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

課號：Psy5261　　教室： 博雅 101

識別碼：227U9340　　時間： 四 234


本週處理聲音資料
!

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？

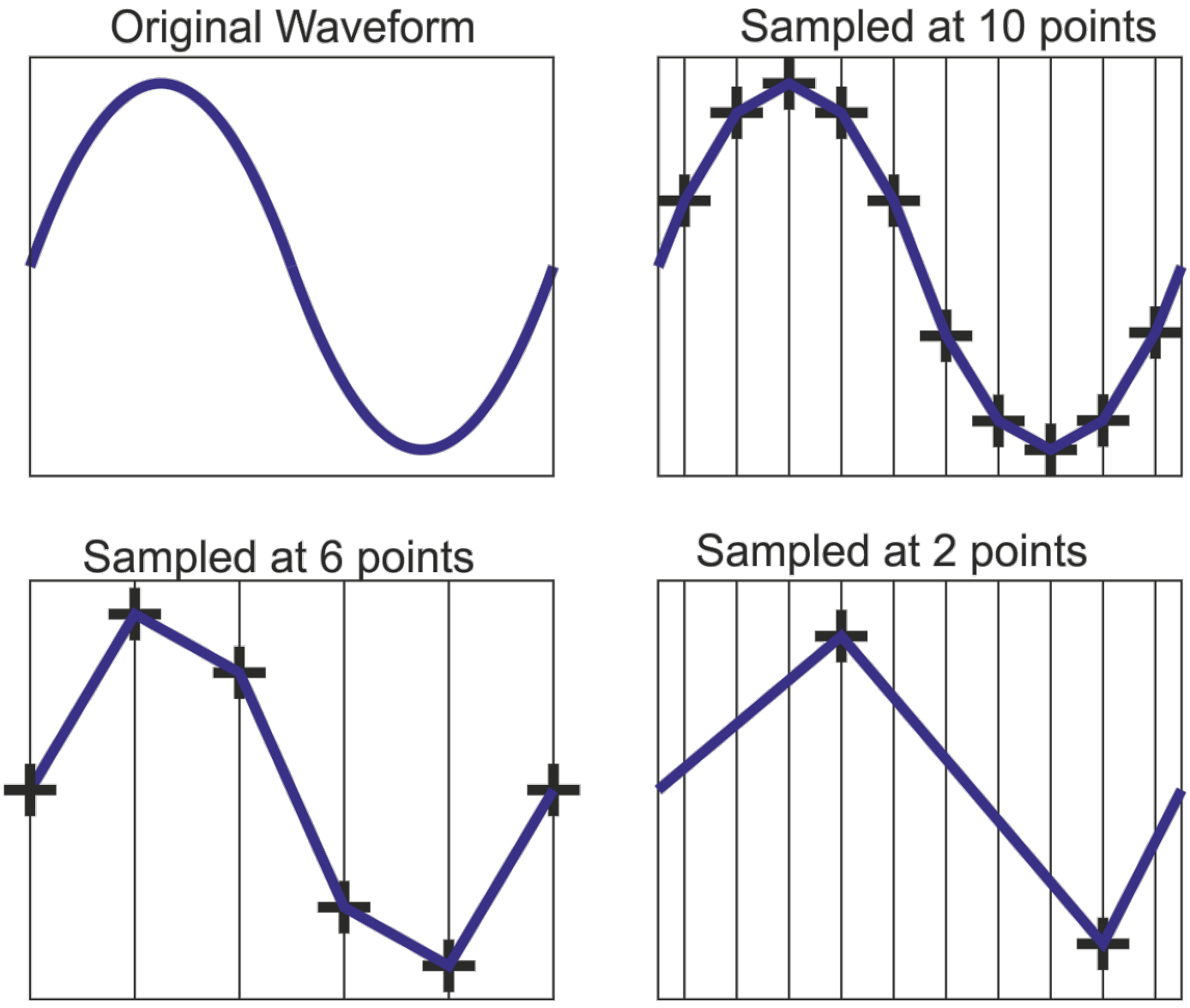
心情 · 12月14日 01:21

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



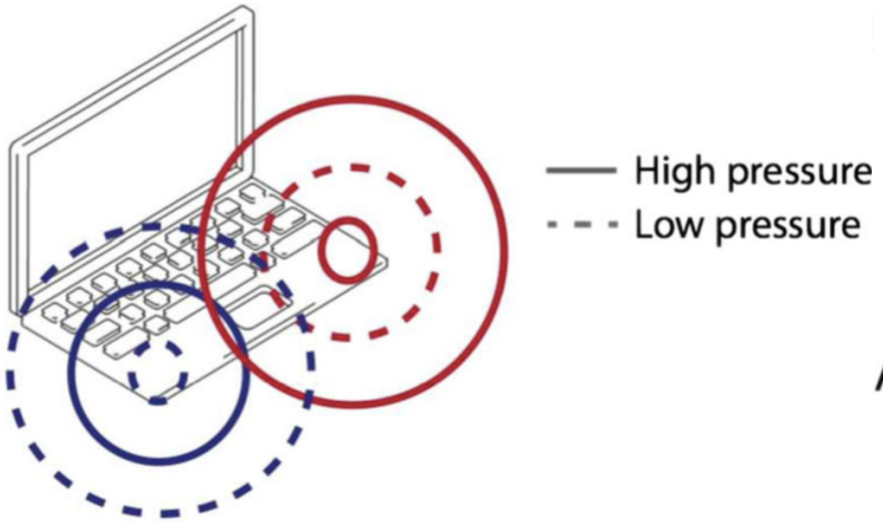
數位訊號處理 (2/3)

