FernUniversität in Hagen

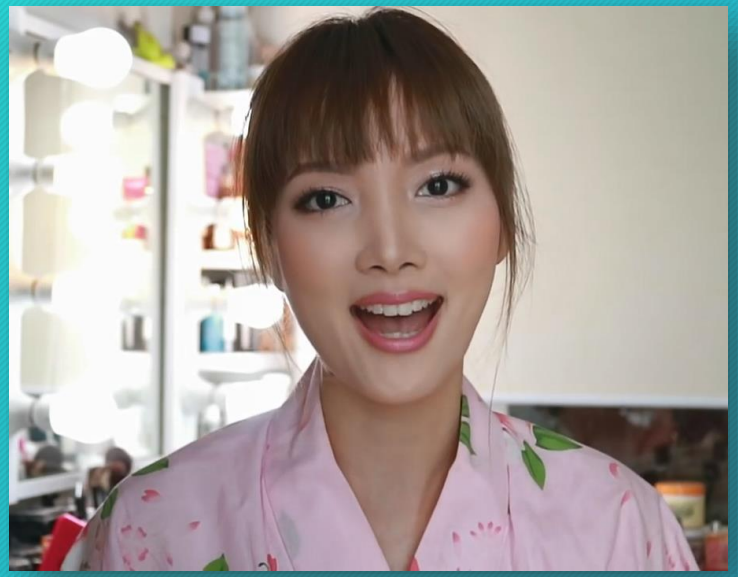
KMUTNB SSRU

Sentiment Analysis of Thai Sound Tones in Social Media Videos

Asst. Prof. Sumitra Nuanmeesri (SSRU)
Lap Poomhiran (KMUTNB)

Introduction

- Different voice of tones in the same word are different feeling
- Detect the emotional of speaker



Happiness?

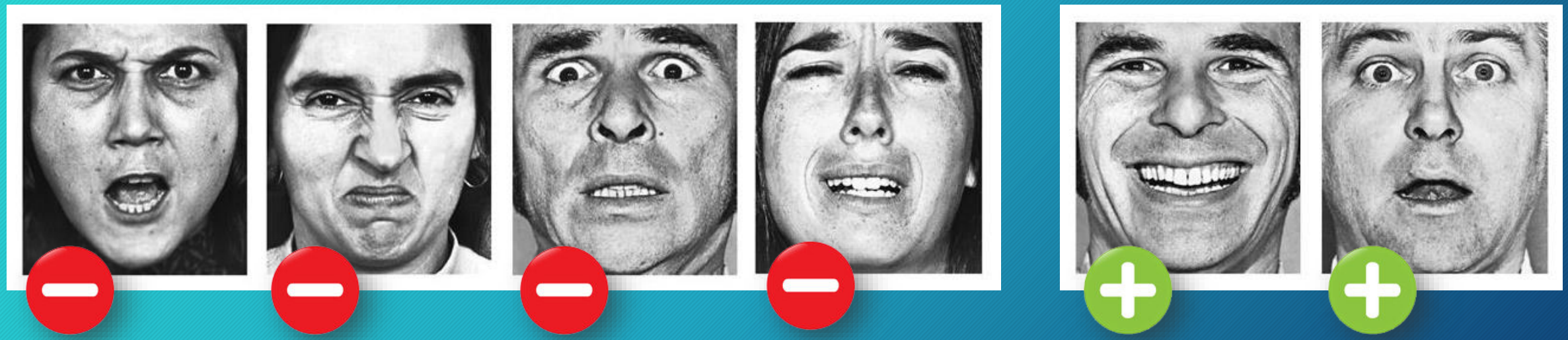
Anger?

Sadness?

Source: www.youtube.com

Six basic emotions – Paul Ekman (1972)

