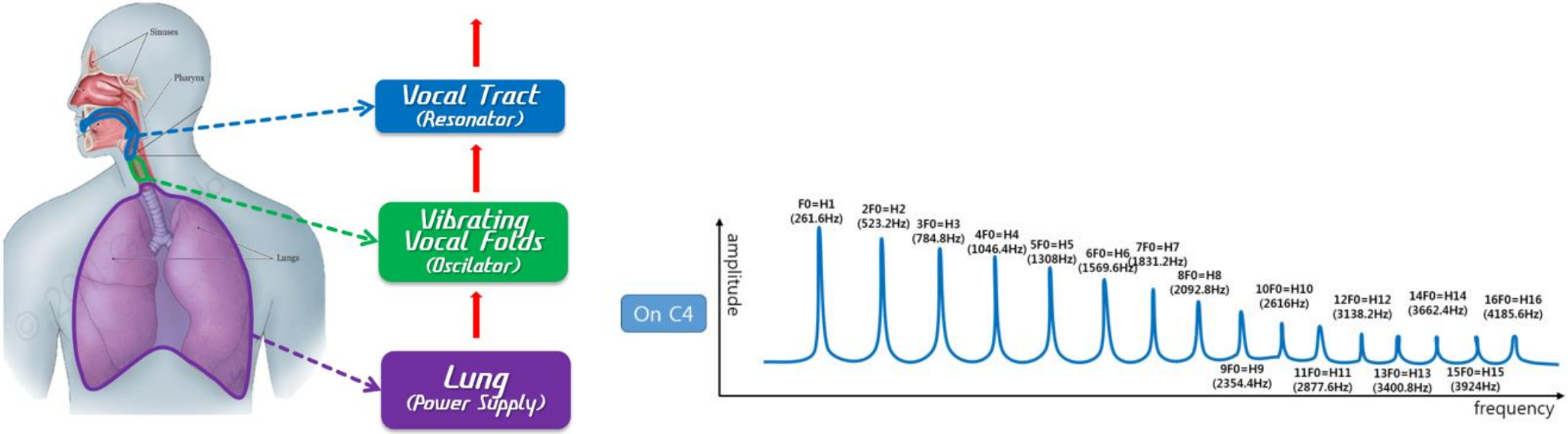
| Vowel | F1(Hz) | F2(Hz) | F3(Hz) |
|---|---|---|---|
| iː | 280 | 2620 | 3380 |
| ɪ | 360 | 2220 | 2960 |
| e | 600 | 2060 | 2840 |
| æ | 800 | 1760 | 2500 |
| ʌ | 760 | 1320 | 2500 |
| ɑː | 740 | 1180 | 2640 |
| ɒ | 560 | 920 | 2560 |
| ɔː | 480 | 760 | 2620 |
| ʊ | 380 | 940 | 2300 |
| uː | 320 |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

Adult male formant frequencies in Hertz collected by J.C.Wells around 1960.
Note how F1 and F2 vary more than F3.

# Formant (포먼트)

## 남성의 포먼트 예시

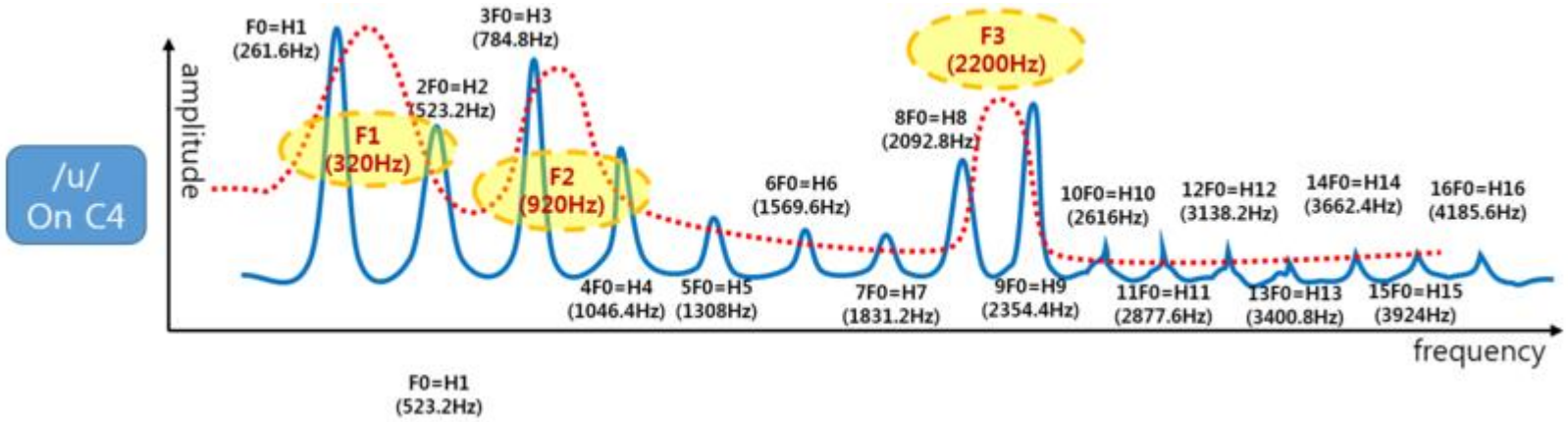
다음과 같은 원음이  허파의 공기와 성대의 진동을 통해 발생했다고 가정하자.



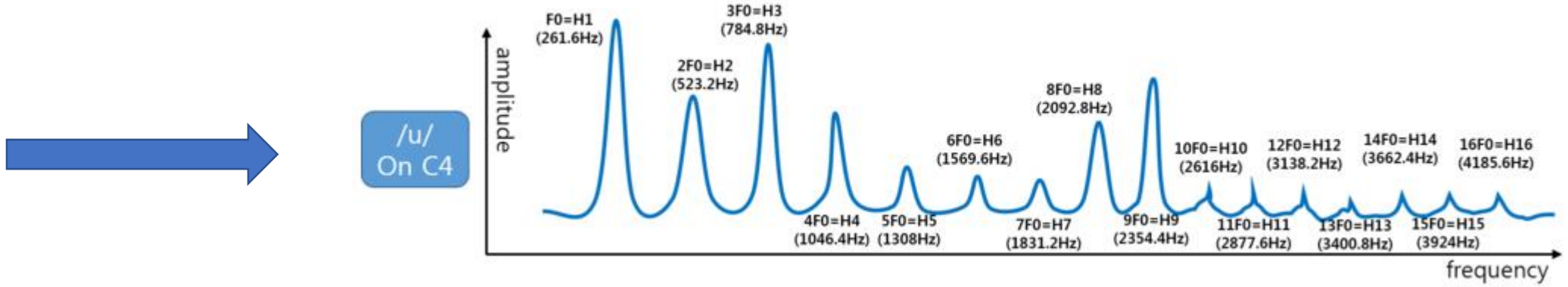
# Formant (포먼트)

