心理與神經資訊學
(Psychoinformatics & Neuroinformatics)

課號: Psy5261

識別碼: 227U9340

教室:彷彿在雲端

時間: 一789


請先裝librosa, pyaudioanalysis
speech_recognition, gtts

心理學案例研究: 情緒辨識

視覺較易判斷類別; 聽覺較易判斷強度

應用上多為跨感官整合判斷

心理學案例研究:男女誰多話?

Sample	Year	Location	Duration	Age range (years)	Sample size (*N*)		Estimated average number (SD) of words spoken per day	
					Women	Men	Women	Men
1	2004	USA	7 days	18–29	56	56	18,443 (7460)	16,576 (7871)
2	2003	USA	4 days	17–23	42	37	14,297 (6441)	14,060 (9065)
3	2003	Mexico	4 days	17–25	31	20	14,704 (6215)	15,022 (7864)
4	2001	USA	2 days	17–22	47	49	16,177 (7520)	16,569 (9108)
5	2001	USA	10 days	18–26	7	4	15,761 (8985)	24,051 (10,211)
6	1998	USA	4 days	17–23	27	20	16,496 (7914)	12,867 (8343)
					Weighted average		16,215 (7301)	15,669 (8633)

Mehl et al., 2007, *Science*

心理學案例研究:逐字稿

為什麼逐字稿這麼多人搶著做？

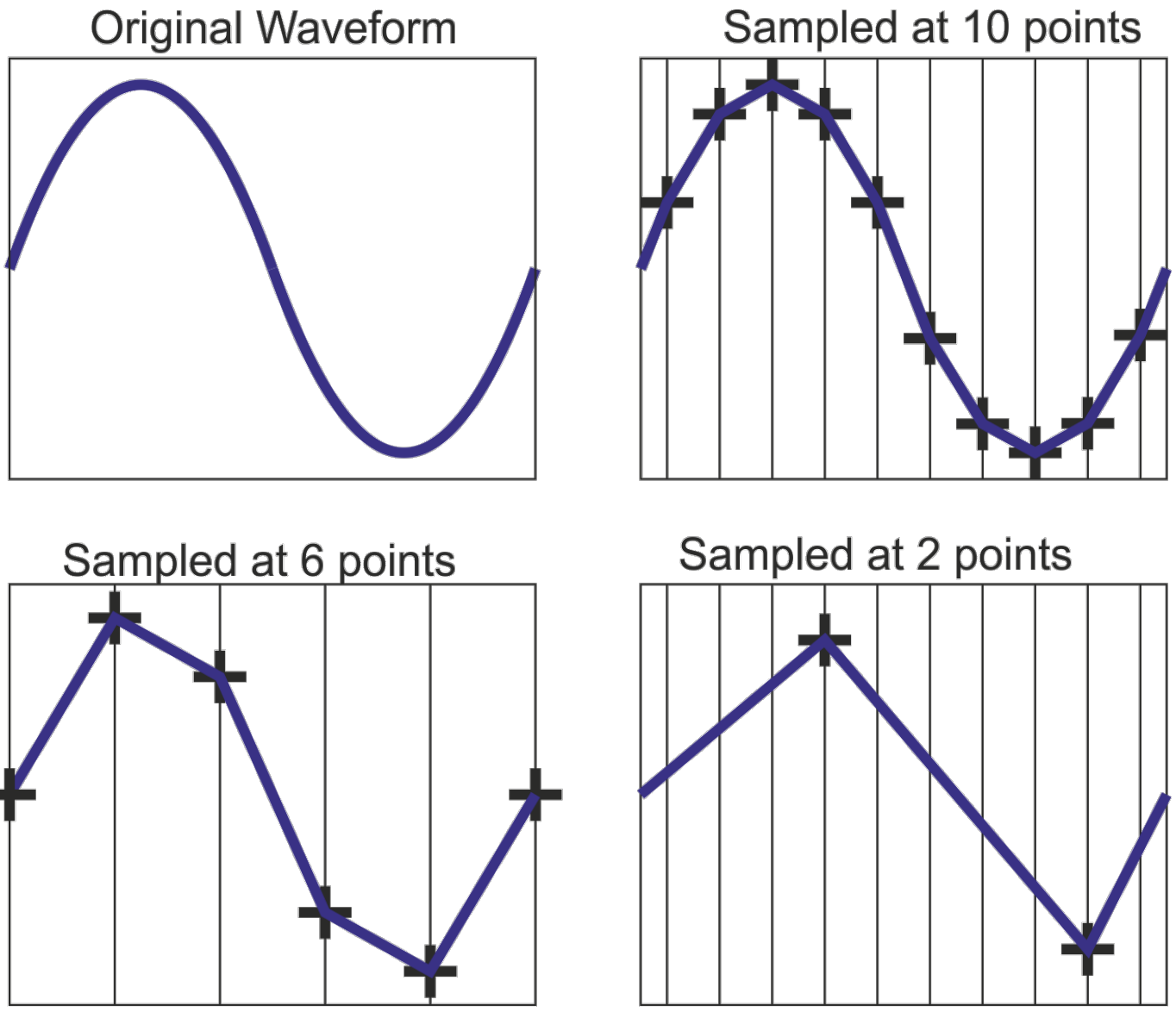
心情 · 12月14日 01:21

PTT打工板只要是逐字稿的工作
下面大約五六個留言已寄信
加上沒留言的我猜至少有十個人應徵
我現在是在某處做行政工讀
有時後主管會請工讀生打逐字稿
自己目前已做過三個檔案總共五小時四十分
超痛苦😵
不知道是不是因為自己打字太慢
自覺不是很能勝任這項事情
一小時的case含休息大概要打16、17個小時左右
就兩個工作天
主管就說我給別的工讀生只要一天就好
哭哭感覺要被fire力😱
一開始沒碰過覺得很新鮮
但現在一聽到要打就覺得很煩😫

音訊資料處理
(Audio Processing)

數位訊號處理(1/3)

原始聲音訊號在time domain



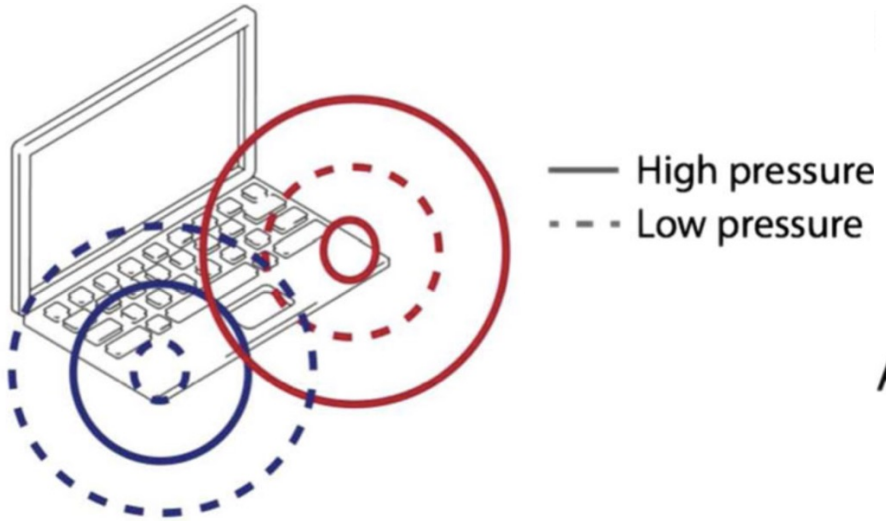
數位訊號處理(2/3)