Anger Disgust Fear Sadness Happiness Surprise



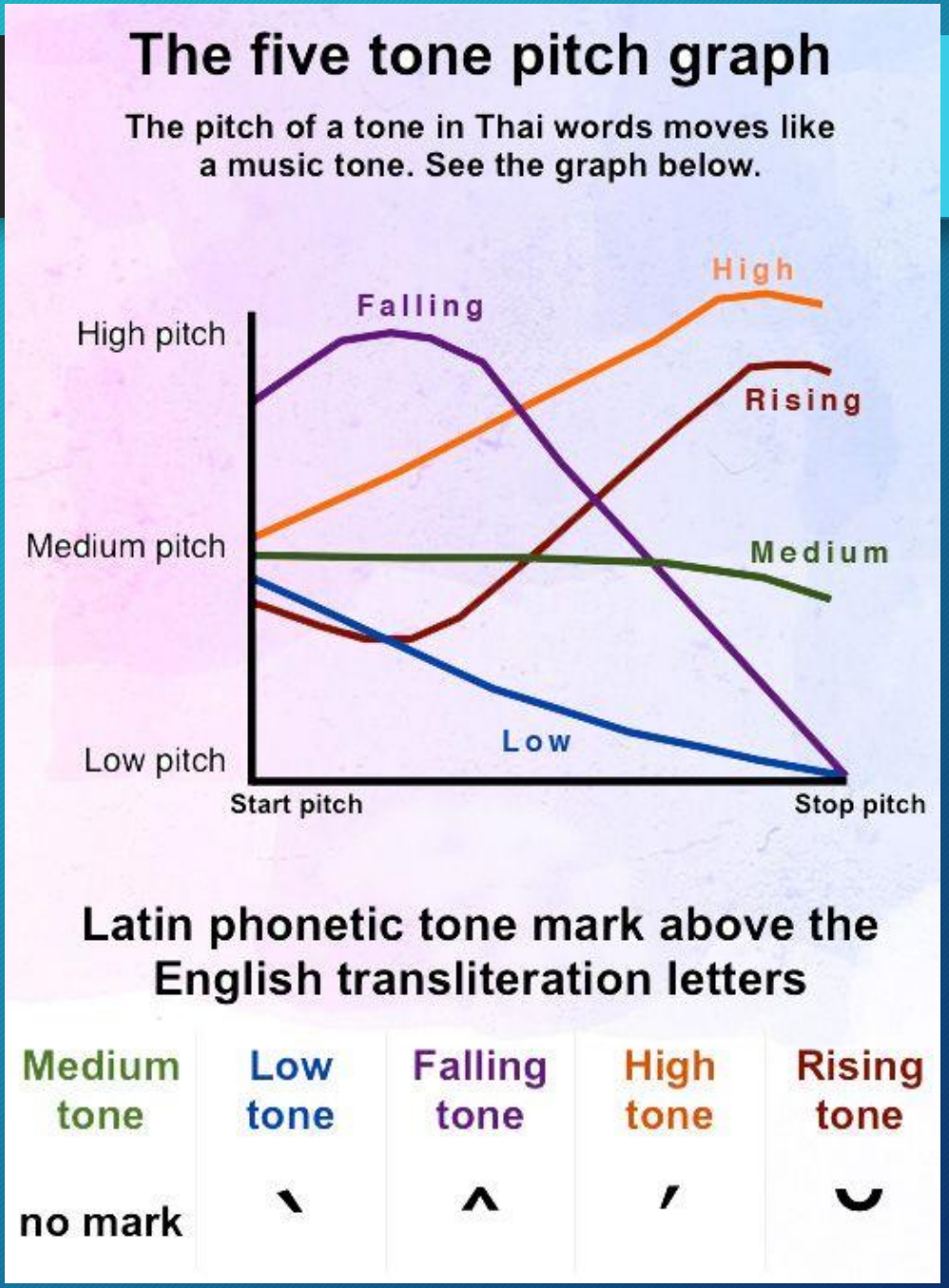
= Positive emotion

= Negative emotion

Photographer by: Paul Ekman

Sound tone

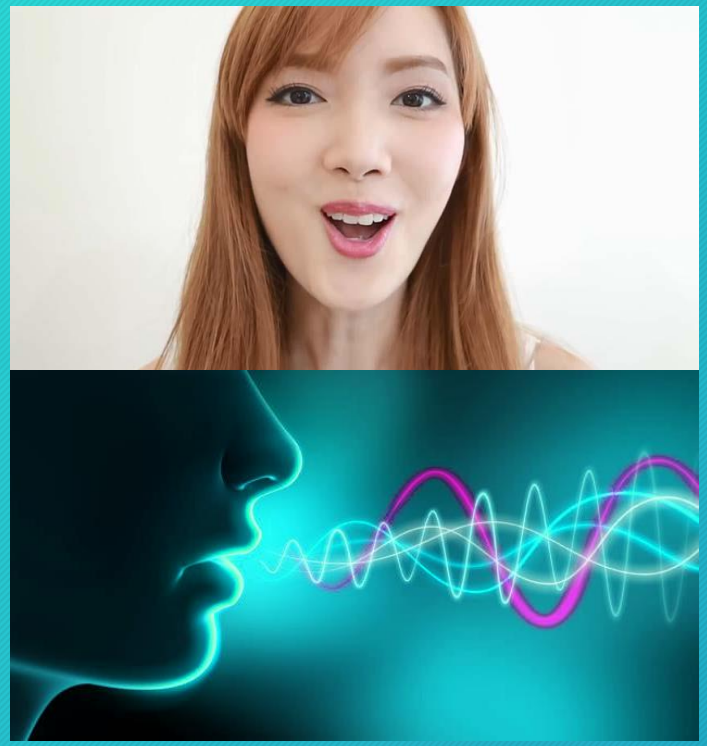
- Every sound has a tone.
- Tones are used for different purposes in different languages.
- There are five tones in Thai language.
- Tones in Thai are used to indicate meaning of words.