## 남성의 포먼트 예시

### /u/ 모음으로 발음시

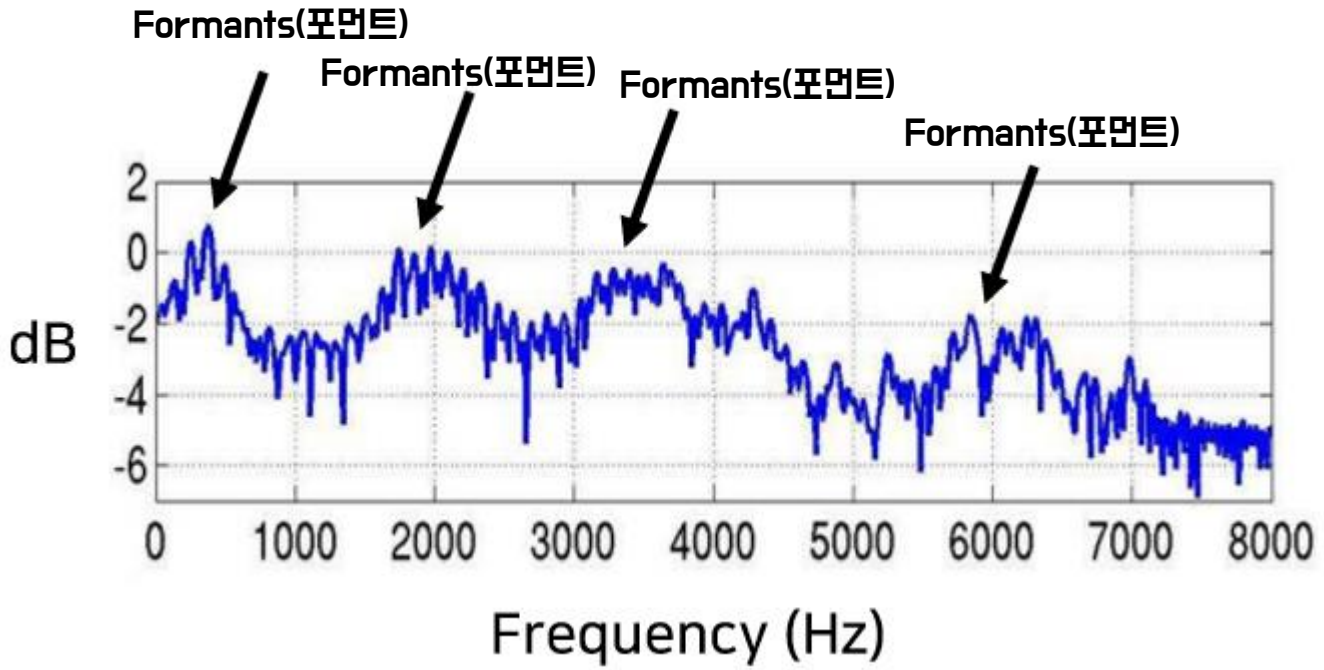


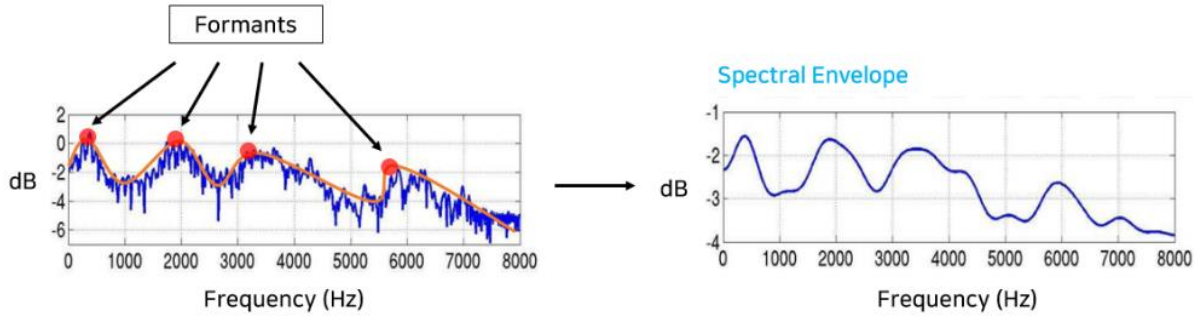

F0=H1
(523.2Hz)




**Formant (포먼트, 음형대)**을 지나며 포먼트를 제외한 다른 주파수 대역은 감쇠가 일어나고 F1 ~ F3에 해당하는 배음들이 부스팅 된다.

# Formant (포먼트)




## ※ Formant (포먼트, 음형대)

: 소리가 공명되는 특정 주파수 대역

- 사람의 음성은 성대에서 형성되어 성도를 거치며 변형된다.
- 성대가 진동하고 이 소리는 목을 통해 전달되어 입 밖으로 나온다.,
- 소리는 **성도**를 지나며 **포먼트**를 만나 **증폭**되거나 **감쇠**된다.
- Formant는 Harmonics와 만나 소리를 풍성하게, 선명하게 만드는 **필터** 역할을 한다.

좌측처럼 포먼트들을 연결한 곡선(**Spectral Envelope**)과 **Spectrum**을 분리하는 일을 해야 하며, **MFCC**는 둘을 분리하는 과정에서 도출된다. 이때 사용하는 개념과 수학적 알고리즘이 **log**와 **Inverse FFT**이다.

# Mel Spectrum

## Mel Spectrum vs Spectrum

Mel-Frequency Cepstral Coefficient 는 Spectrum이 아닌 Mel Spectrum에서 Cepstral Anlaysis로 추출한다. Mel Spectrum은 다음과 같은 과정을 거쳐 만들어진다.

사람의 청각기관은 고주파수 대역보다 저주파수 대역에 더욱 민감하다. 이런 특성을 반영해 물리적인 주파수와 실제 사람이 인식하는 주파수의 관계를 Mel Scale(멜 스케일)이라 한다.