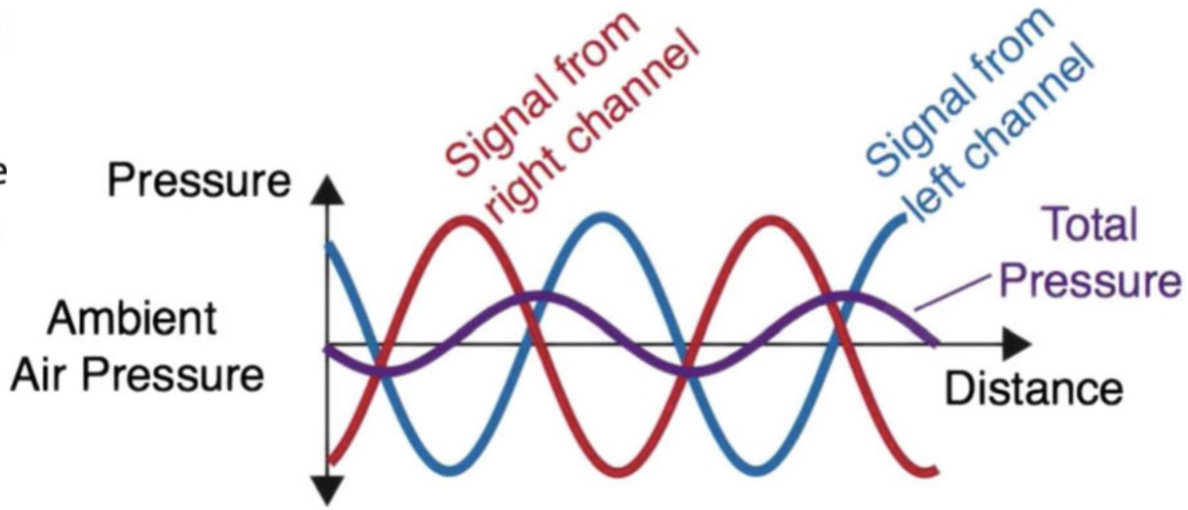
讓雙聲道的訊號從喇叭放出來可以相互抵銷

Atten Percept Psychophys
DOI 10.3758/s13414-017-1361-2

Headphone screening to facilitate web-based auditory experiments

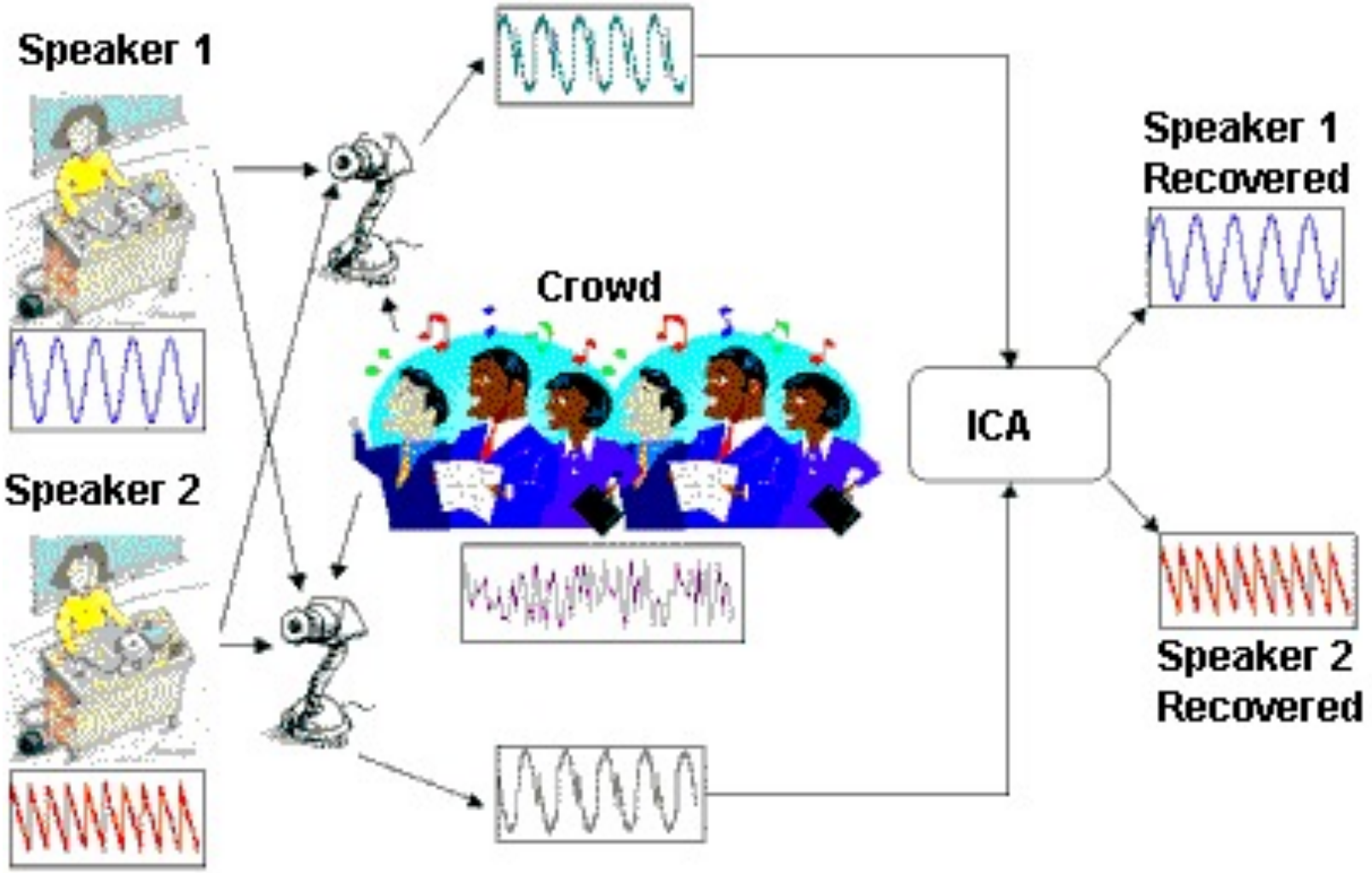
Kevin J. P. Woods[1,2] · Max H. Siegel[1] · James Traer[1] · Josh H. McDermott[1,2]





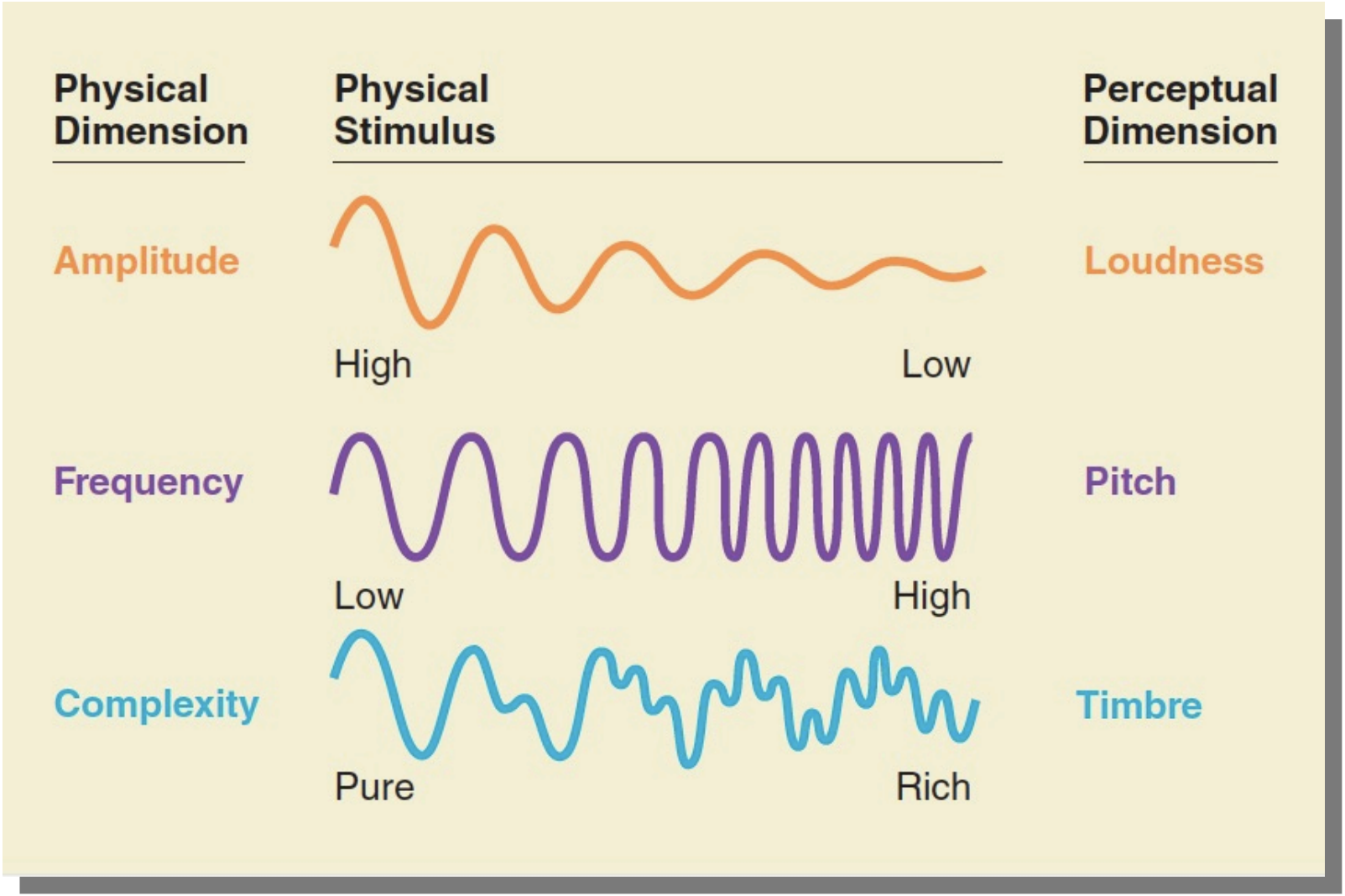
數位訊號處理(3/3)