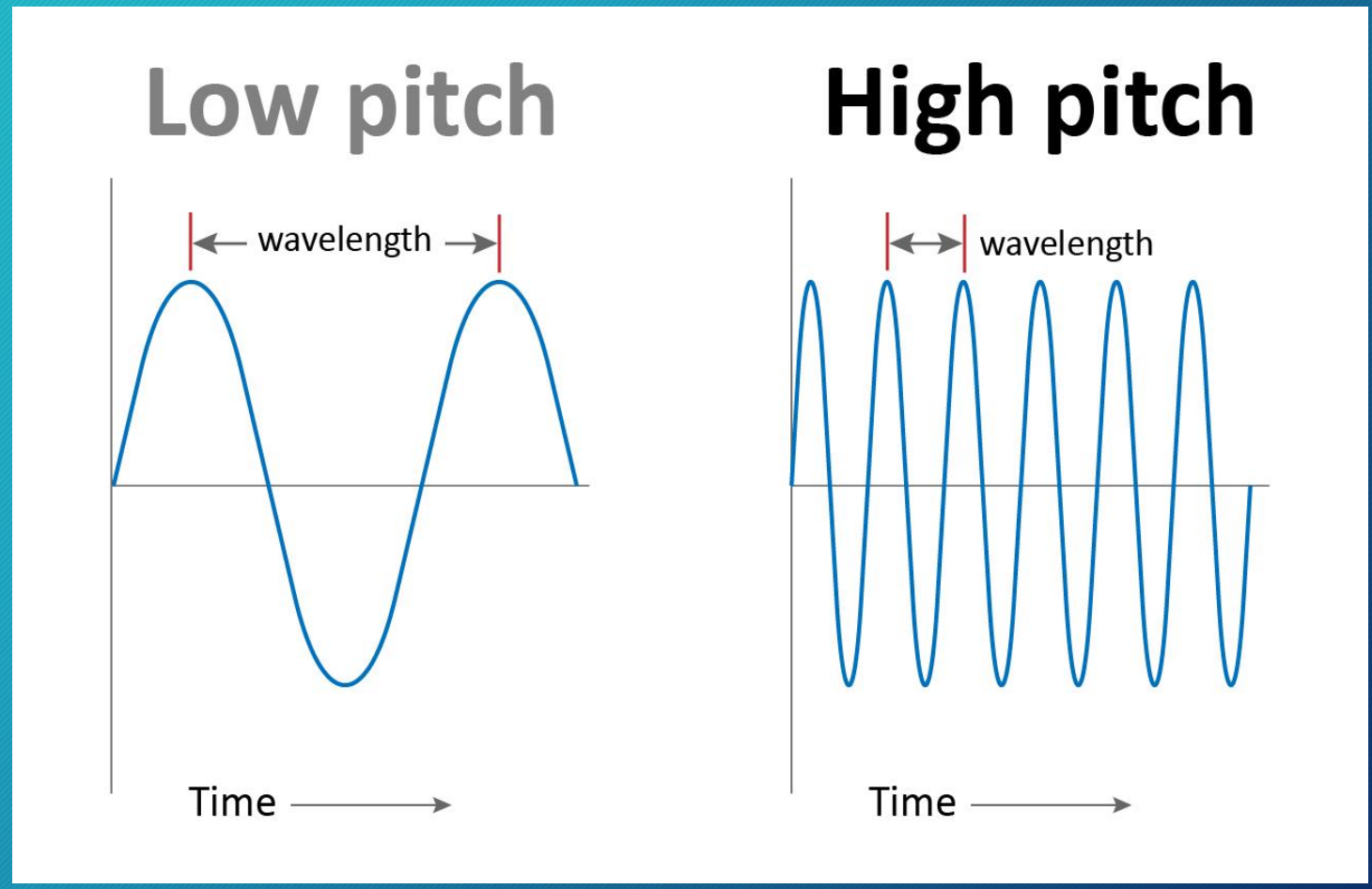


Source: www.thailanguagehut.com

Sound wave: Pitch

Pitch ⇔ Frequency

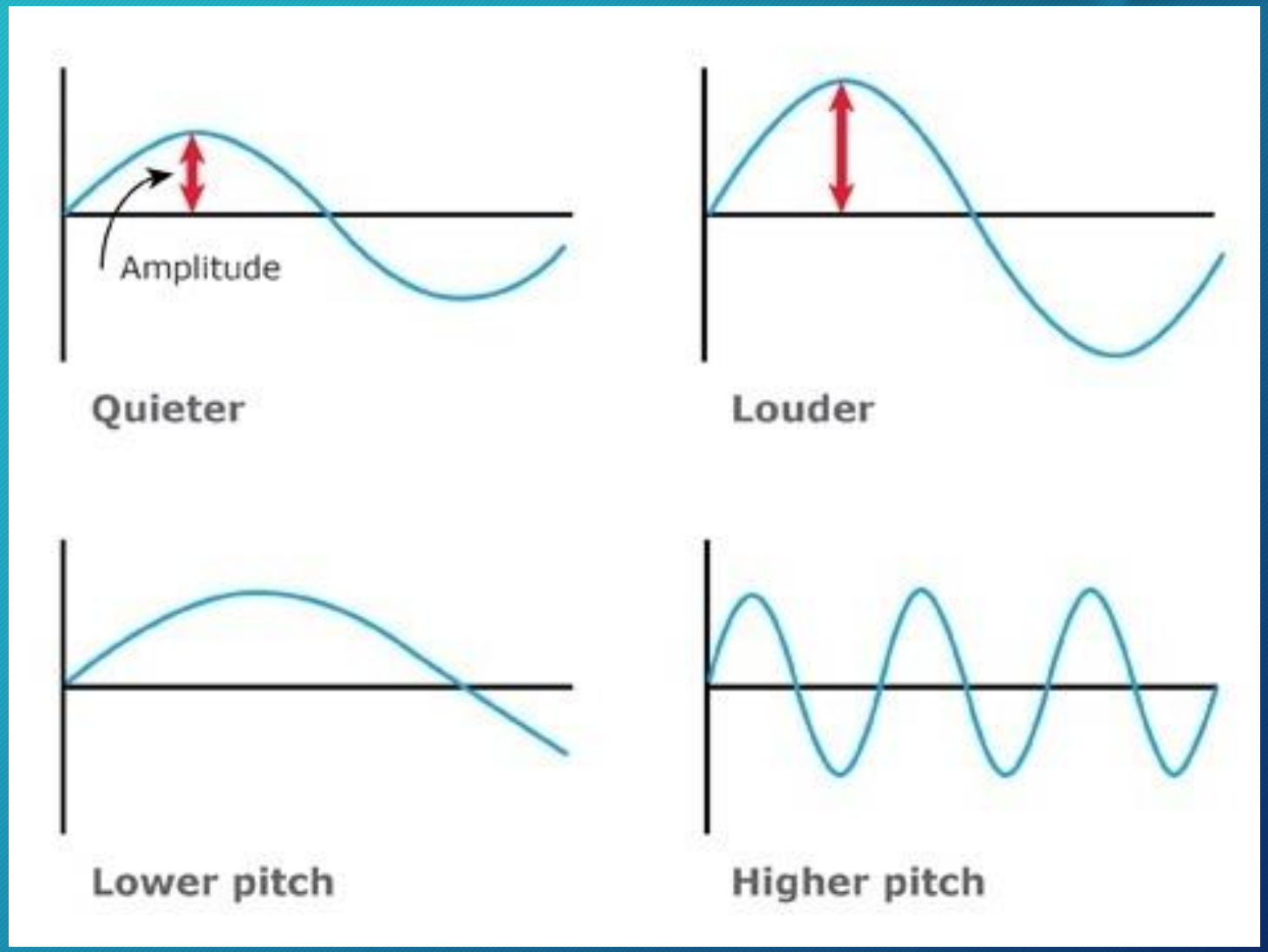




Source: saccoccioscience.wordpress.com

Sound wave: Amplitude

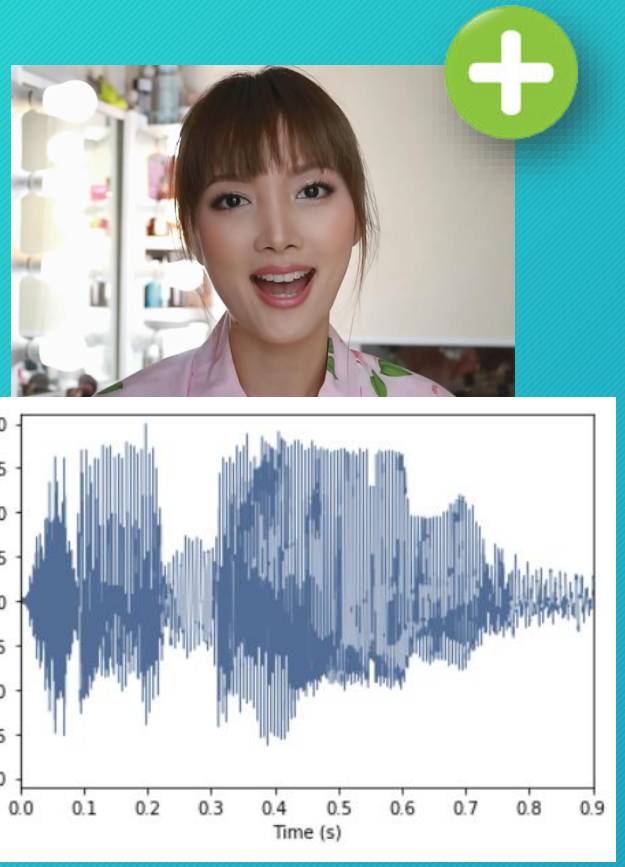
Loudness ⇔ Amplitude



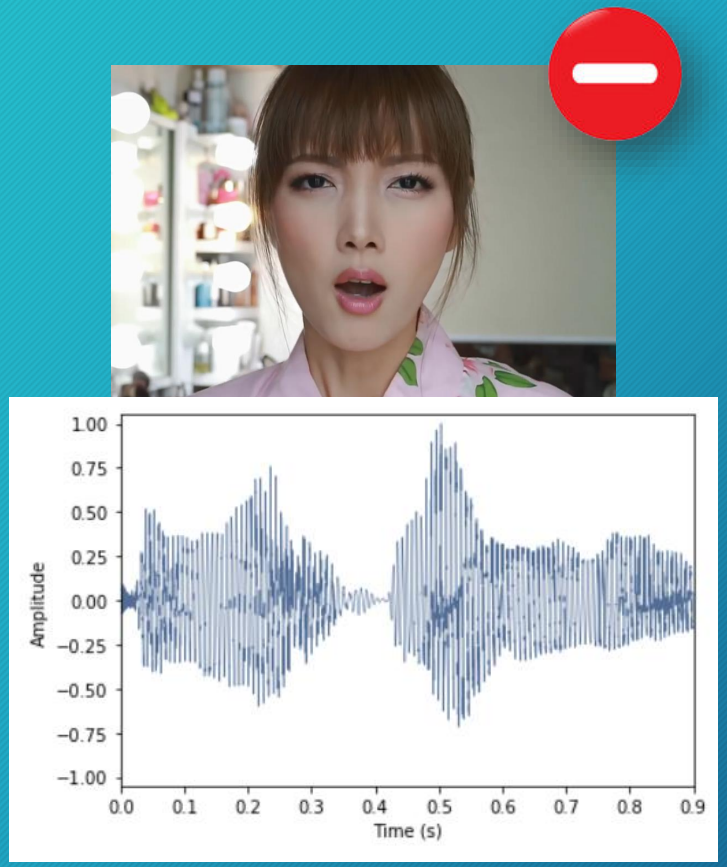
Source: www.scienceteam.org.nz

Tone with sound wave for each emotion

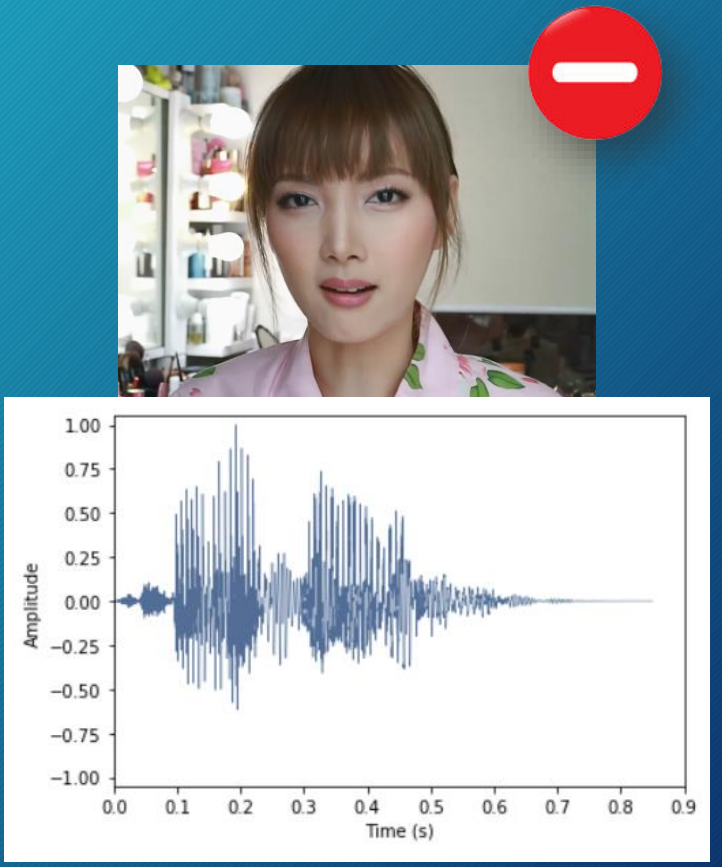
- Same word with audio signal in different sound tone.