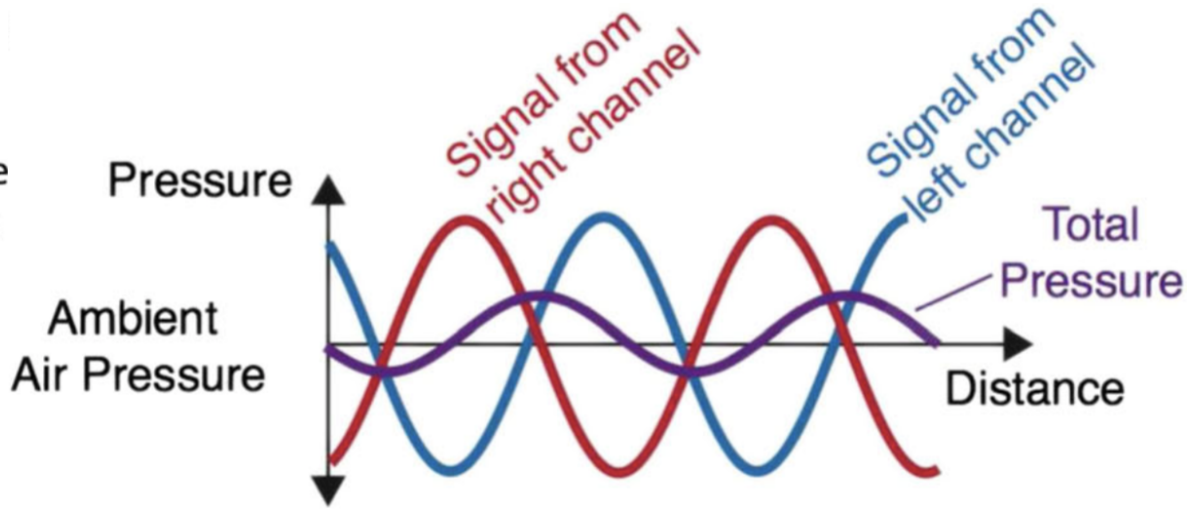
讓雙聲道的訊號從喇叭放出來可以相互抵銷

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

Headphone screening to facilitate web-based auditory experiments

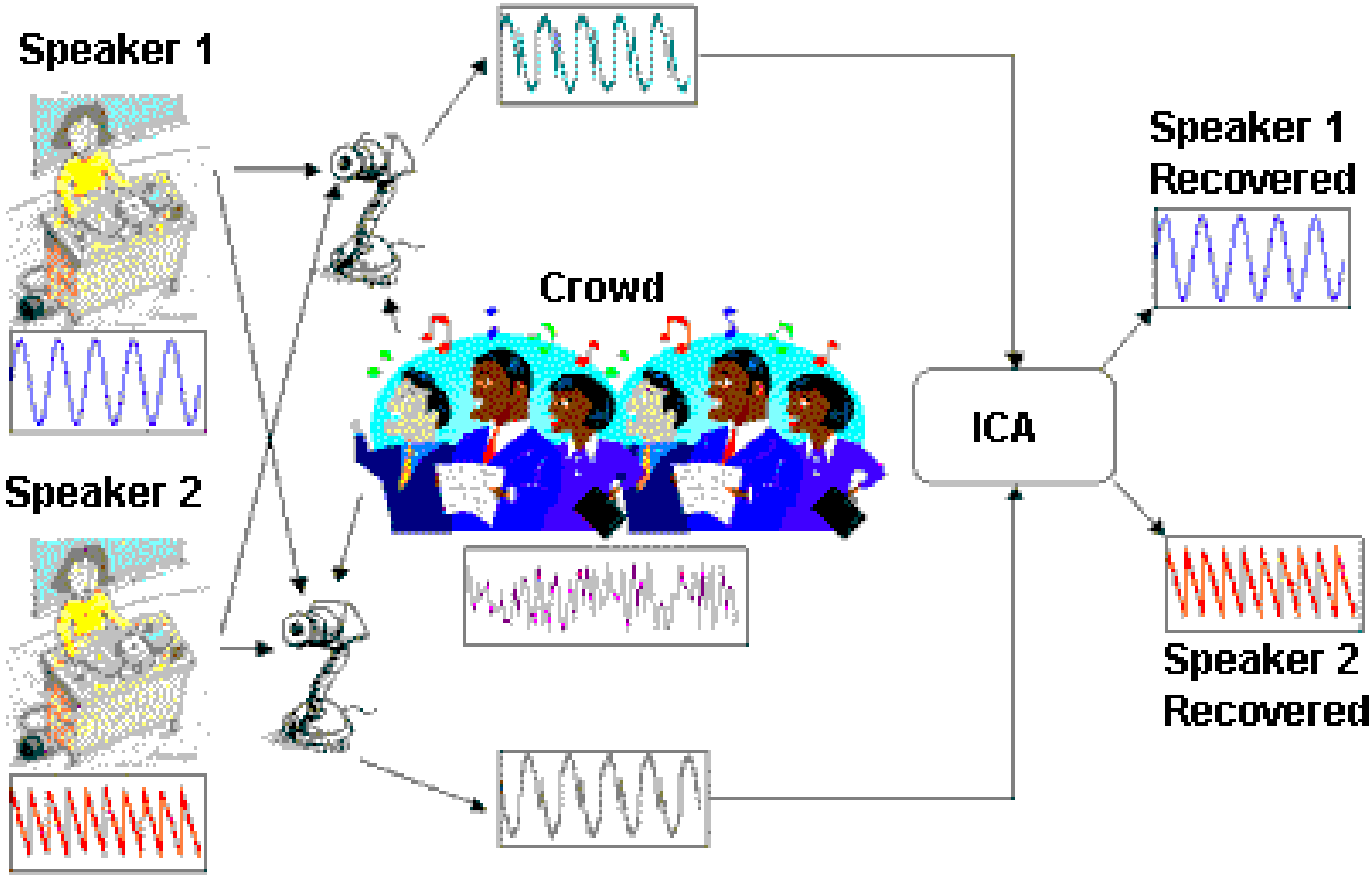
Kevin J. P. Woods[1,2] · Max H. Siegel[1] · James Traer[1] · Josh H. McDermott[1,2]