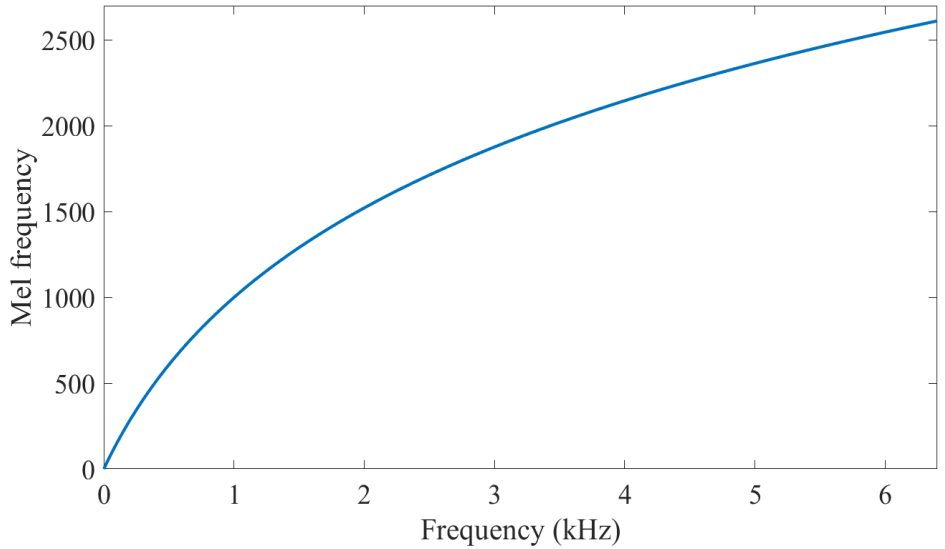
$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

이렇게 만들어진 공식 외에 다음과 같은 형태로 나타나기도 한다.

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) = 1127 \ln\left(1 + \frac{f}{700}\right)$$

그리고 역 관계는 아래와 같이 나타난다.

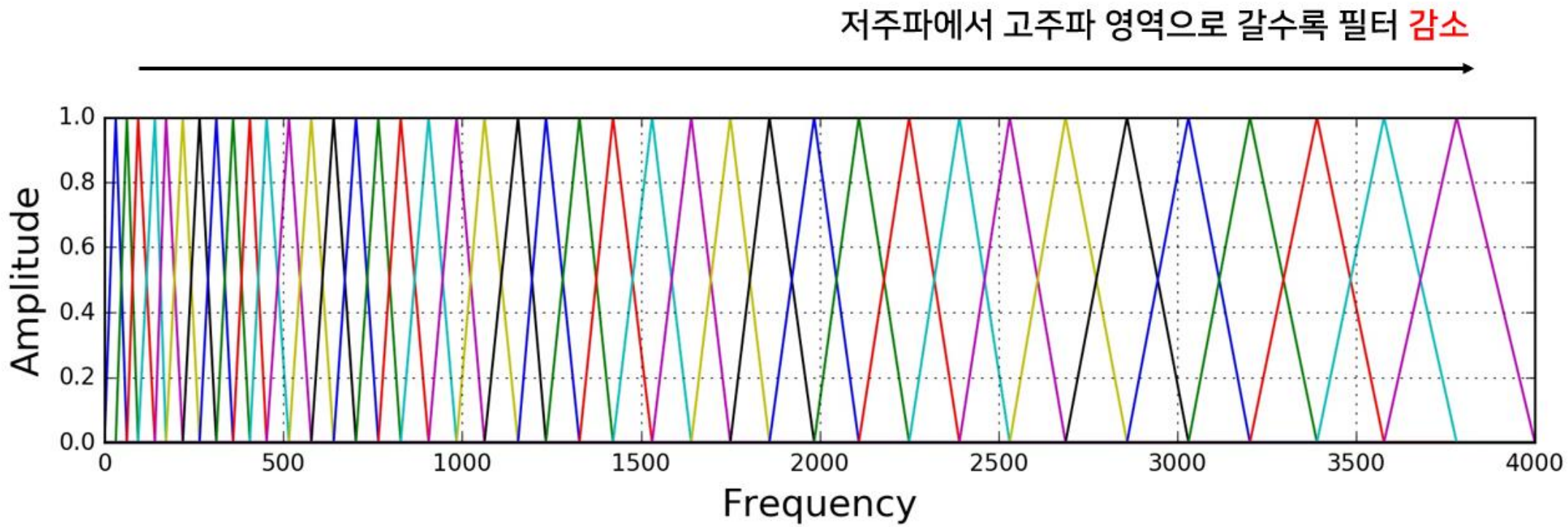
$$f = 700\left(10^{\frac{m}{2595}} - 1\right) = 700\left(e^{\frac{m}{1127}} - 1\right)$$




이 **Mel Scale**에 기반한 **Filter Bank**를 **Spectrum**에 적용하여 도출해낸 것이 **Mel Spectrum**이다. Mel Scale은 즉, Filter Bank를 나눌 때 어떤 간격으로 나눠야 하는지 알려주는 역할을 한다.

# Mel Scale

## Filter Banks on the Mel-Scale




이 **Mel Scale**에 기반한 **Filter Bank**를 **Spectrum**에 적용한 예시이다. Filter Bank의 분포는 사용자마다 상이할 수 있으나 기본적으로 귀의 청각적 특성(달팽이관에서의 주파수 특성)을 고려하여 결정한다. 삼각형 모양의 필터를 사용해 보통 1kHz까지는 선형적으로 Filter Bank를 위치시키고 그 이상에서는 Mel Scale로 분포하는 뱅크로 구성한다.