例如ICA可以把不同的人聲分隔(speaker diarization)



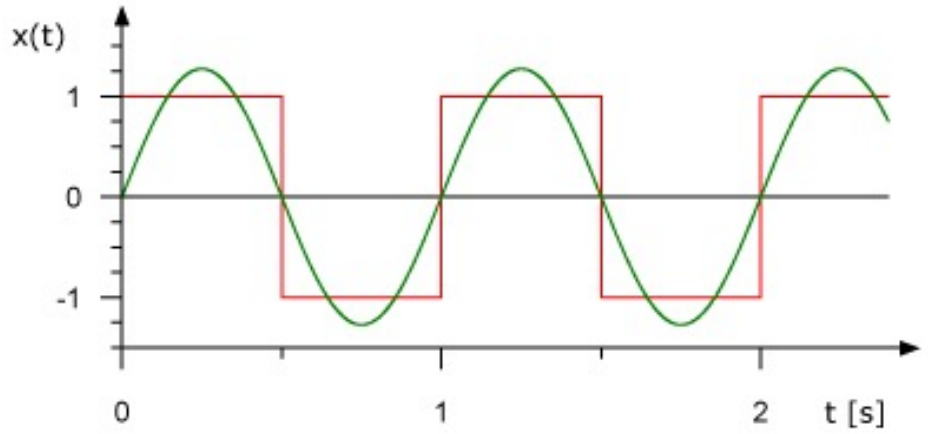
聲音的基本特徵

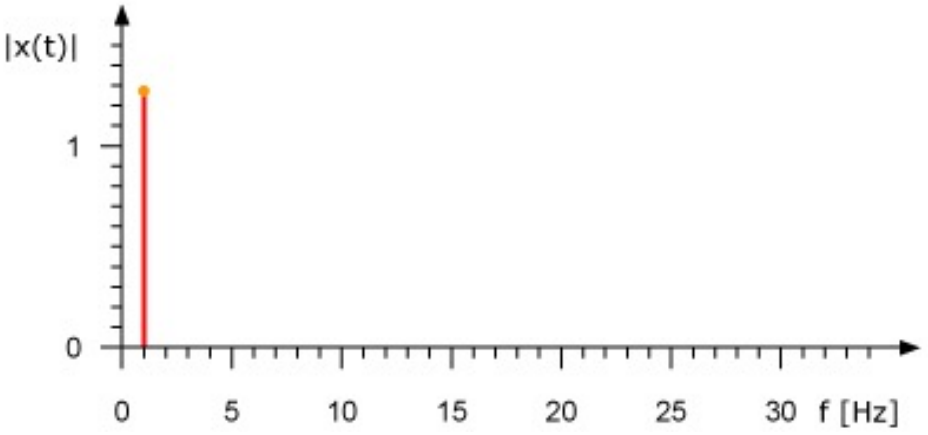
音量、音調/音高、音色



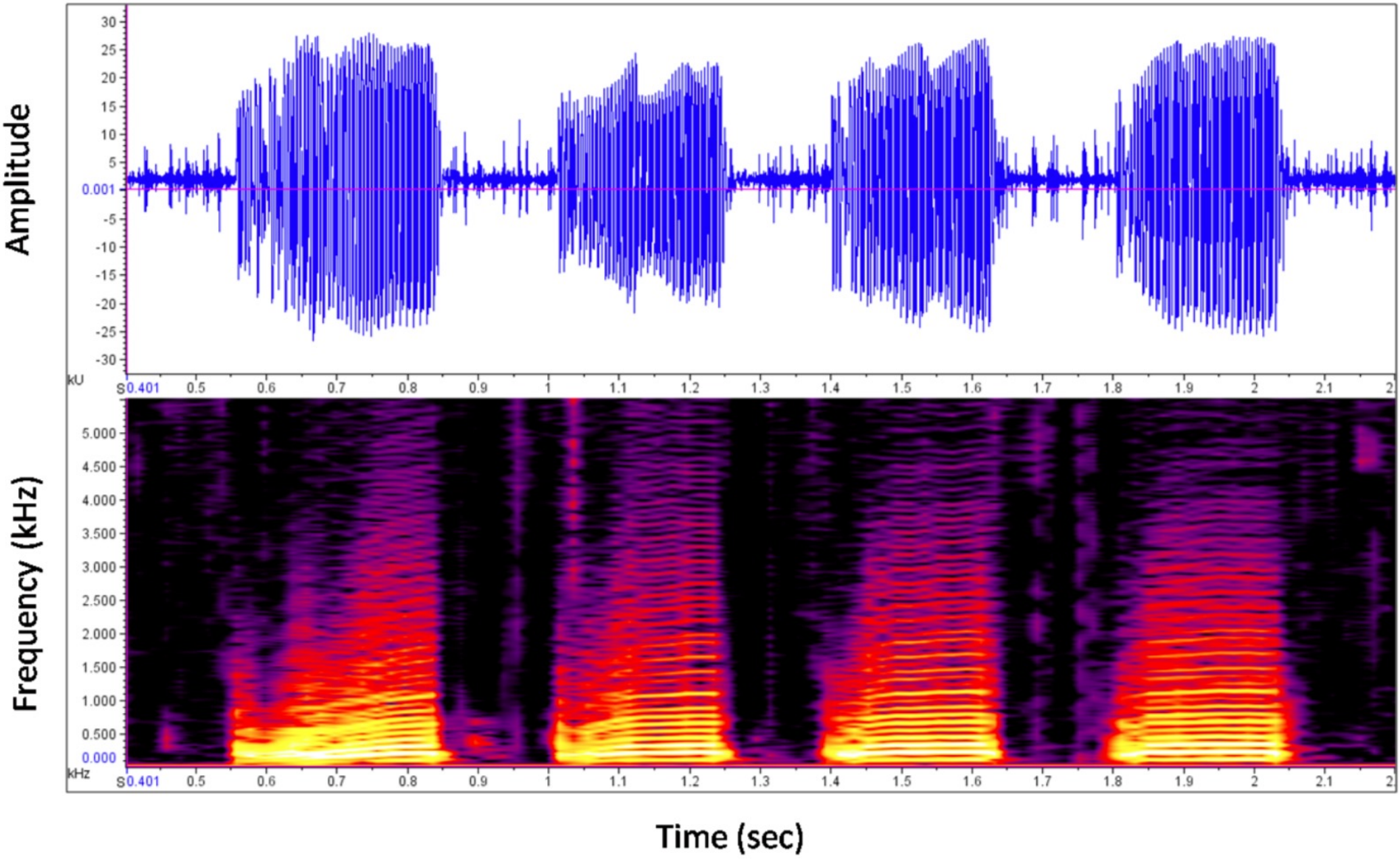
傅立葉分析: Frequency Domain

$$y(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} [a_k \cos(2\pi k f_0 t) - b_k \sin(2\pi k f_0 t)]$$