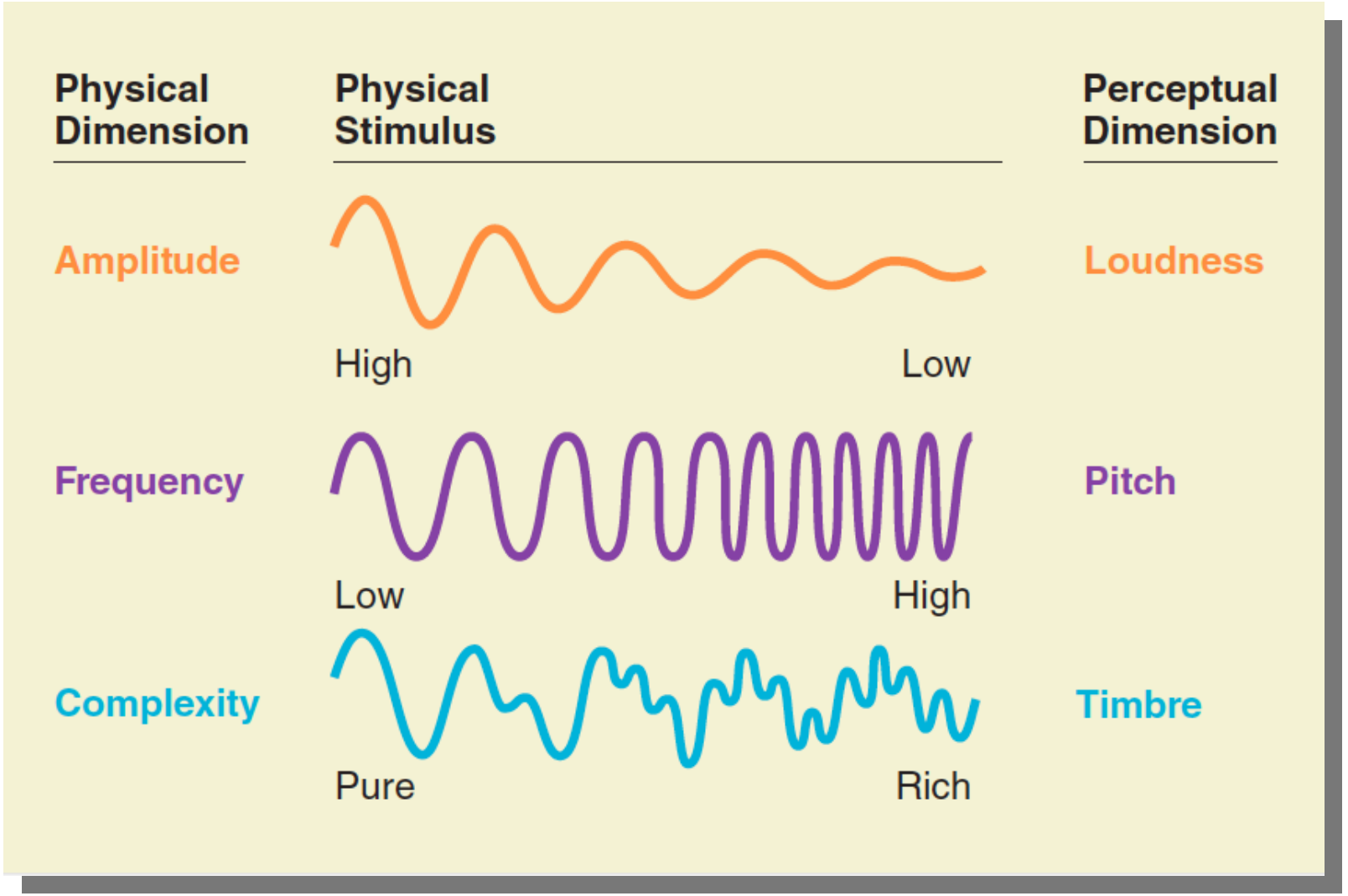


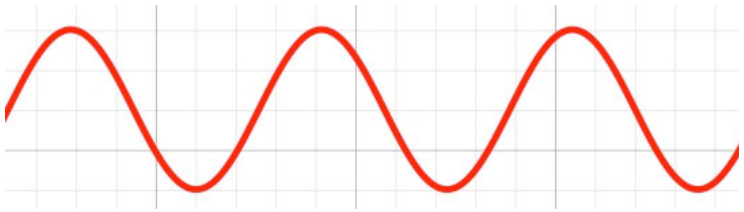
數位訊號處理 (3/3)

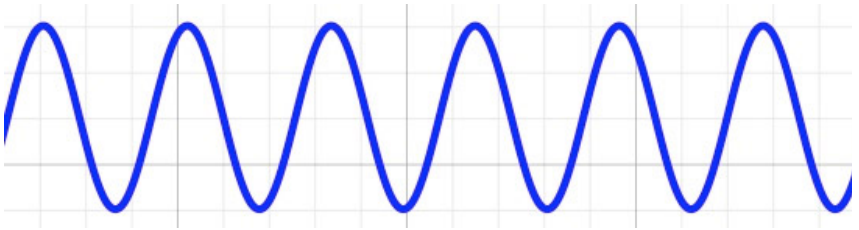
例如 ICA可以把不同的人聲分隔



聲音的特徵

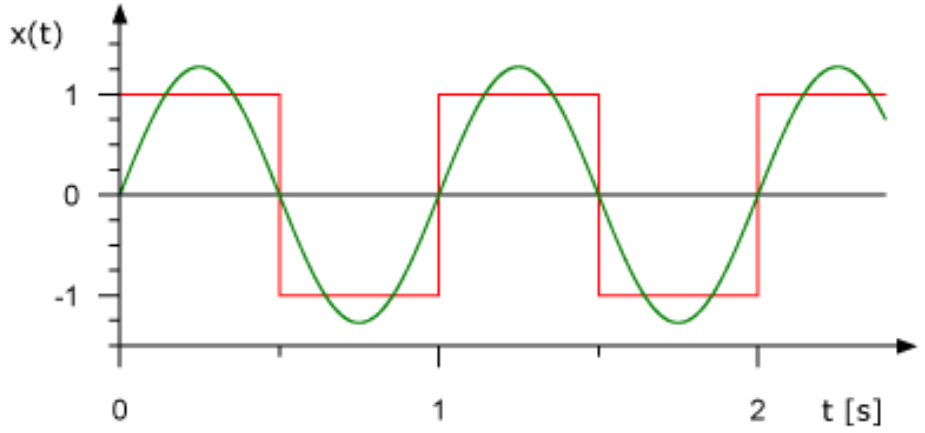

Physical Dimension
Physical Stimulus
Perceptual Dimension
Amplitude
Loudness
High
Low
Frequency
Pitch
Low
High
Complexity
Timbre
Pure
Rich