# Mel Scale

| Index | Bark scale | | Mel scale | |
|---|---|---|---|---|
| | Center freq (Hz) | BW (Hz) | Center freq (Hz) | BW (Hz) |
| 1 | 50 | 100 | 100 | 100 |
| 2 | 150 | 100 | 200 | 100 |
| 3 | 250 | 100 | 300 | 100 |
| 4 | 350 | 100 | 400 | 100 |
| 5 | 450 | 110 | 500 | 100 |
| 6 | 570 | 120 | 600 | 100 |
| 7 | 700 | 140 | 700 | 100 |
| 8 | 840 | 150 | 800 | 100 |
| 9 | 1000 | 160 | 900 | 100 |
| 10 | 1170 | 190 | 1000 | 124 |
| 11 | 1370 | 210 | 1149 | 160 |
| 12 | 1600 | 240 | 1320 | 184 |
| 13 | 1850 | 280 | 1516 | 211 |
| 14 | 2150 | 320 | 1741 | 242 |
| 15 | 2500 | 380 | 2000 | 278 |
| 16 | 2900 | 450 | 2297 | 320 |
| 17 | 3400 | 550 | 2639 | 367 |
| 18 | 4000 | 700 | 3031 | 422 |
| 19 | 4800 | 900 | 3482 | 484 |
| 20 | 5800 | 1100 | 4000 | 556 |
| 21 | 7000 | 1300 | 4595 | 639 |
| 22 | 8500 | 1800 | 5278 | 734 |
| 23 | 10500 | 2500 | 6038 | 843 |
| 24 | 13500 | 3500 | 6954 | 969 |

가장 흔히 사용되는 필터뱅크 분석 주파수 및 대역폭

■ Filter Bank Analysis

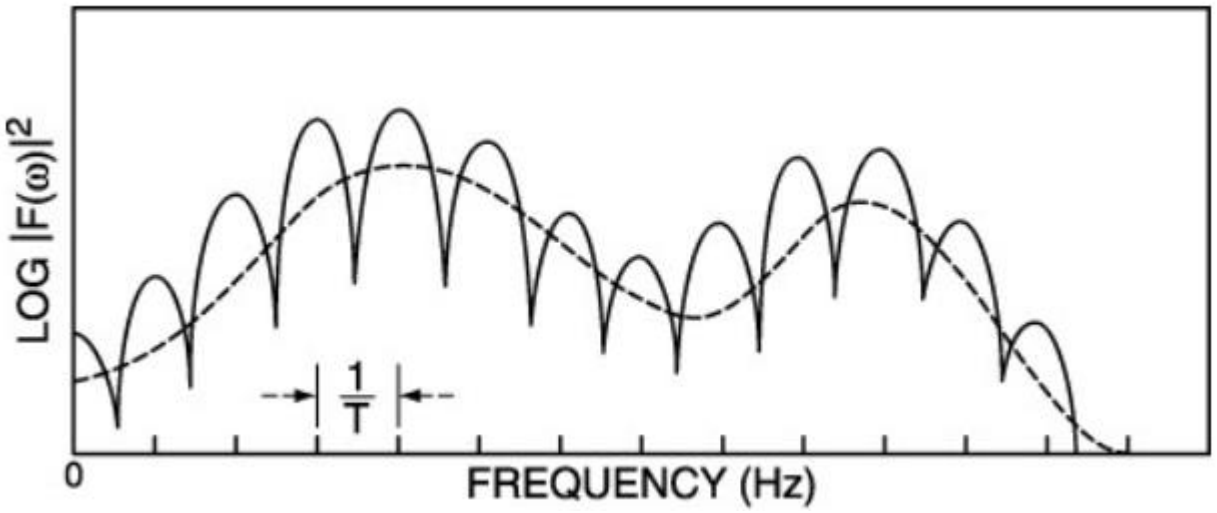
- 필터뱅크의 중심 주파수는 Bark 또는 Mel 단위로 위치하고 대역폭은 Critical Bandwidth에 따라 결정된다.
- Bark = 13 atan(0.76f/1000)+3.5 atan(f^2/7500^2)
- MelFreq = 2595 log10(1+f/700)
- CriticalBW = 25+75[1 + 1.4 (f/1000)^2]^0.69,  f>1000
- CriticalBW = 100, f<1000
- 가중치 모양은 Triangular, Rectangular, Trapezoidal, Gaussian 등이 있는데, Triangular Shape이 Spectrum을 Smoothing 하는 효과를 갖는다.
- 이 가중치를 통계적으로 최적화하는 분야도 하나의 연구 주제로 각광받고 있다.

# Cepstrum

## Cepstrum vs Spectrum

**둘의 관계는 역으로 정의할 수 있다**

### Spectrum 의 표현

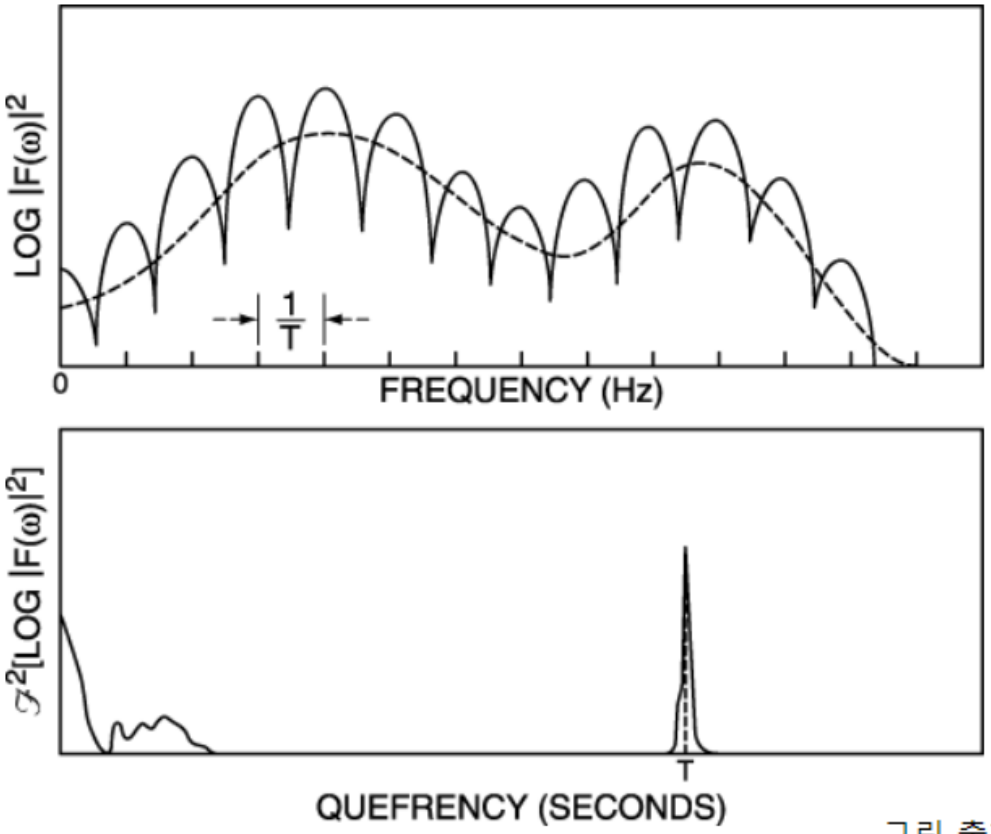



그림에서는 2가지 유형의 마루를 확인 가능

- 좁은 마루는 성대 진동에 의한 기본 주파수의 정수배이다.
- 넓은 마루는 Formant에 의해 공명한 뒤 나타나는 마루이다.
- 우리가 관측하는 결과는 두 번째 큰 마루이다.
- 증폭과 감쇠가 된 최종 스펙트럼으로부터 Harmonics를 추출하는 방법을 Cepstral 분석이라 한다.
- Harmonics는 기본 주파수의 정수배로 나타나기 때문에 Pitch를 잘 **추정**할 수 있게 된다.