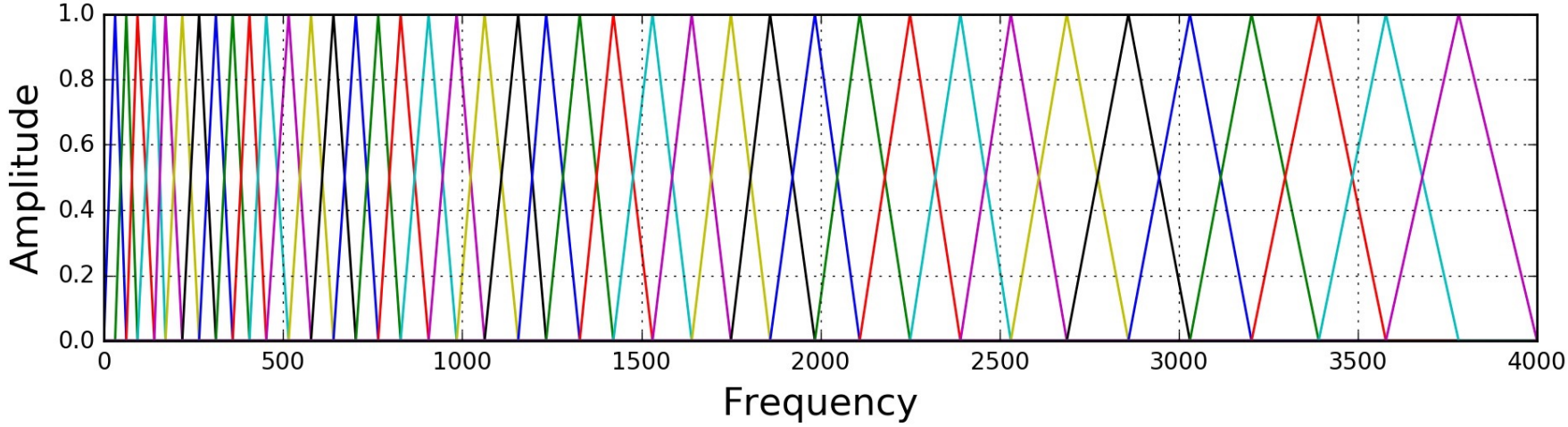


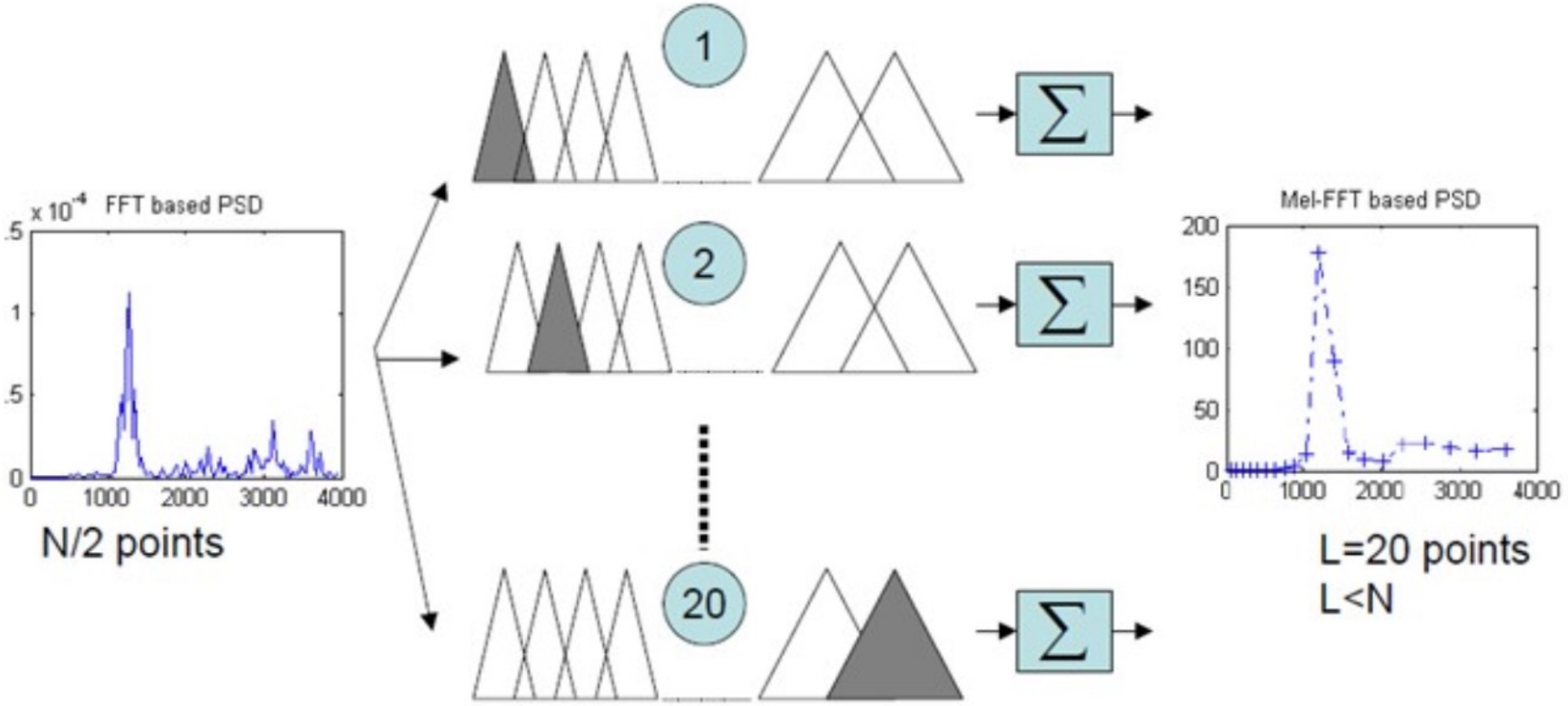


時頻譜(Spectrogram)



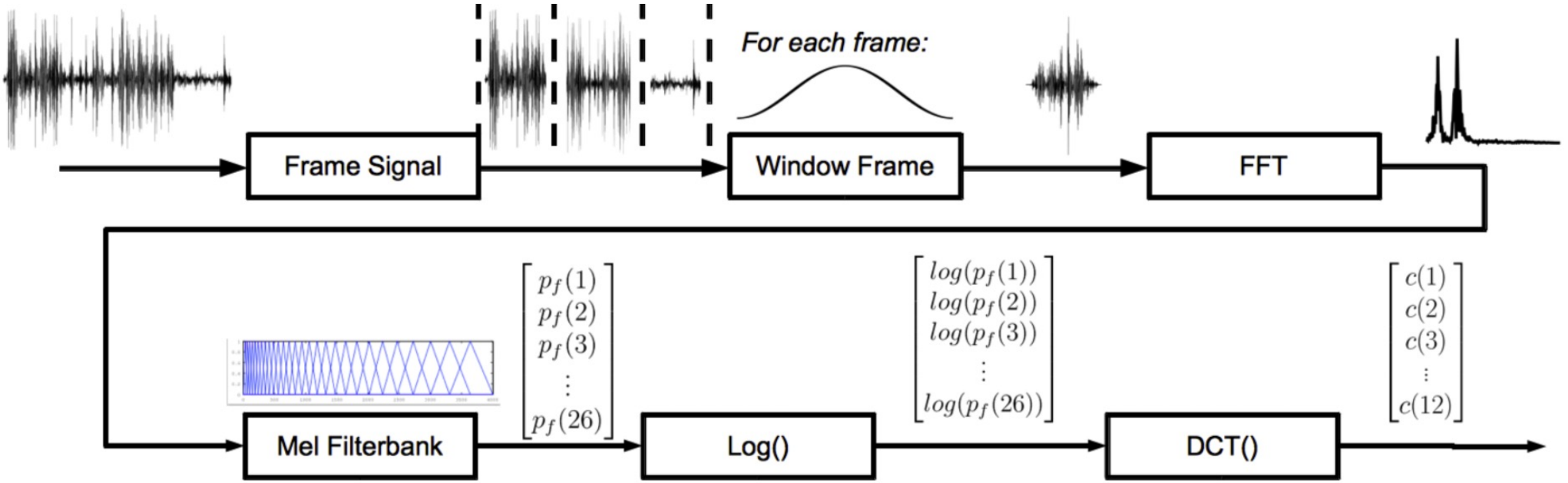
Mel Filters:聽不到的頻率就少採樣





梅爾倒頻譜(MFCC)

把filter coefficients再透過DCT做進一步壓縮



1976就提出的MFCC雖經典但仍是很有效的特徵

各種聲音特徵

openSMILE:) by audEERING™ 可以幫忙萃取出各種聲學特徵

2.5 Default feature sets

For common tasks from the Music Information Retrieval and Speech Processir some example configuration files in the `config/` directory for the following fi ture sets. These also contain the baseline acoustic feature sets of the 2009–201 challenges on affect and paralinguistics:

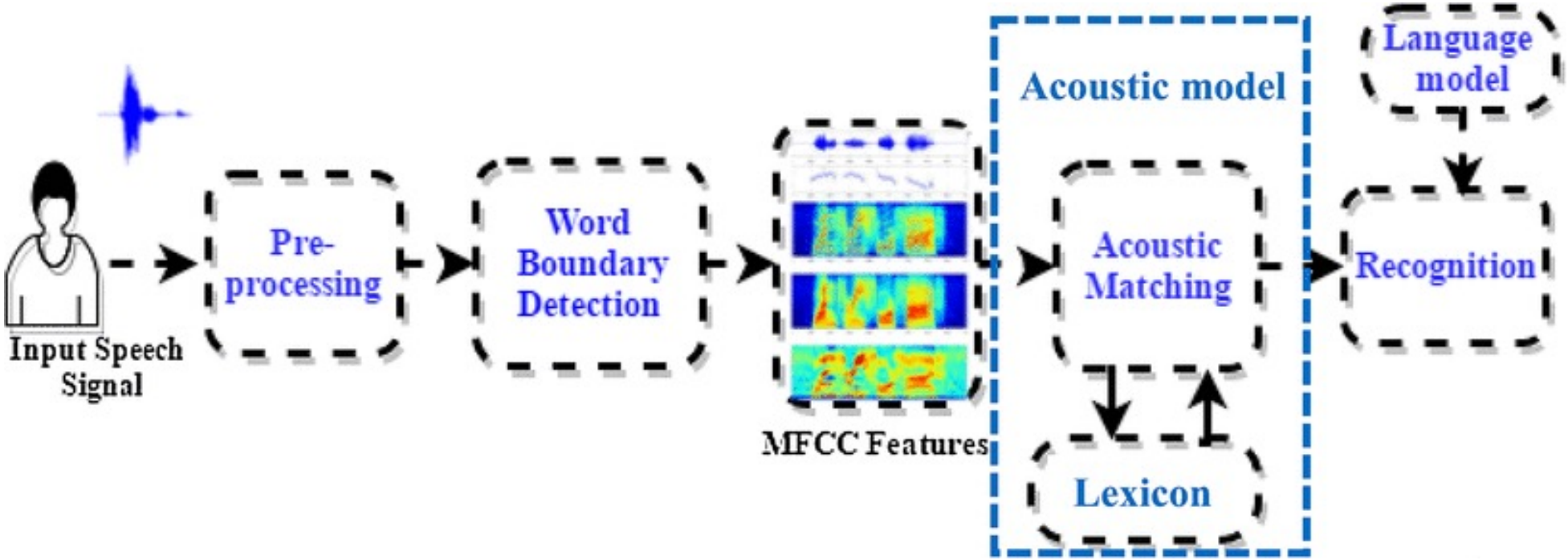
- Chroma features for key and chord recognition
- MFCC for speech recognition
- PLP for speech recognition
- Prosody (Pitch and loudness)
- The INTERSPEECH 2009 Emotion Challenge feature set
- The INTERSPEECH 2010 Paralinguistic Challenge feature set
- The INTERSPEECH 2011 Speaker State Challenge feature set
- The INTERSPEECH 2012 Speaker Trait Challenge feature set
- The INTERSPEECH 2013 ComParE feature set
- The MediaEval 2012 TUM feature set for violent scenes detection.

Acoustic LLDs	
Low-level Descriptors (LLDs)	Type
zero-crossing rate, log energy, probability of voicing, F_0	prosodic
MFCC 0-12, spectral flux, spectral centroid, max, min, spectral bands 0-4 (0-9KHz), spectral roll-off (0.25, 0.5, 0.75, 0.9)	spectral
Functionals applied to LLDs/ΔLLDs/ΔΔLLDs	
position of min/max, range, max − arithmetic mean, arithmetic mean − min	extremes
linear regression slope, offset, error, centroid, quadratic error, quadratic regression a, b offset, linear error, quadratic error (contour & quadratic regression)	regression
percentile range (25%, 50%, 75%), 3 inter-quartile ranges (25% - 50%, 50%-75%, 25%-75%)	percentiles
mean value of peaks, distance between peaks, mean value of peaks − arithmetic mean	peaks
arithmetic means, absolute value of arithmetic mean (original, non-zero values), quadratic mean (original, non-zero values), geometric mean (absolute values of non-zero values), number of non-zero values	means
relative duration LLD above 25%, 50%, 75%, 95% range, relative duration LLD is rising/falling, relative duration LLD has left/right curvature	temporal

語音資料處理
(Speech Processing)