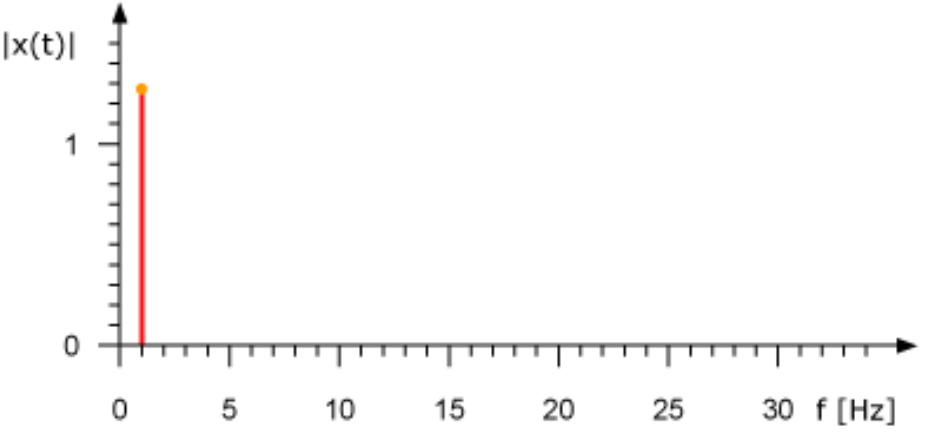




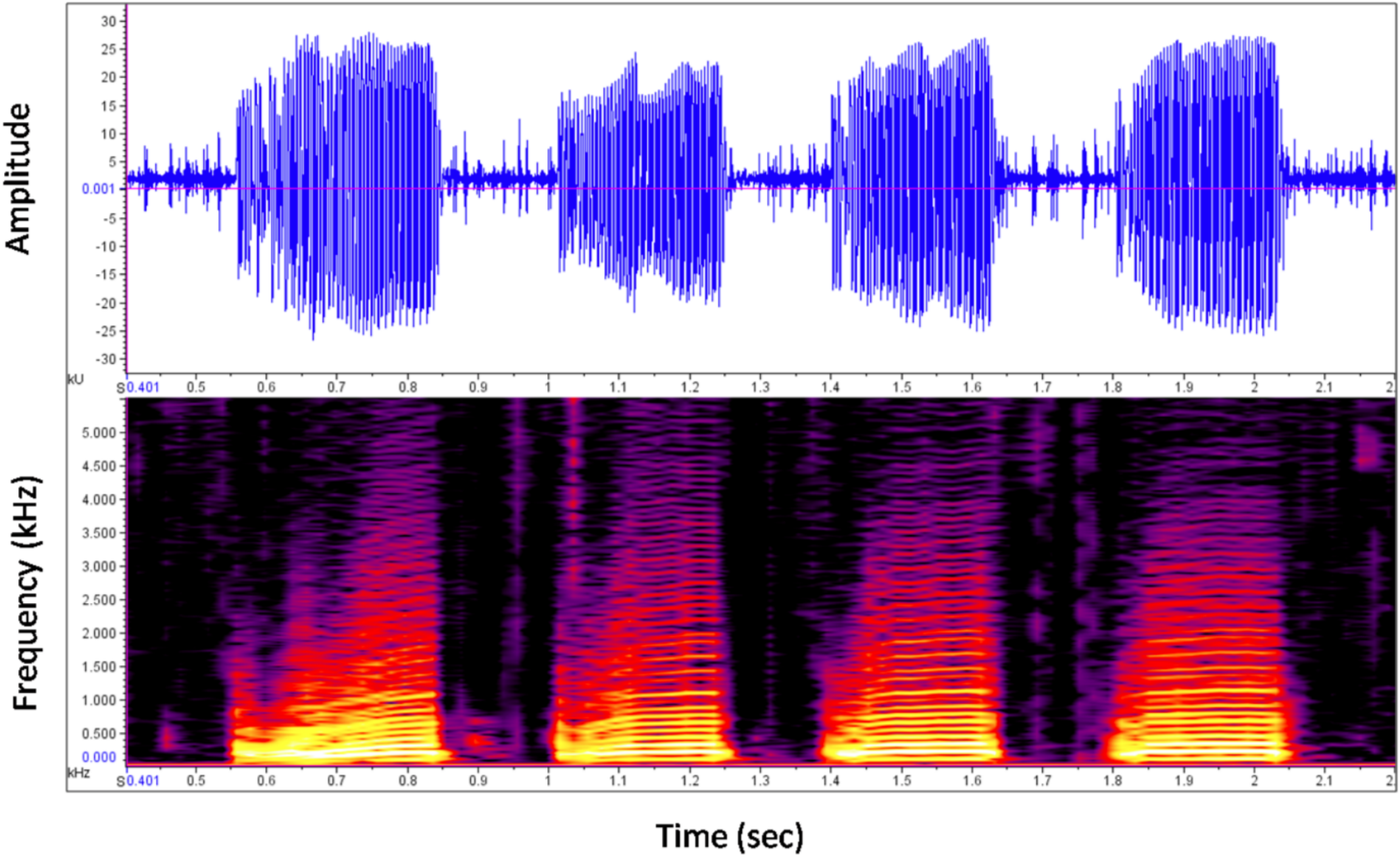
傅立葉分析：Frequency Domain