# Cepstrum

**이전에 구한 스펙트럼 신호의** Log **값에** Inverse Fourie Transfom**을 하면 캡스트럼이 된다**.




■ **Cepstrum**에서 용어

Frequency : **Que**frency (큐프렌시)

Harmonics : **Rah**monics (라모닉스)

Filter : **Lif**ter (리프터)

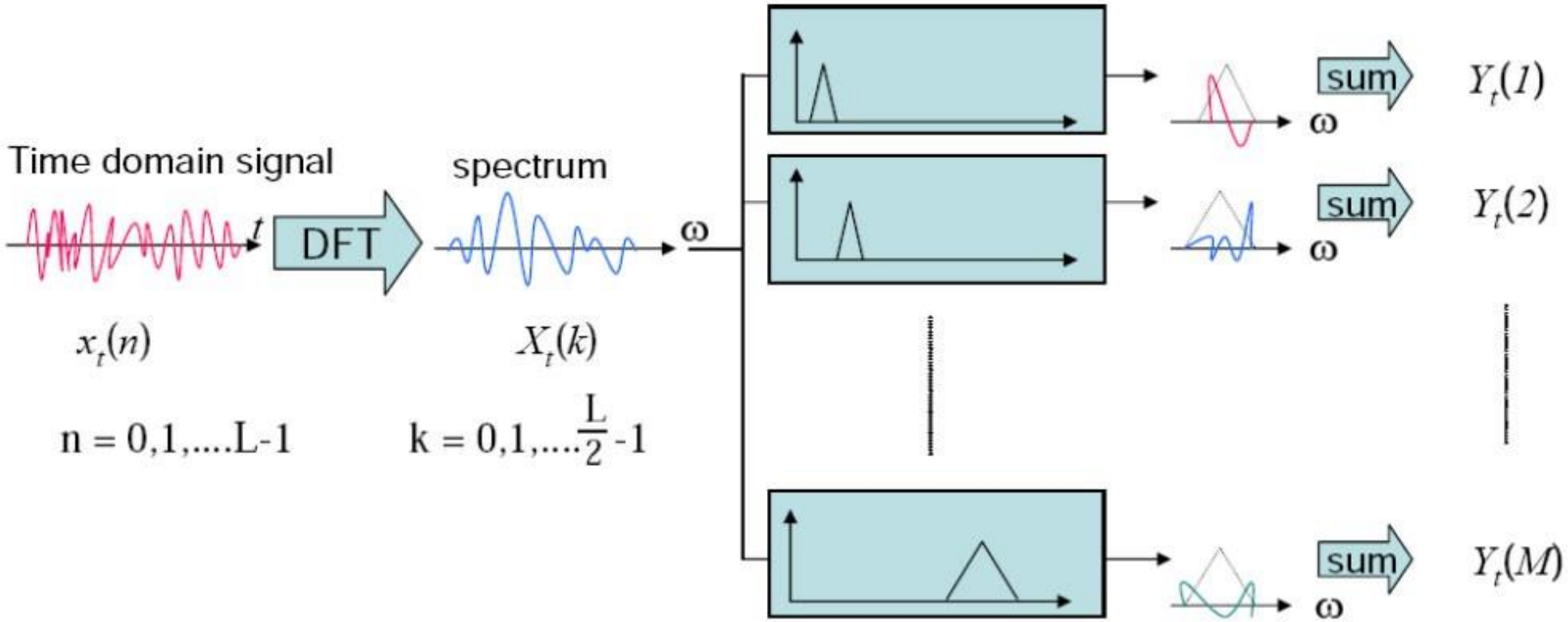
Magnitude : **Gam**nitude (감니튜드)

Phase : **Shap**e (셰이프)