語音辨識(1/3): 考慮字

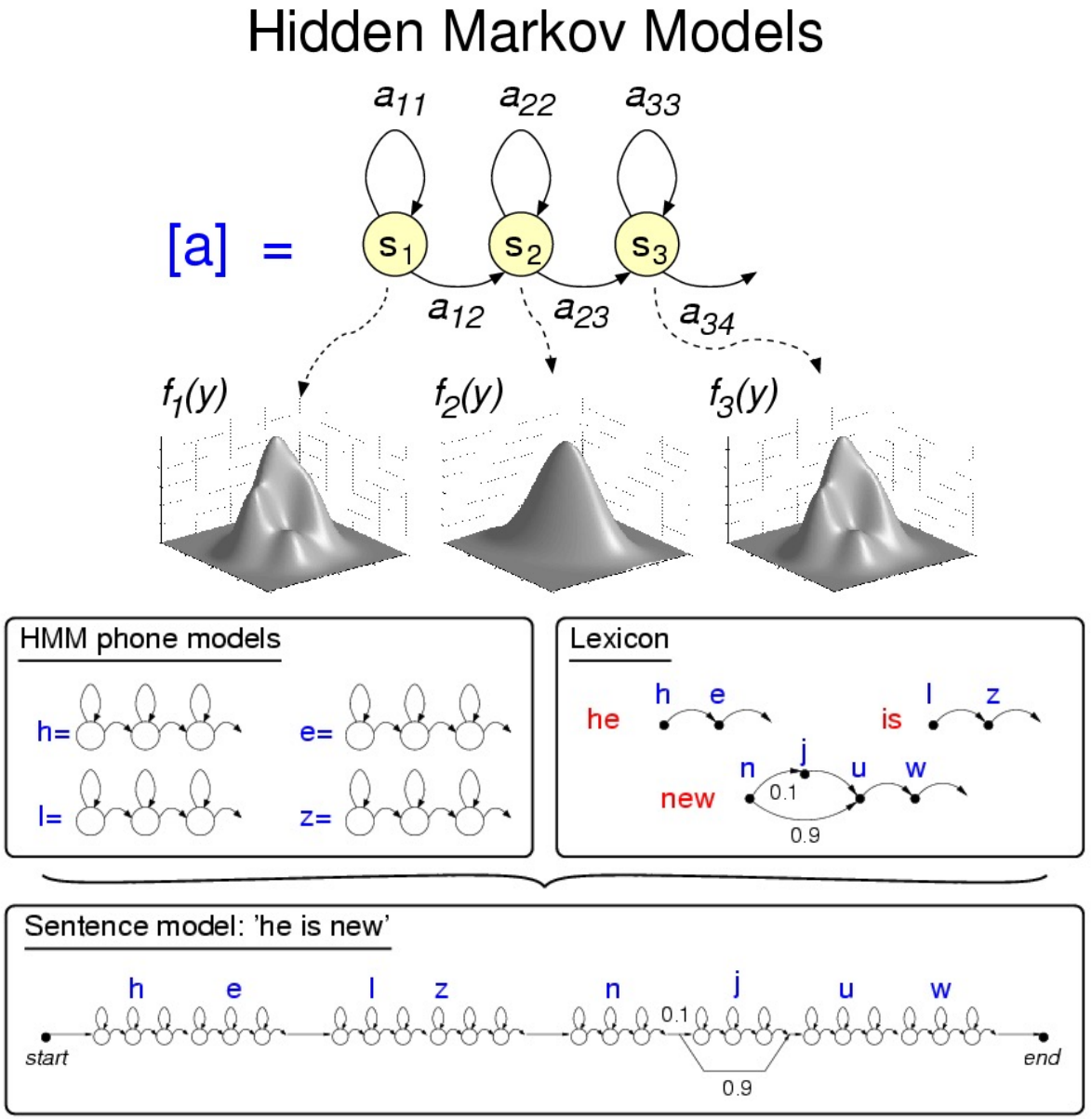
最簡單的想法是用bottom-up訊號比對字的頻譜特徵



若有language model可提供top-down協助

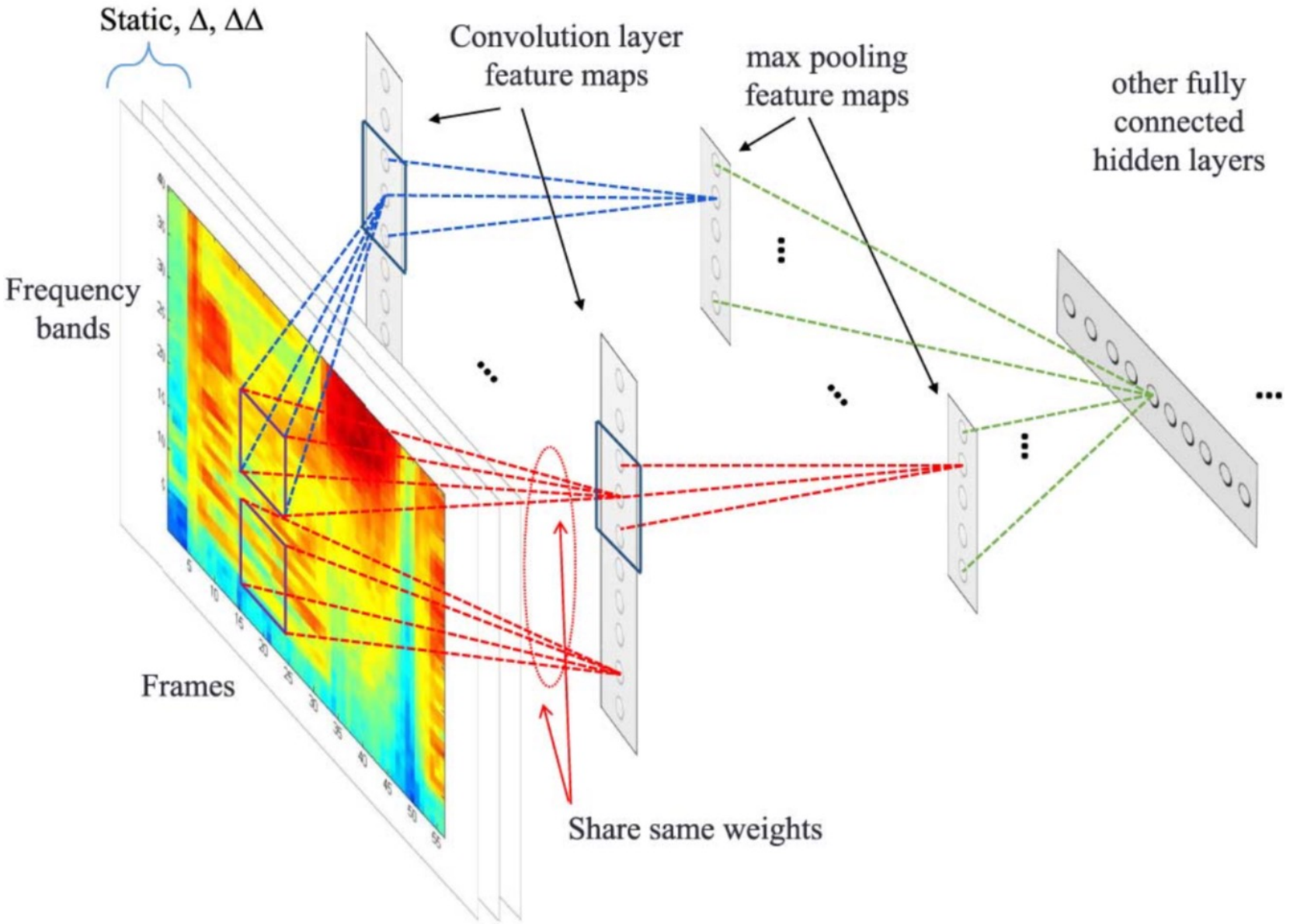
語音辨識(2/3): 考慮字+前字

HMM可model音素→字元、字元→單字、單字→片語



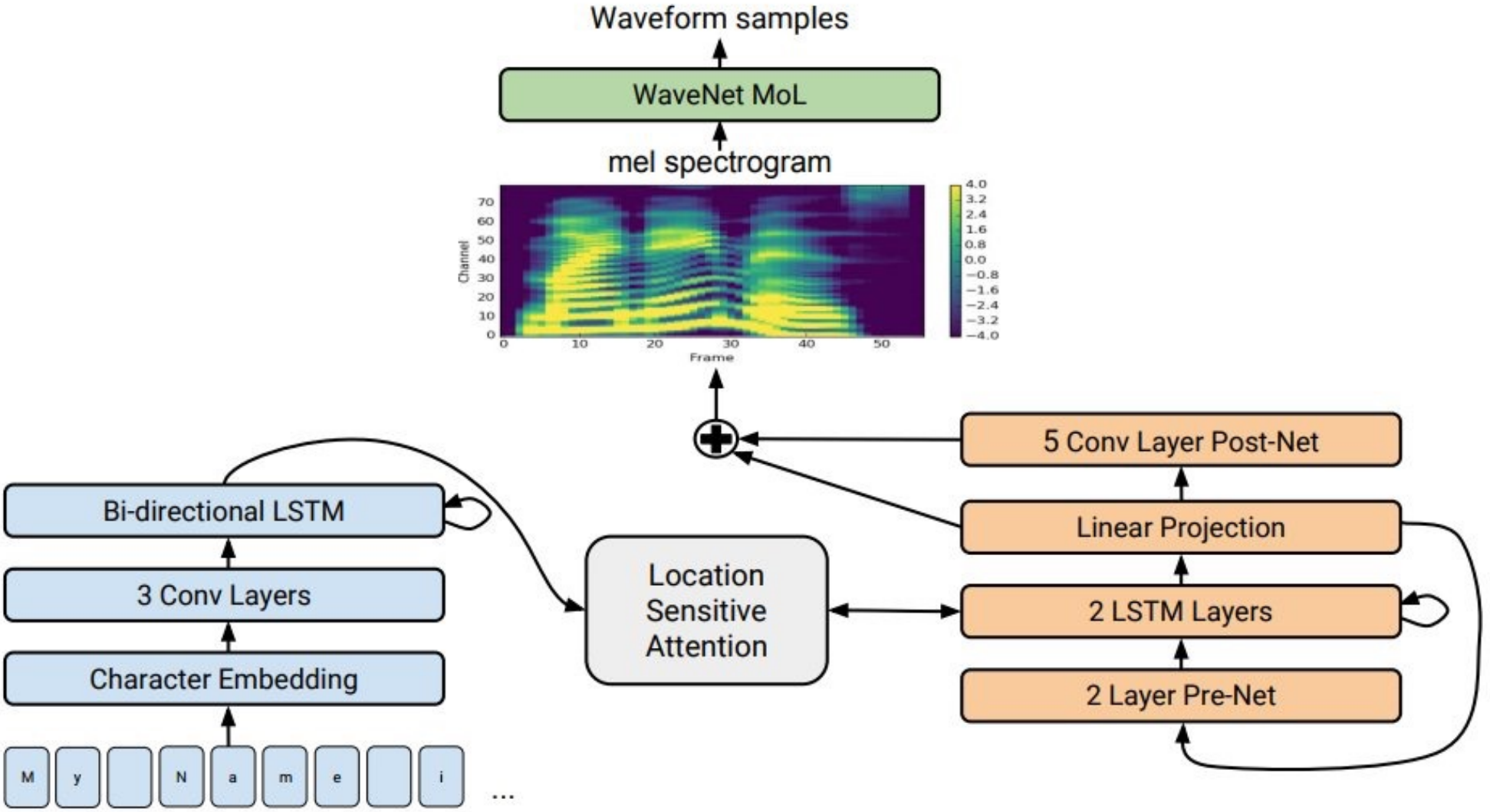
語音辨識(3/3): 考慮字+前字

可把spectrogram/MFCC當作圖片用CNN處理



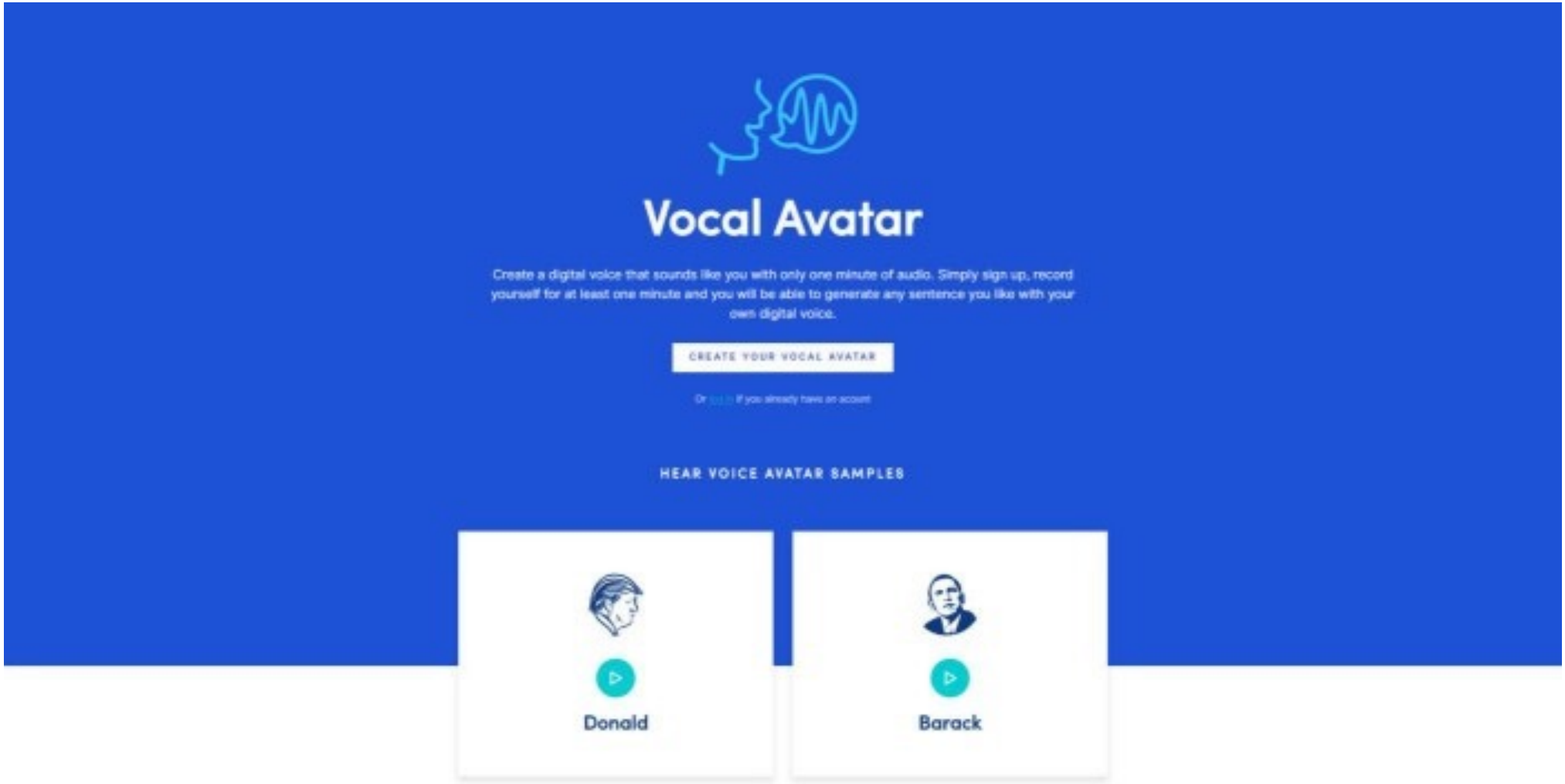
語音合成(1/2)