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

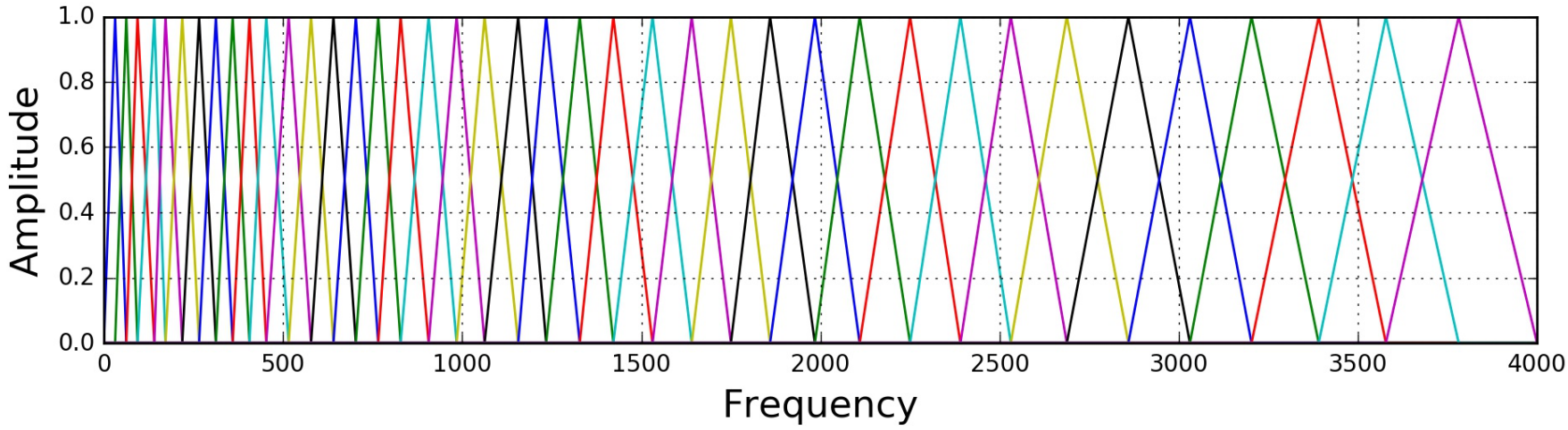


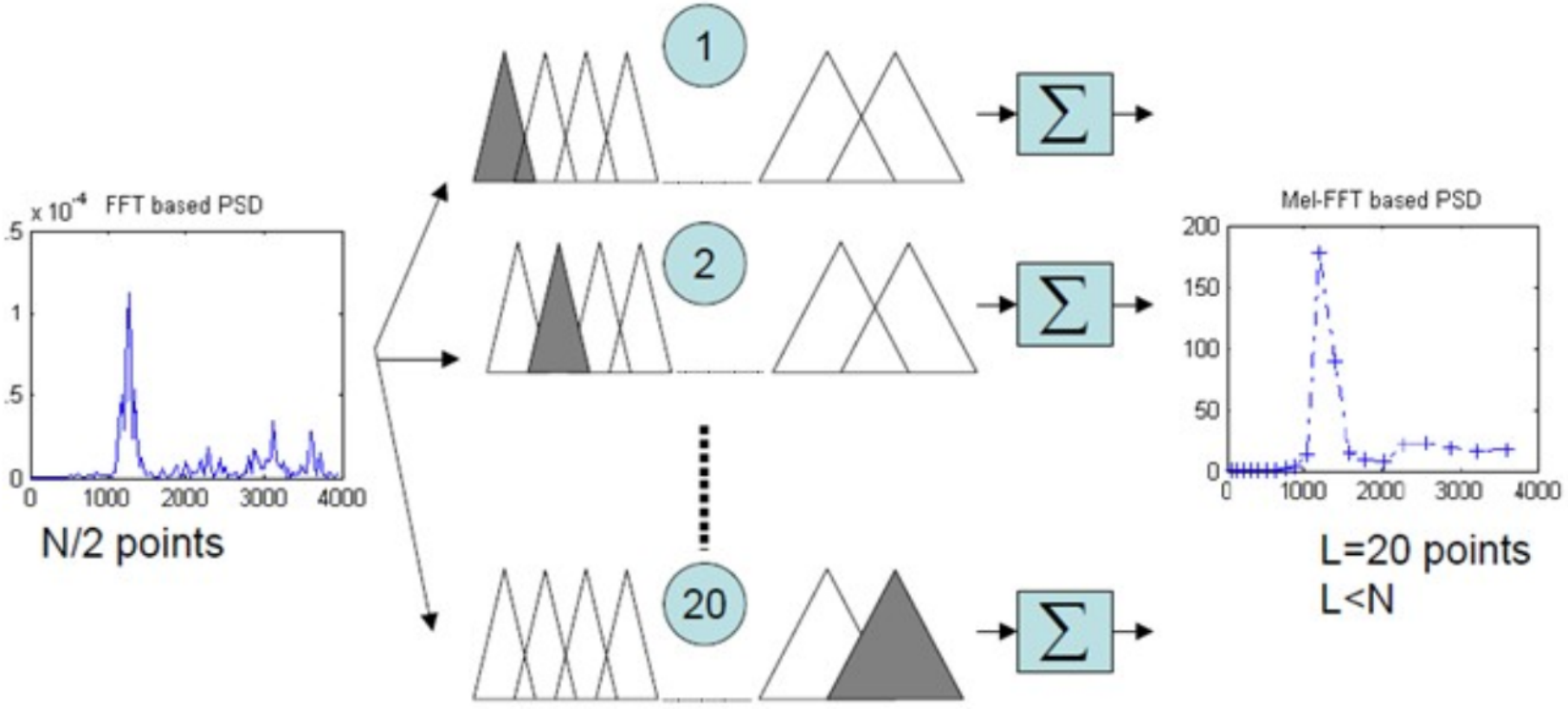


時頻譜 (Spectrogram)