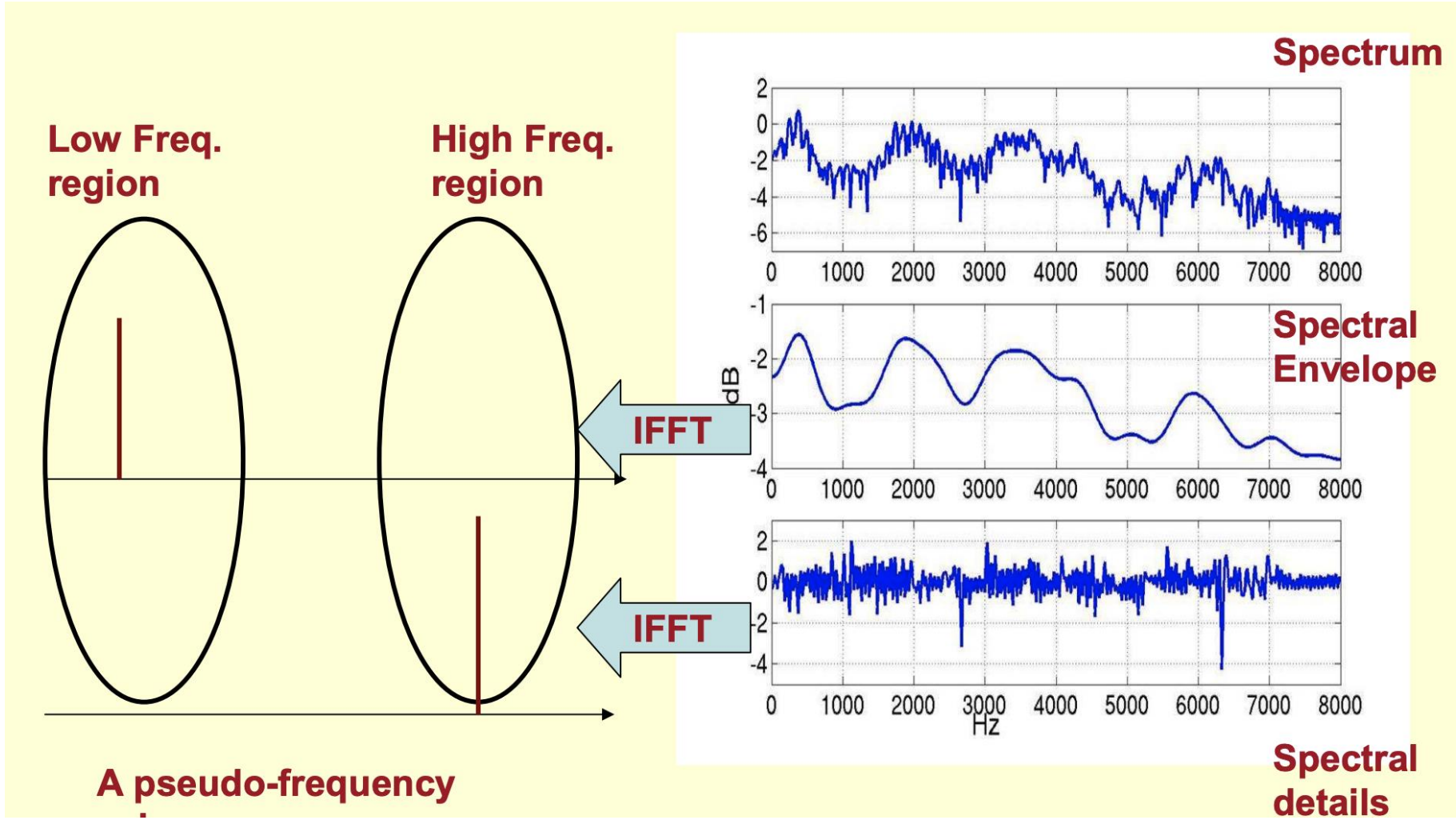
그림 출처

A. M. Noll, "Cepstrum Pitch Determination," J. Acoustic Society of America, vol. 41, 1967, p. 293.

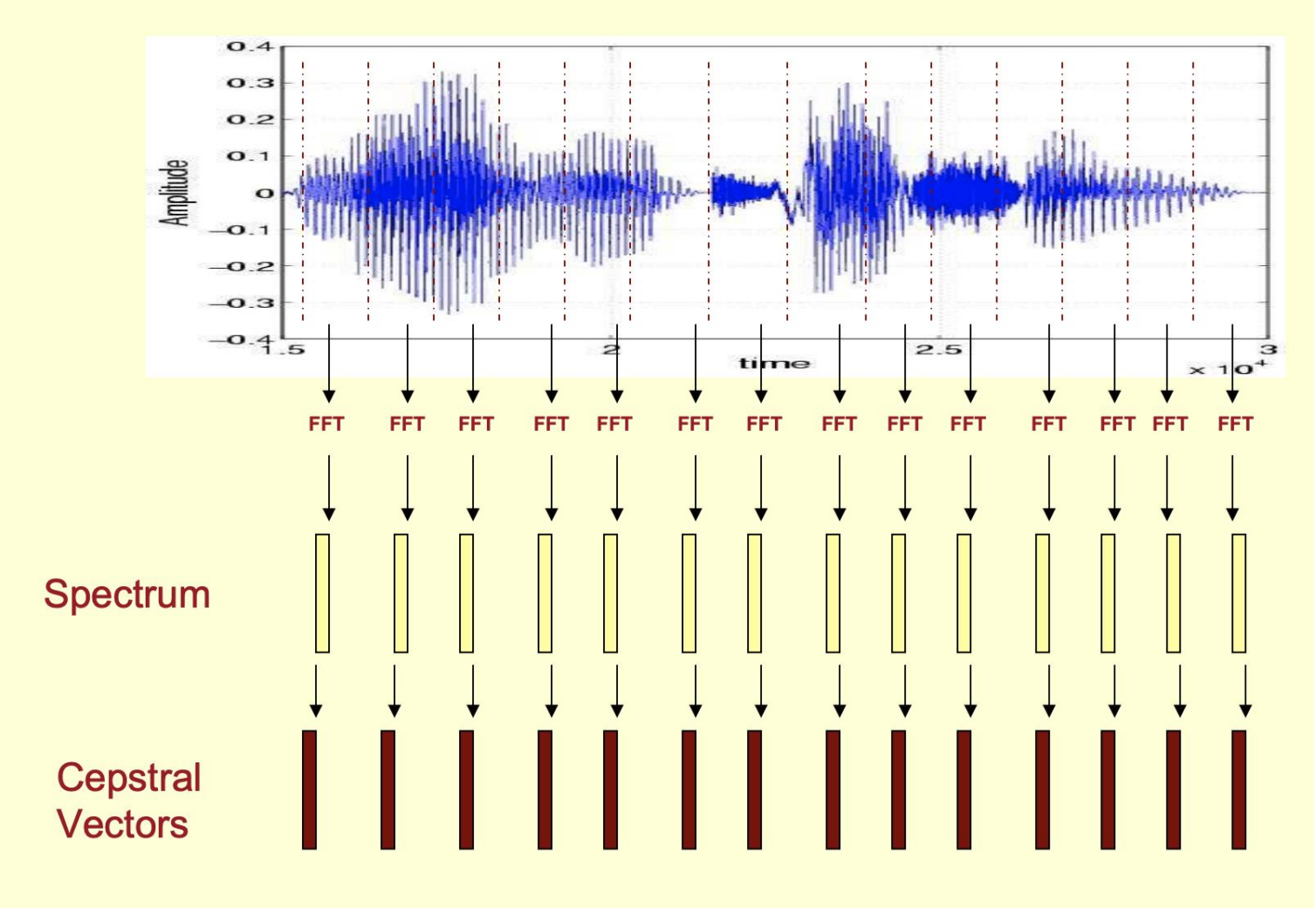
# MFCC 정리


Time domain signal
t
DFT
spectrum
ω
$x_t(n)$
$X_t(k)$
n = 0,1,....L-1
k = 0,1,....$\frac{L}{2}$-1
sum
$Y_t(1)$
sum
$Y_t(2)$
sum
$Y_t(M)$