Emotion = Happiness

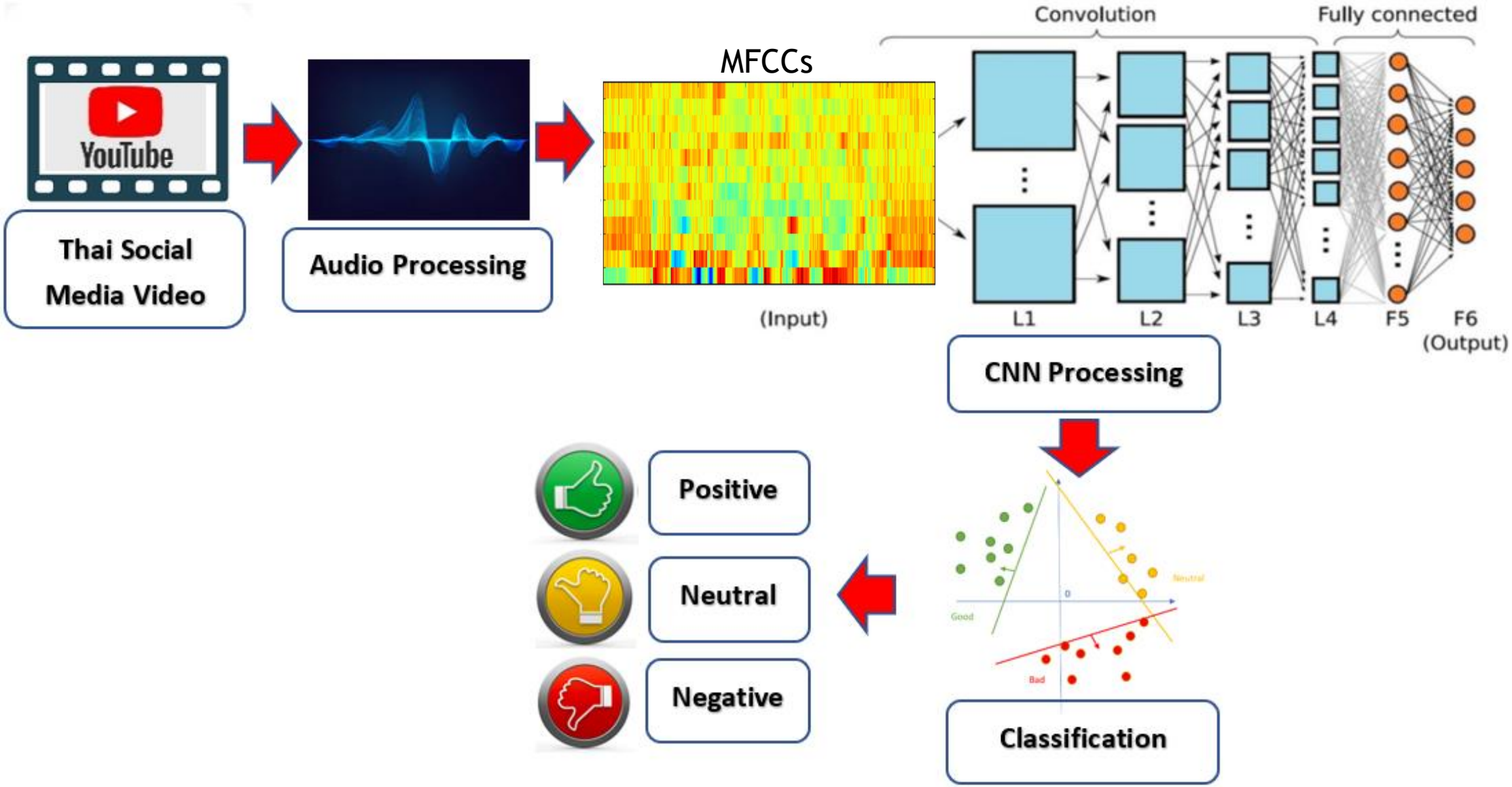


Emotion = Fear



Emotion = Disgust

Conceptual Model



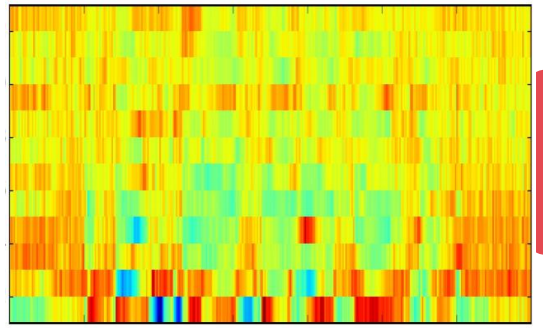
Audio processing

Thai Social Media Video

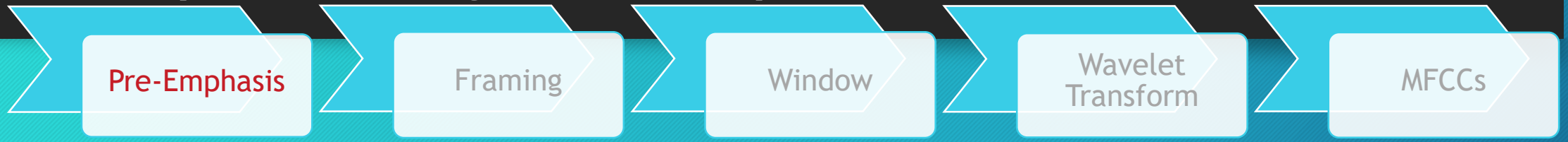
Split Audio

Audio Processing

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs



Audio processing: Pre-Emphasis



A pre-emphasis filter is useful in several ways:

(1) Balance the frequency spectrum.

(2) To improve the Signal-to-Noise Ratio (SNR).

The pre-emphasis filter can be applied to a signal x using the first order filter in the following equation:

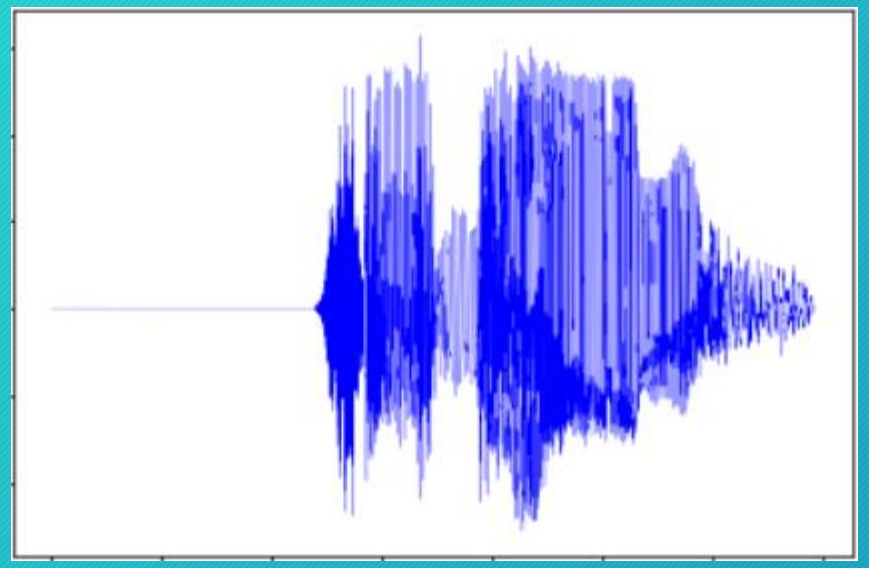
$$y(t) = x(t) - \alpha x(t-1)$$

Typically the filter coefficient (α) are 0.95 or 0.97.