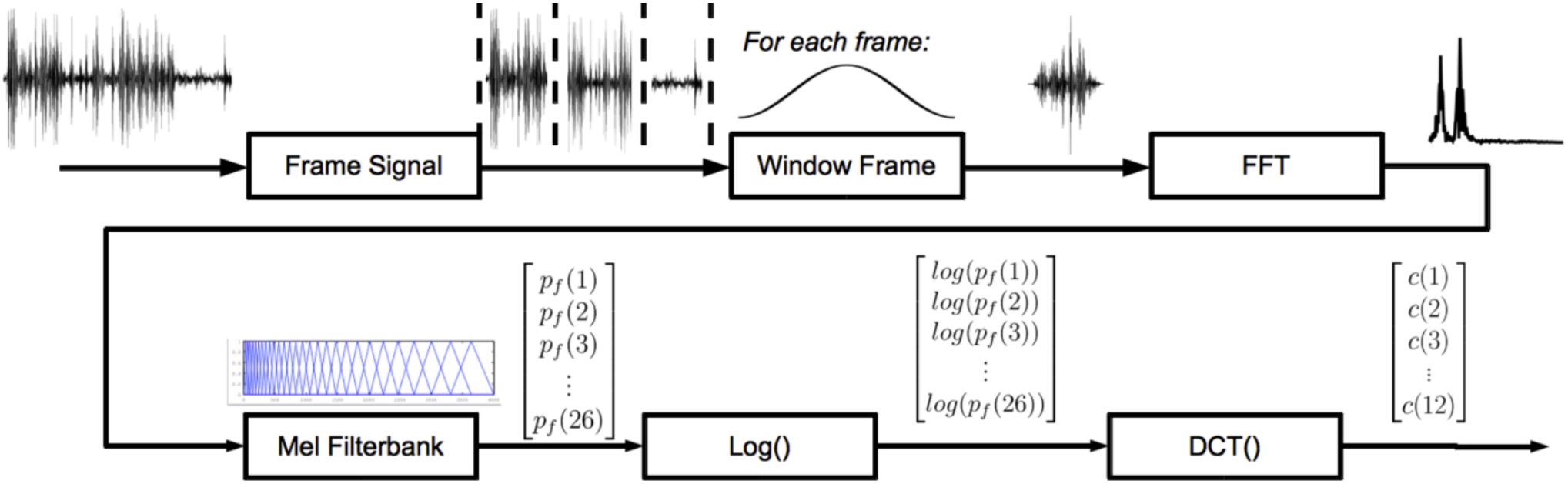
Mel Filters: 聽不到的頻率就少採樣





梅爾倒頻譜 (MFCC)