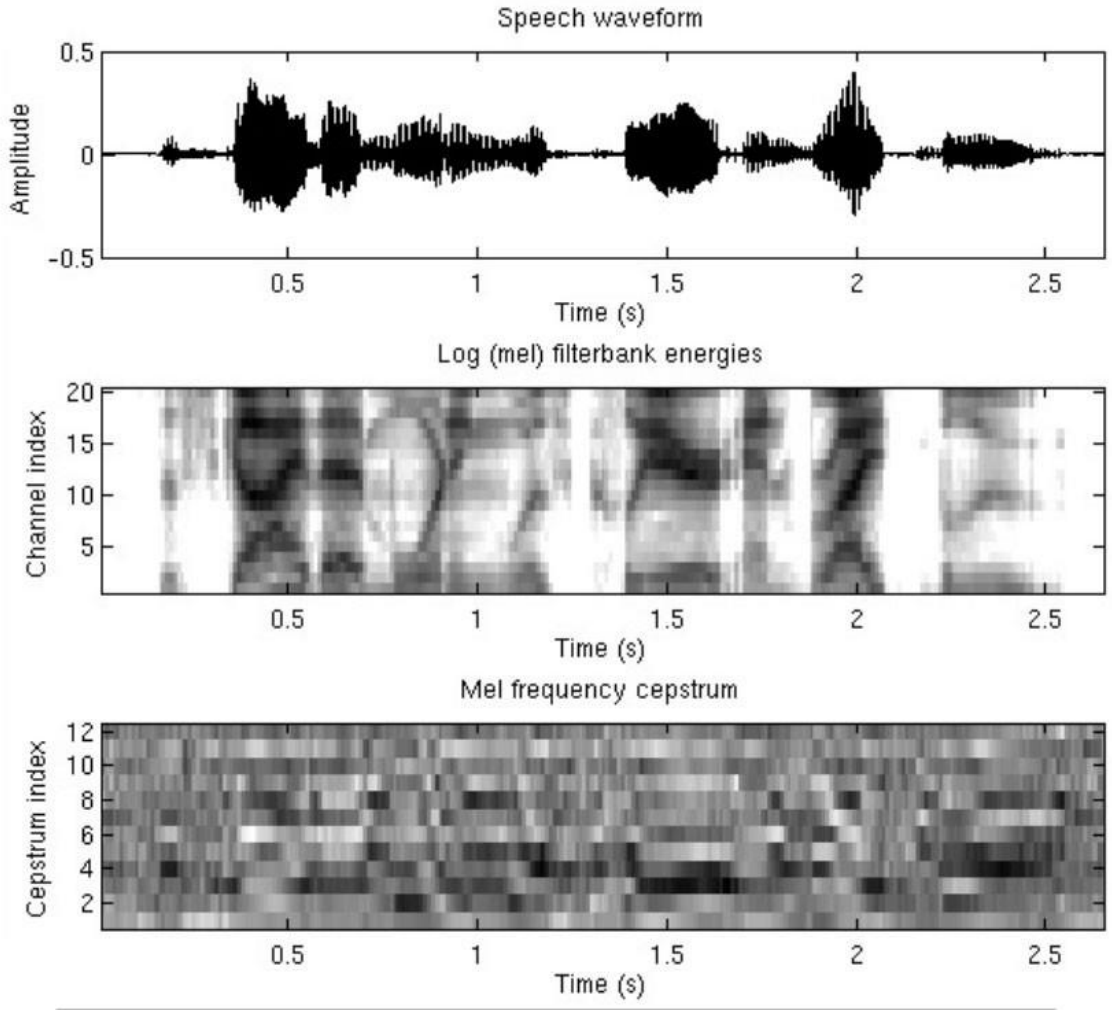
# MFCC 정리


Low Freq. region
High Freq. region
IFFT
IFFT
A pseudo-frequency
Spectrum
Spectral Envelope
Spectral details
dB
Hz

# MFCC 정리



# MFCC 정리


Speech waveform
Amplitude
0.5
0
-0.5
0.5
1
1.5
2
2.5
Time (s)
Log (mel) filterbank energies
Channel index
20
15
10
5
0.5
1
1.5
2
2.5
Time (s)
Mel frequency cepstrum
Cepstrum index
12
10
8
6
4
2
0.5
1
1.5
2
2.5
Time (s)

# 신호처리

## Winter Vacation Capstone Study

TEAM Kai.Lib

발표자 : 원철황

2020.01.13 (MON)

# Fourier Series

아래의 기함수에 대한 푸리에 급수. 간단하게 **푸리에 사인 급수(Fourier sine series)**라고도 한다.

$$f(t+T) = f(t), f(-t) = -f(t)$$

$$\Rightarrow \quad f(t) = \sum_{m=1}^{\infty} F_m \sin(m\omega_0 t) \text{ when } \omega_0 = \frac{2\pi}{T}$$

푸리에 급수의 특징은 임의의 주기 함수를 삼각 함수의 무한 급수로 표현할 수 있다는 점이다.

$$\int_{-T/2}^{T/2} (\cdot) \sin(l\omega_0 t)\ dt$$

$$\Rightarrow \quad \int_{-T/2}^{T/2} \sin(m\omega_0 t) \sin(l\omega_0 t)\ dt$$

$$= \begin{cases} 0 & \text{if } m \neq l \\ \frac{T}{2} & \text{if } m = l \end{cases}$$

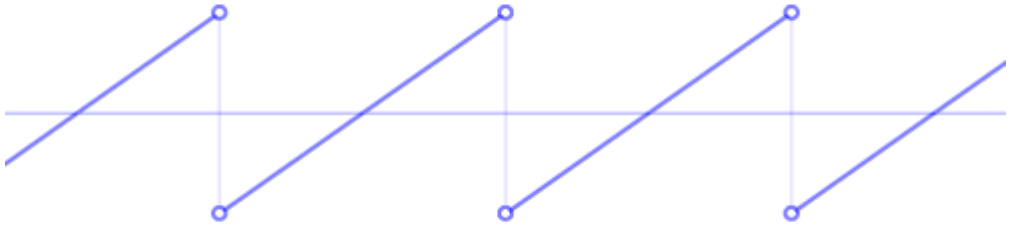
$$F_m = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin(m\omega_0 t)\ dt$$

$$\Rightarrow \quad = \frac{4}{T} \int_{0}^{T/2} f(t) \sin(m\omega_0 t)\ dt$$

- (.) 에 임의의 주기 함수를 한 주기 T에 대해 적분한 것이다.
- 기함수를 한 주기에 대해 적분하면 둘째식에 있는 Sine 함수의 직교성에 의해 아래 식이 나타난다.
- (.)에 우함수를 적용하면 관계없이 모두 0이 된다.
- 그 아래는 푸리에가 제안한 푸리에 계수를 결정하는 방법이다.