Source: haythamfayek.com

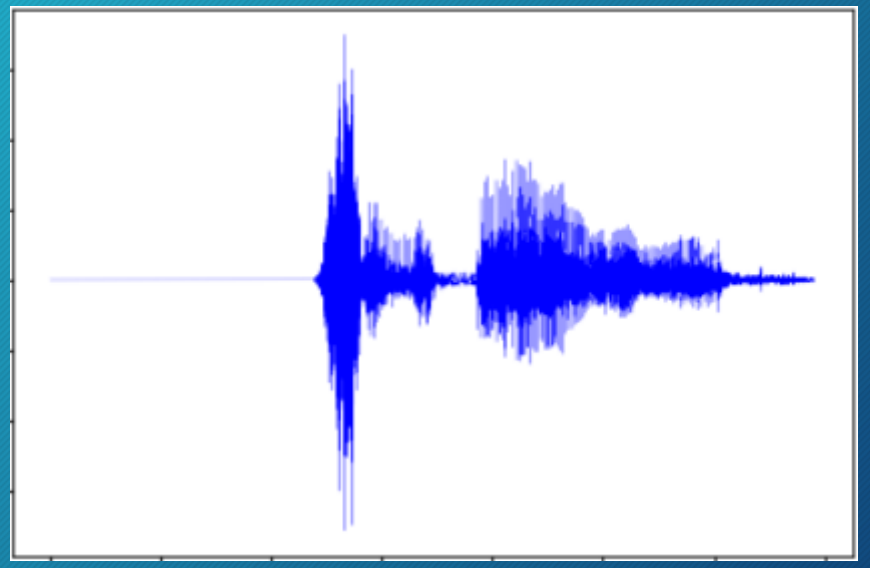
Audio processing: Pre-Emphasis

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs



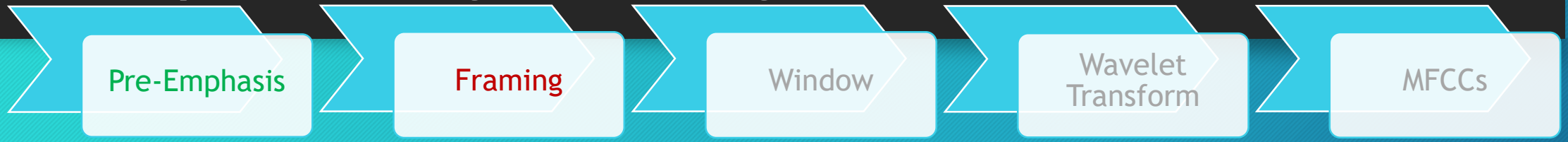
Before

Filtering



After pre-emphasis filtering

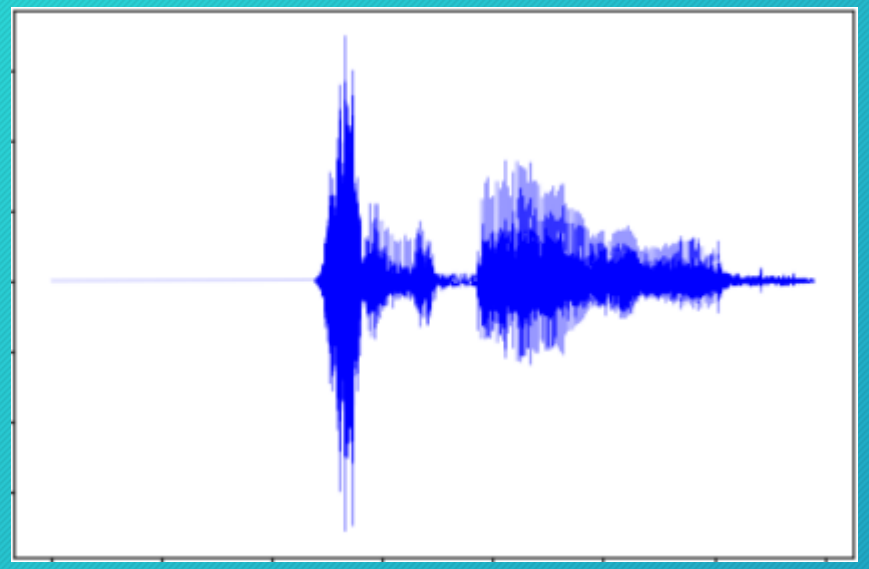
Audio processing: Framing



- To avoid losing the frequency contours of the signal over time.
- To obtain a good approximation of the frequency contours of the signal by concatenating adjacent frames.
- Typical frame sizes in speech processing range from 20 ms to 40 ms with 50% (+/-10%) overlap between consecutive frames.
- Popular settings are 25 ms for the frame size, and a 10 ms stride (15 ms overlap).

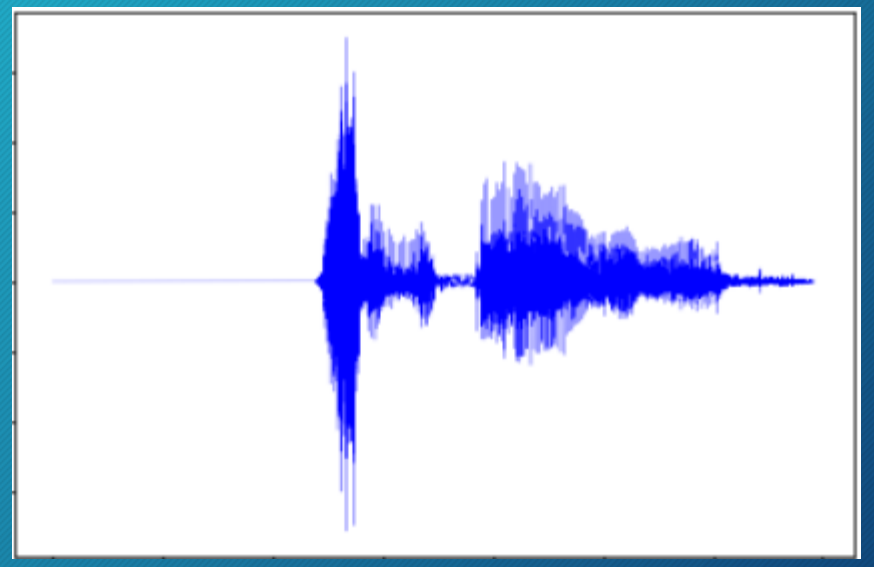
Audio processing: Framing

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs



Before

Framing



After framing process

Audio processing: Window

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs

- To apply a Hamming window function to the frames, accompt that the data is infinite and to reduce spectral leakage.
- A Hamming window has the following form:

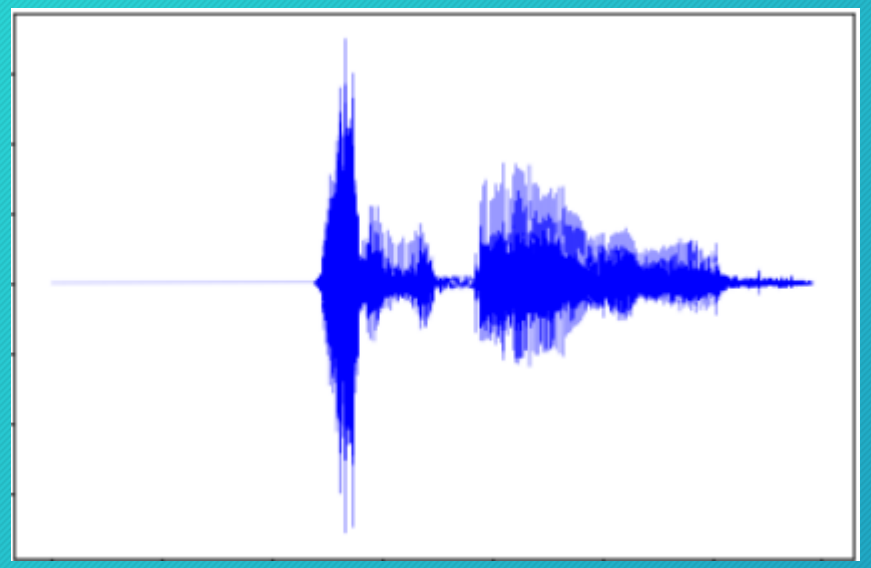
$$w[n] = 0.54 - 0.46cos(\frac{2\pi n}{N-1})$$

where, 0≤n≤N−1, N is the window length.

Source: haythamfayek.com

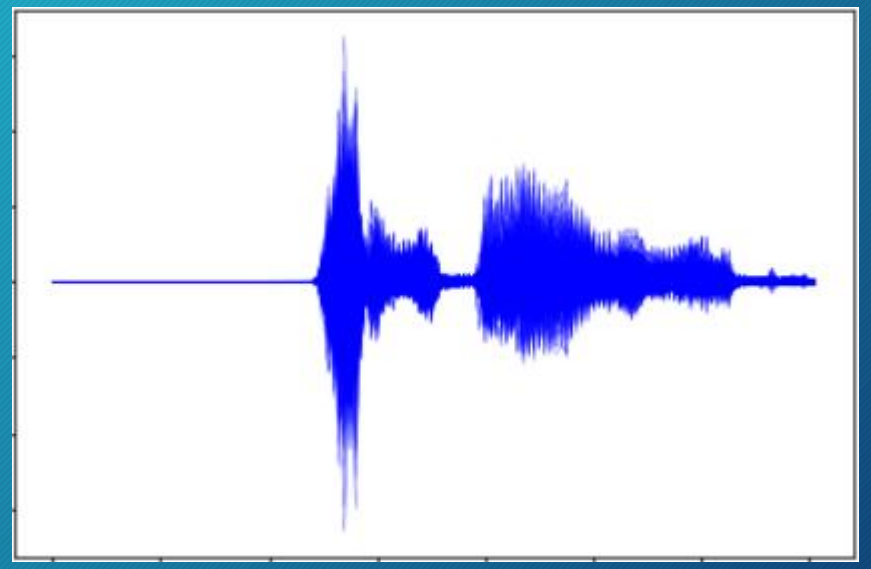
Audio processing: Window

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs



Before

Hamming window

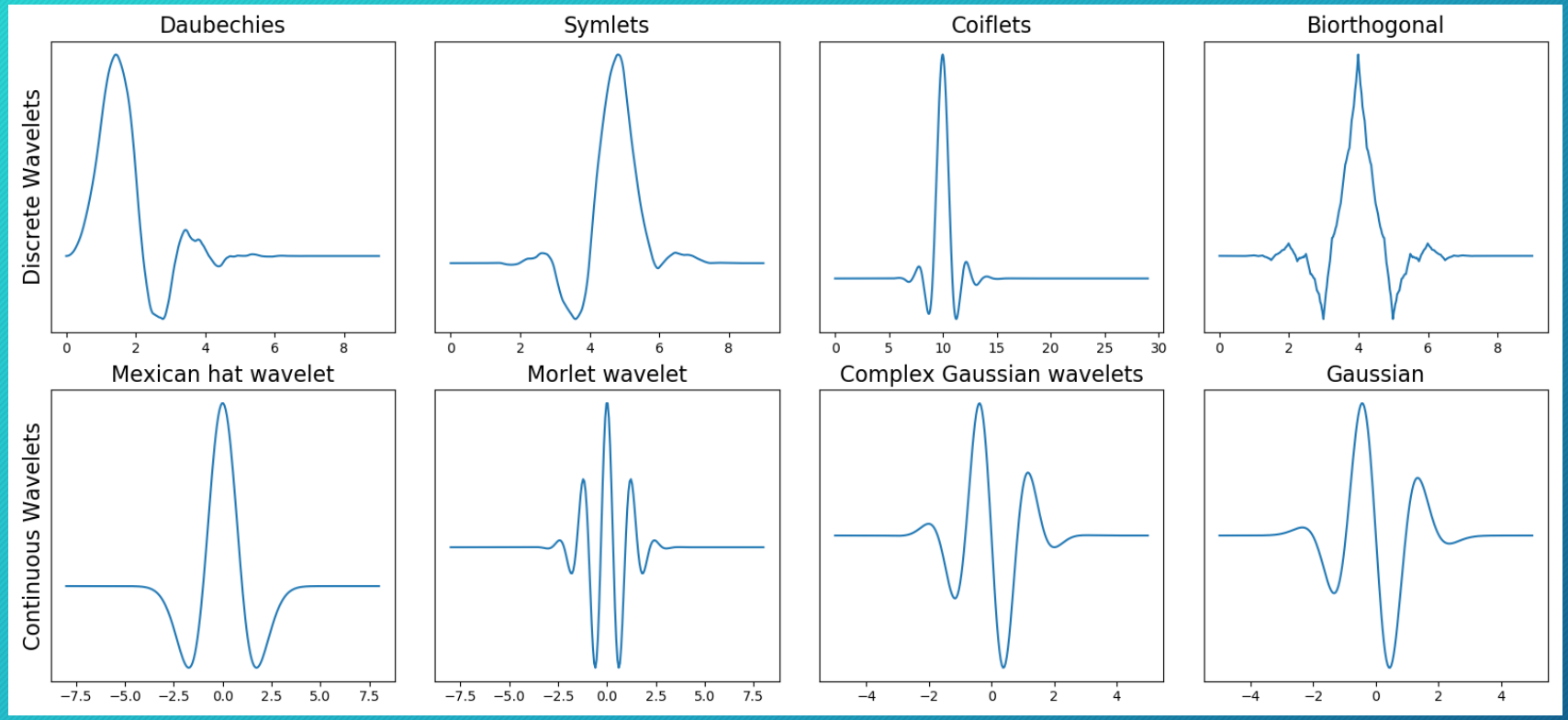


After hamming window

Audio processing: Wavelet Transform

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs

Wavelet families (DWT and CWT)