把 filter coefficients 再透過 DCT做進一步壓縮

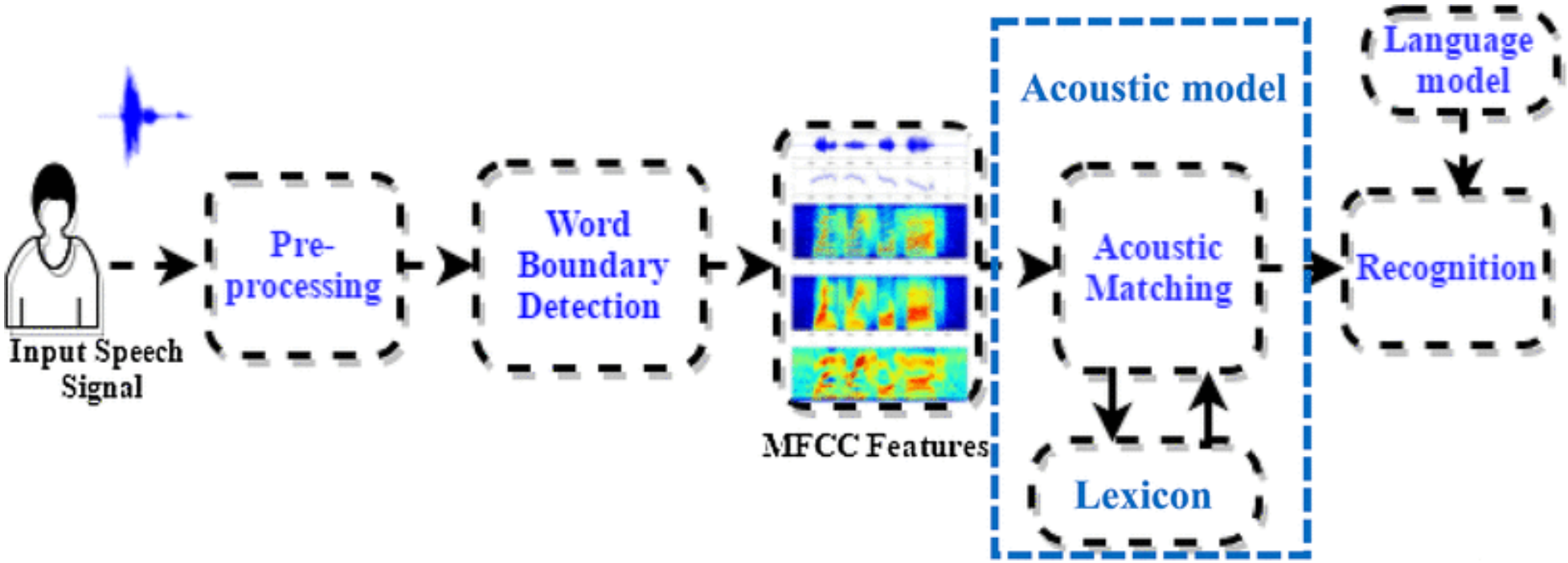


1976 就提出的 MFCC 雖經典但仍是很有效的特徵

語音資料處理
(Speech Processing)