# Fourier Series

주기함수 f(t)를 그대로 사용하는 것이 쉽다고 생각할 수 있지만 우리는 아는 내용이 없다. 반면 Fourier가 제안한 식에서는 **Fourier Coefficient**만 구해지지 않았지 다른 형태는 알고 있다. 따라서 시간적으로 변하는 f(t)를 **푸리에 계수만으로 결정**할 수 있다면 **신호를 분석**할 수 있다.



N = 0

다음과 같은 **불연속적으로 존재하는 톱니 파**는 **무한한 정현파의 합**으로 구현할 수 있다. 그러나 중간의 불연속 부분은 함수 값을 특정할 수 없으나 푸리에 급수는 이를 0으로 표현하여 주기성을 지니게 하고 있다. 따라서 불연속 함수에 관해서는 푸리에 급수 식이 이를 만족하지 않는다. 따라서 **불연속 주기 함수에도 적용 가능하도록 보완할 필요성**이 있다.

# Fourier Series

푸리에 급수를 확장하여 기함수와는 다른 특징을 갖는 우함수에 대해 푸리에 급수를 정의한다. 우함수에 대한 푸리에 급수는 푸리에 코사인 급수라고 한다.

$$g(t+T) = g(t), g(-t) = g(t)$$

$$\Rightarrow \quad g(t) = \sum_{m=0}^{\infty} G_m \cos(m\omega_0 t) \text{ when } \omega_0 = \frac{2\pi}{T}$$

이를 적용하여 이전과 같이 전개하면 다음과 같은 특징을 얻는다.

$$\int_{-T/2}^{T/2} (\cdot) \cos(l\omega_0 t)\ dt$$

$$\Rightarrow \quad \int_{-T/2}^{T/2} \cos(m\omega_0 t) \cos(l\omega_0 t)\ dt$$

$$= \begin{cases} 0 & \text{if } m \neq l \\ T & \text{if } m = l = 0 \\ \frac{T}{2} & \text{otherwise} \end{cases}$$

$$G_0 = \frac{2}{T} \int_0^{T/2} g(t)\ dt$$

$$\Rightarrow \quad G_m = \frac{2}{T} \int_{-T/2}^{T/2} g(t) \cos(m\omega_0 t)\ dt$$

$$= \frac{4}{T} \int_0^{T/2} g(t) \cos(m\omega_0 t)\ dt \text{ when } m \neq 0$$

※ 오일러 공식

*Euler's Formula*

$$e^{i\phi} = \cos\phi + i\sin\phi$$

*Euler's identity*

$$e^{i\pi} + 1 = 0$$

# Fourier Series

다음처럼 모든 함수는 기함수와 우함수의 합으로 표현할 수 있다. 따라서 이전에 정의했던 두 식을 합치면 기함수와 우함수의 특성을 모두 지니는 함수를 정의할 수 있기 때문에 임의의 주기 함수에 대한 복소 푸리에 급수를 얻을 수 있다.

$$
\begin{aligned}
f(t) &= \underbrace{\frac{f(t)+f(-t)}{2}}_{f_{\text{even}}(t)} + \underbrace{\frac{f(t)-f(-t)}{2}}_{f_{\text{odd}}(t)} \\
&= f_{\text{even}}(t) + f_{\text{odd}}(t)
\end{aligned}
$$

**복소 푸리에 급수**는 기함수 또는 우함수 여부에 관계없이 **모든 주기 함수**에 적용할 수 있는 **일반적 기법**이다. 정리하면 임의의 주기 함수는 무한한 주기함수로 표현할 수 있다. 그리고 이 주기 함수는 우함수 또는 기함수로 표현될 수 있으며 그 가장 기본적인 꼴은 Sine과 Cosine이며 각각의 계수를 가진다. 그리고 그 합을 통해 임의의 주기함수를 표현할 수 있다. 그리고 sine 과 cosine은 오일러 공식으로 표현되어 나타날 수 있다.

$$
\begin{aligned}
h(t) &= \sum_{m=1}^{\infty} F_m \sin(m\omega_0 t) + \sum_{m=0}^{\infty} G_m \cos(m\omega_0 t) \\
&= \sum_{m=1}^{\infty} F_m \frac{e^{im\omega_0 t} - e^{-im\omega_0 t}}{2i} + \sum_{m=0}^{\infty} G_m \frac{e^{im\omega_0 t} + e^{-im\omega_0 t}}{2} \\
&= G_0 + \frac{1}{2}\sum_{m=1}^{\infty} \left[(G_m - iF_m)e^{im\omega_0 t} + (G_m + iF_m)e^{-im\omega_0 t}\right]
\end{aligned}
$$

# Fourier Transform

음성(speech), 음악(music) 등의 음향(sound) 데이터에서 특징(feature)을 추출하는 방법

모든 신호는 **주파수(frequency)**와 **크기(magnitude)**, **위상(phase)**이 다른 **정현파(sinusolida signal)**의 조합으로 나타낼 수 있다. **퓨리에 변환**은 **조합된 정현파의 합(하모니) 신호**에서 **그 신호를 구성하는 정현파들을 각각 분리해내는 방법**이다.

$$f(t) = \sum_{m=-\infty}^{\infty} F_m e^{im\omega_0 t}$$

$$\text{where } F_m = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-im\omega_0 t}\,dt,\ \omega_0 = \frac{2\pi}{T}$$

푸리에 급수는 유용하지만 함수가 반드시 주기적이어야 한다. 이점을 보완하기 위해 제안된 적분 변환이 바로 푸리에 변환이다. 즉, 주기 T를 무한대로 보내 모든 시간에 대해 독립된 한 Signal을 주기 함수로 간주하는 것이다. 한 주기를 무한한 시간으로 확장하는 것을 Idea로 한다.

$$F(\omega) = \lim_{T\to\infty} F_m T,\ \omega = \lim_{T\to\infty} m\omega_0$$

$$\Rightarrow \quad F(\omega) = \lim_{T\to\infty} F_m T = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt$$