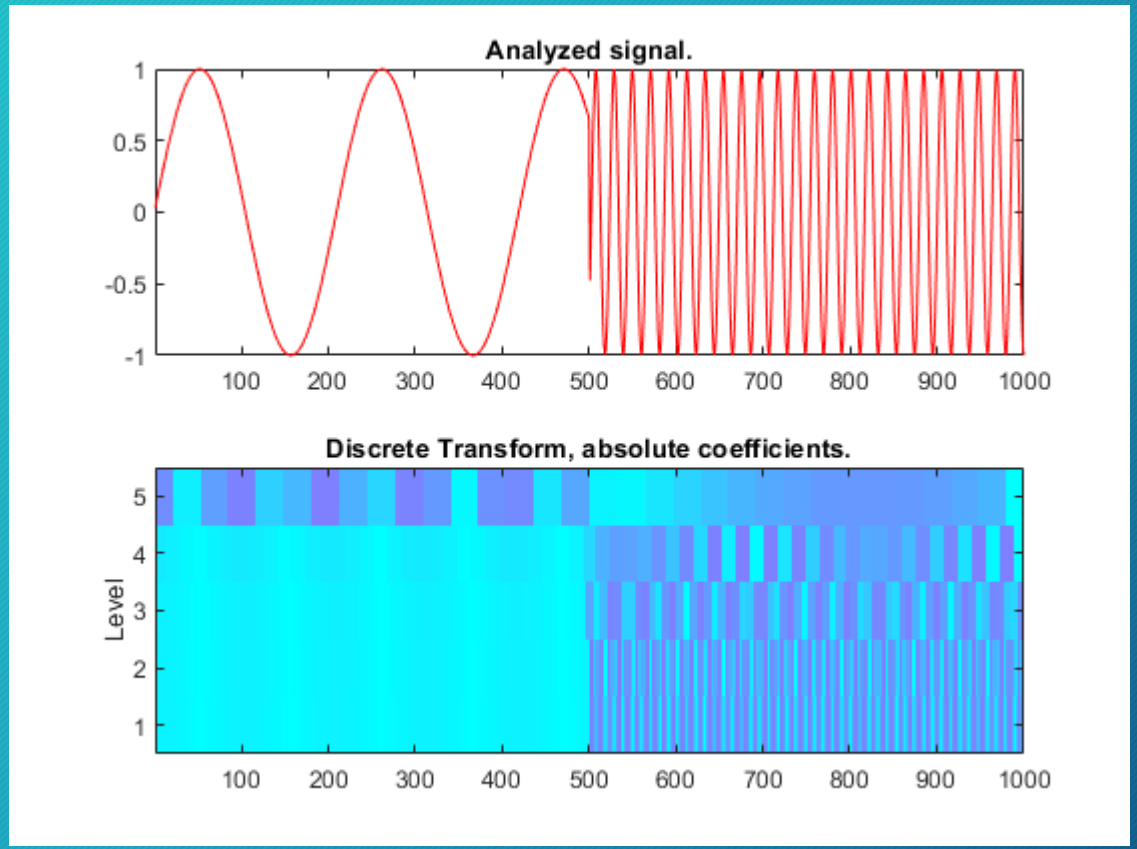
- CWT ⇔ easier than DWT
- DWT ⇔ Denoising, Compression, Restoration
- DWT ⇔ faster than CWT
- Daubechies is suitable for high/low frequency.

Wavelet method is better than Fourier Transform both stationary and non-stationary characters.

Audio processing: Wavelet Transform

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs

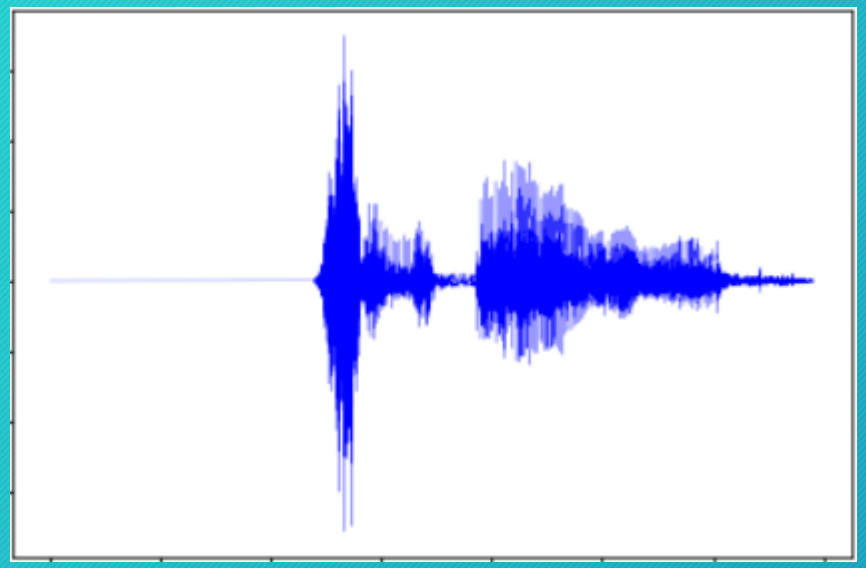
Audio signal is transformed to wavelet.



Source: www.mathworks.com

Audio processing: Wavelet Transform

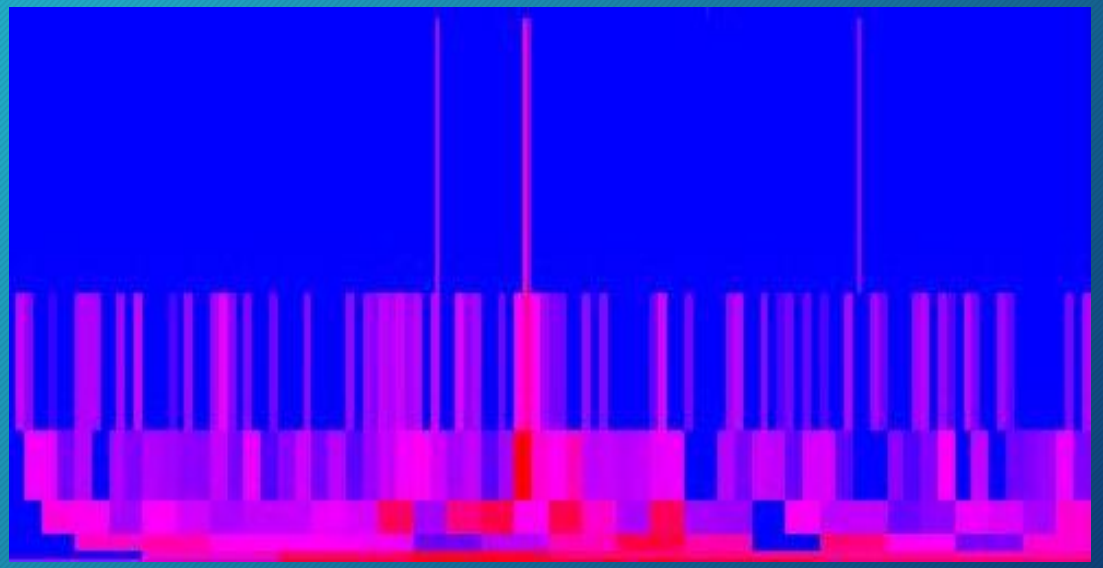
Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs



Before

Wavelet Transform

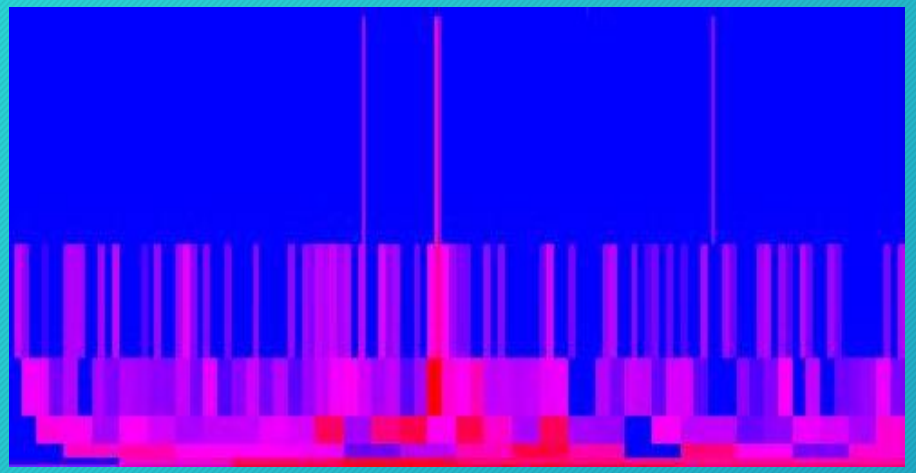
Wavelet Transform (Power Spectrum) of signal



After applied DWT

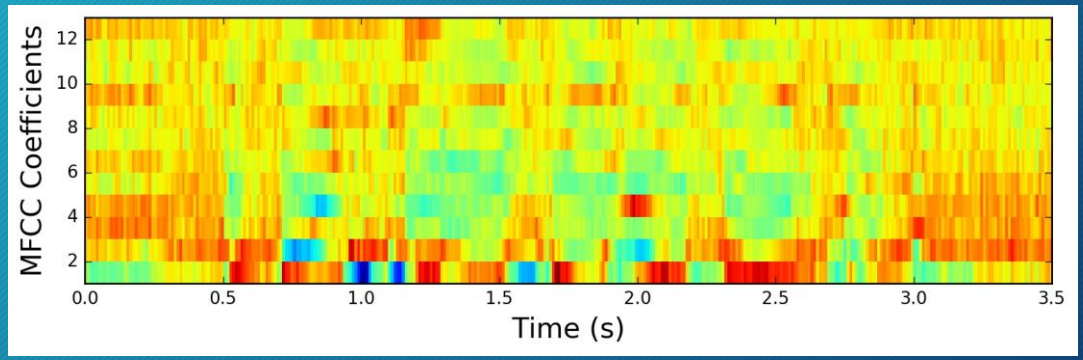
Audio processing: Mel-Frequency Cepstral Coefficients

Pre-Emphasis → Framing → Window → Wavelet Transform → MFCCs



DWT

Feature extraction



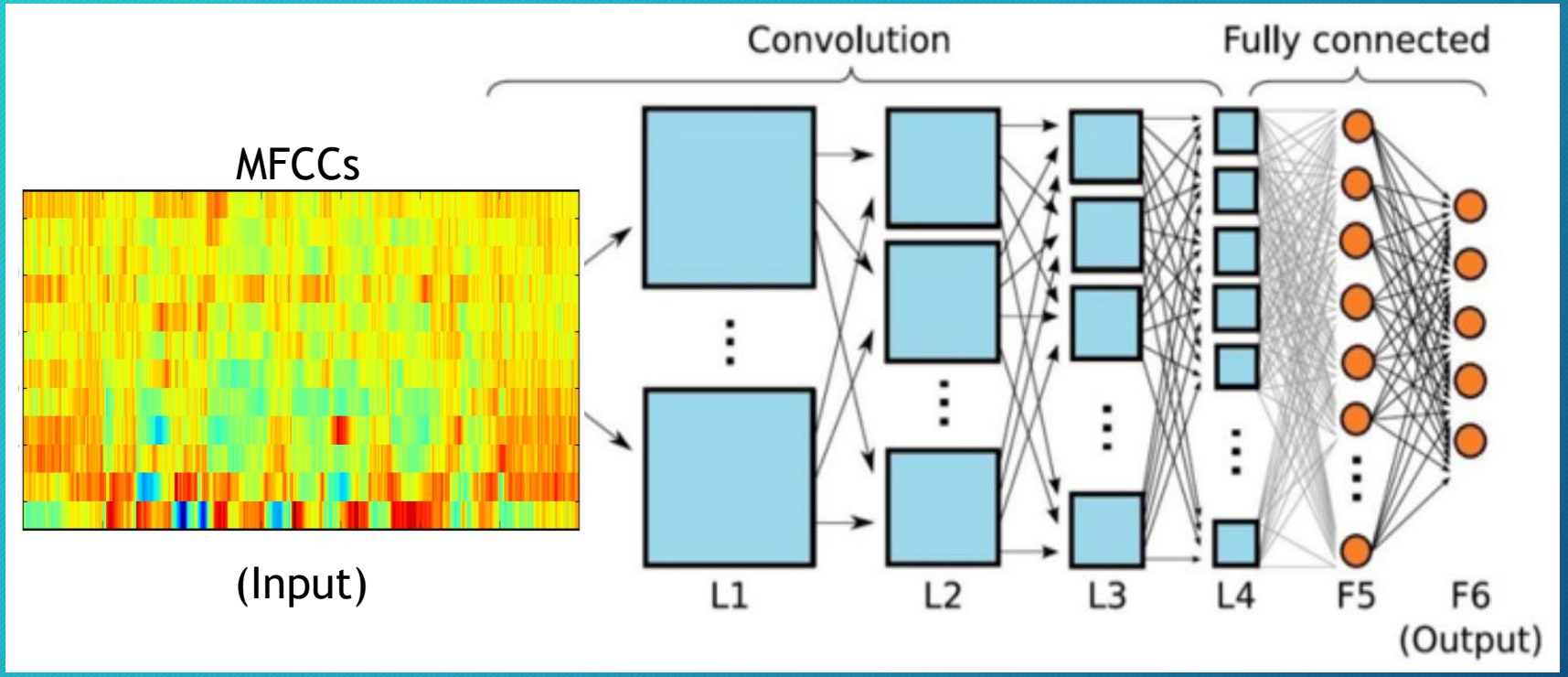
MFCCs

Use MFCCs to the machine learning algorithm is susceptible to correlated input.

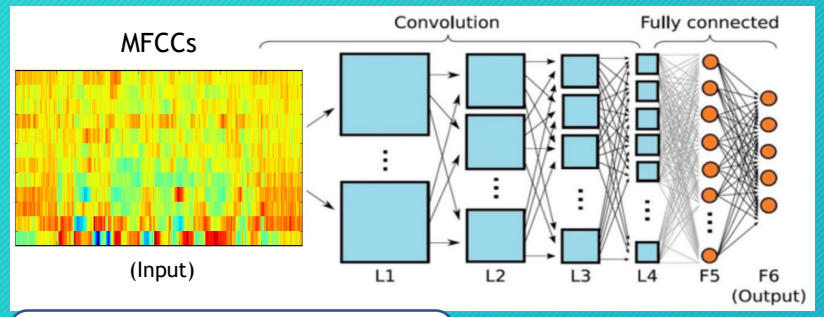
Source: haythamfayek.com

CNN Processing

- Applied MFCCs to Convolutional Neural Network for learning Thai sound tone.

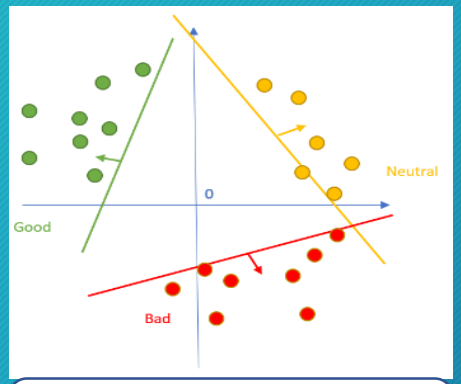


Classification: Support Vector Machine (SVM)



CNN

Classification



SVM

Sentiment Output

Emotion

Positive

Neutral

Negative

THANK YOU.