就是speech to text反過來的text to speech



語音合成(2/2)

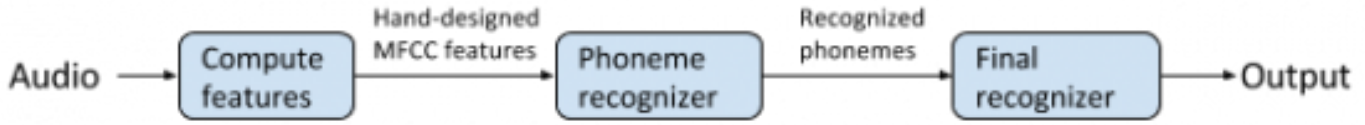
套入個人聲音特徵即speaker normalization的相反



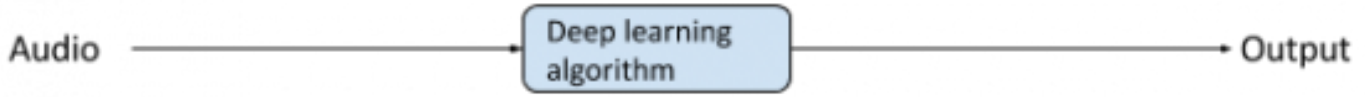
Why big data? (1/2)

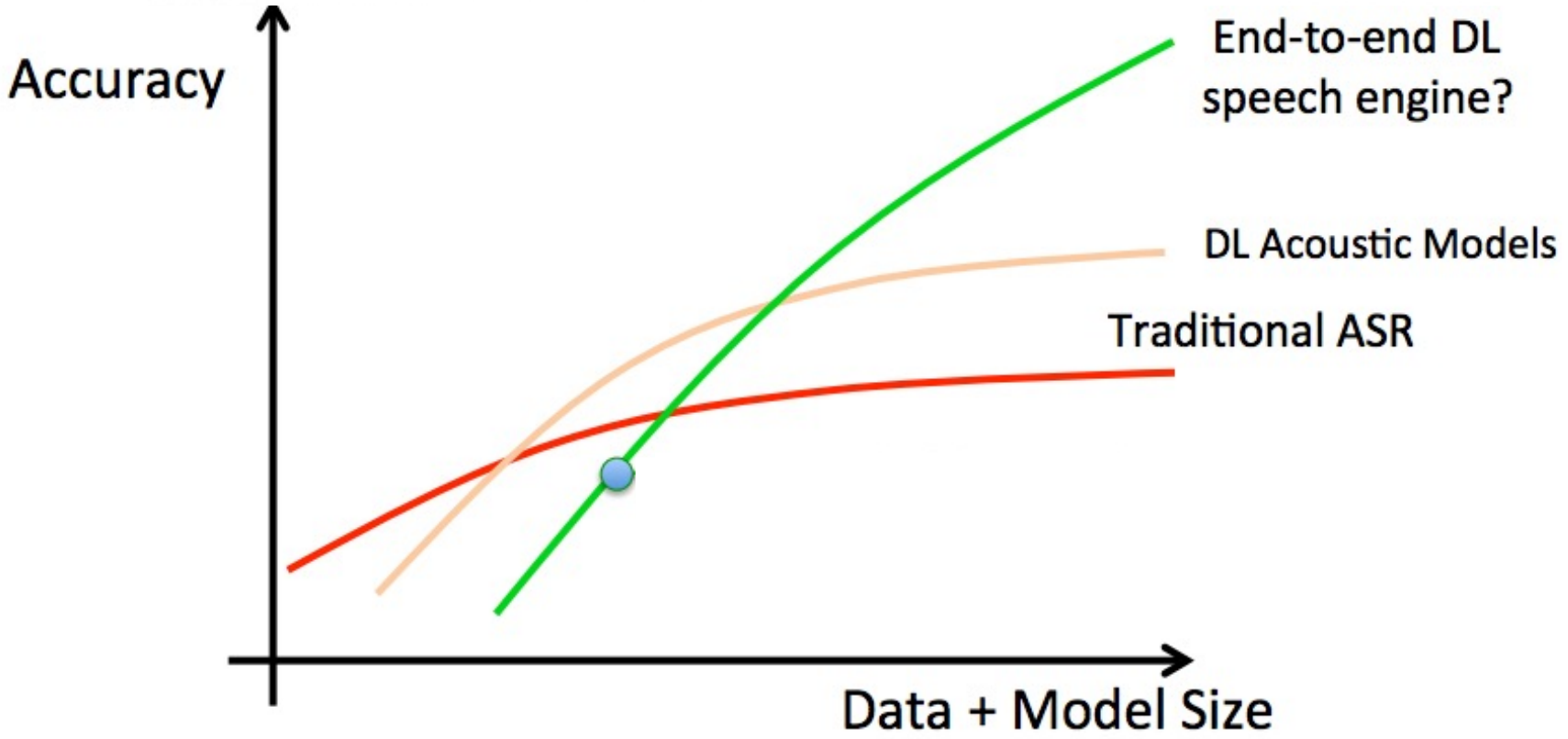
Speech recognition

Traditional model:



End-to-end learning:

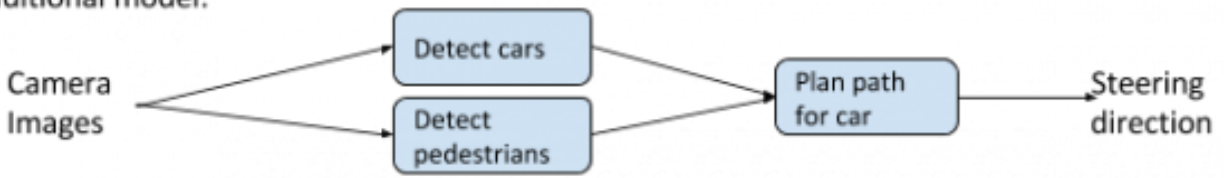




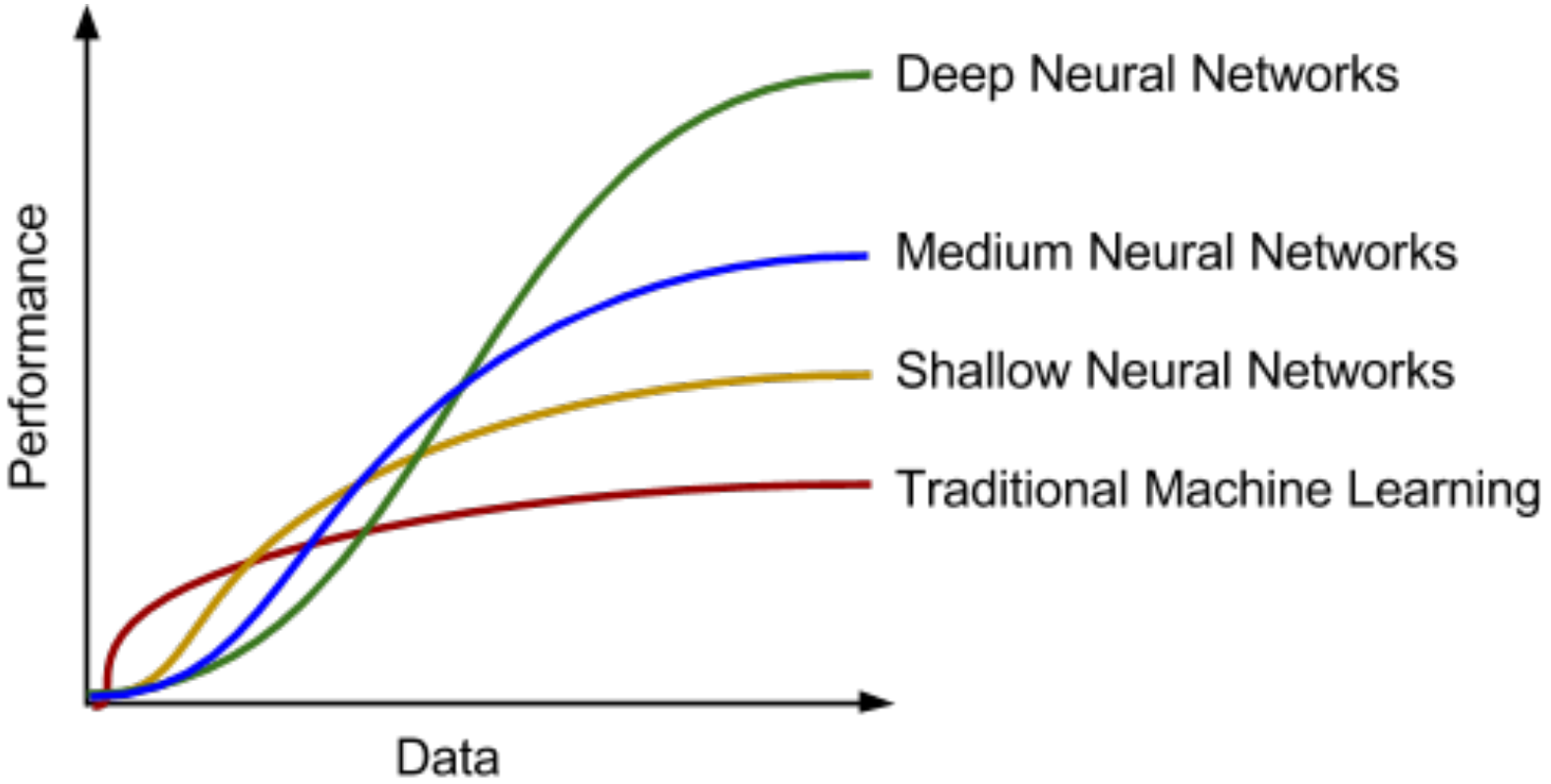
Why big data? (2/2)

Autonomous driving

Traditional model:

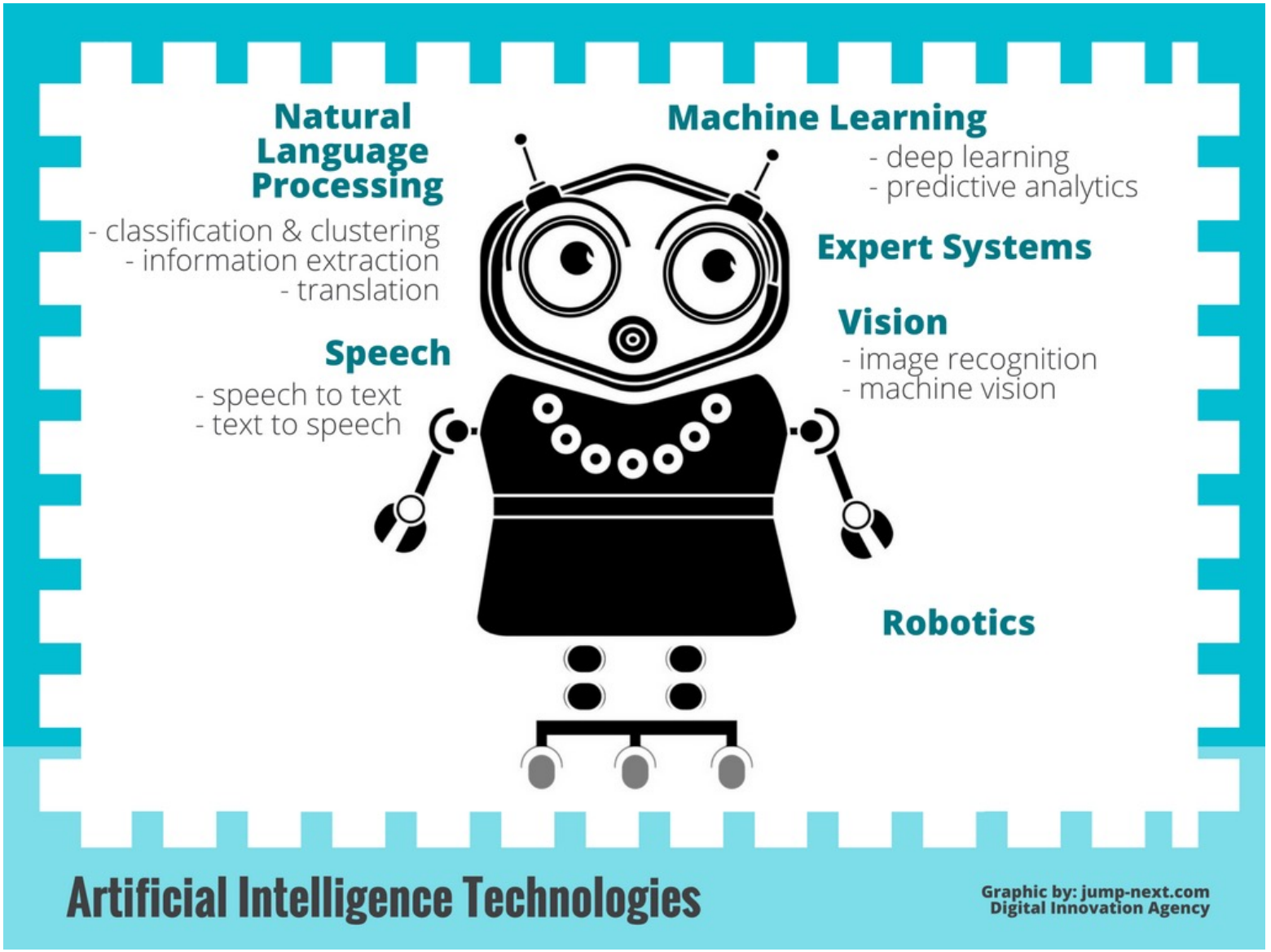


End-to-end learning:



非結構化資訊:影像+聲音+文字

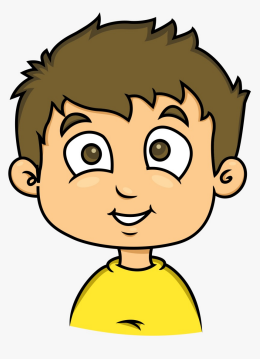
From Data Science to Artificial Intelligence

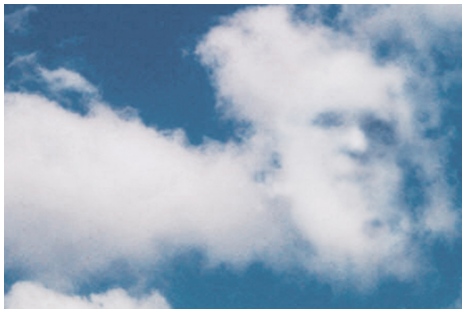


本週作業

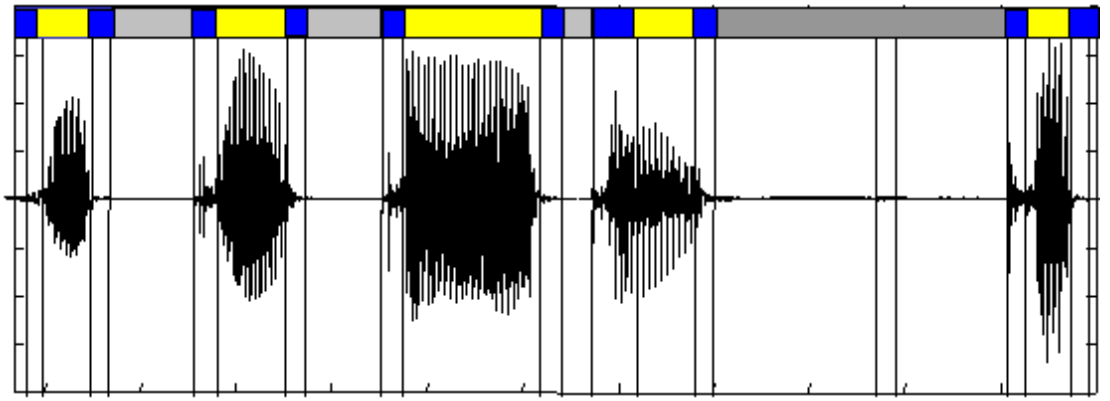
進一步研究影像與音訊處理

1.能偵測到這兩張臉嗎? Why or why not?





2.每個檔案幾個語音片段?總長多久?



Game Over