語音辨識 (1/3)

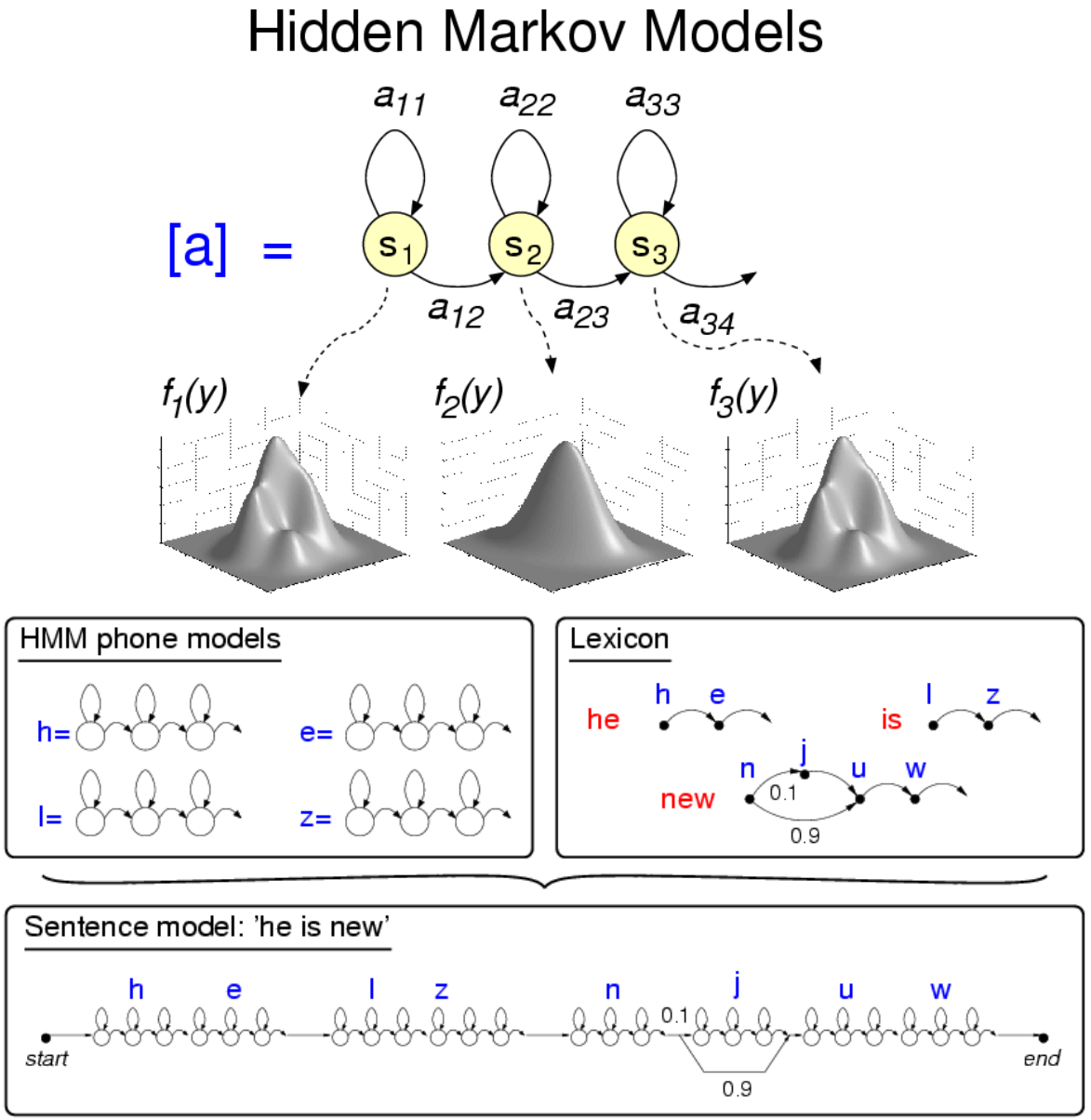
最簡單的想法是用 bottom-up 訊號比對字的頻譜特徵



若有 language model 可提供 top-down 協助

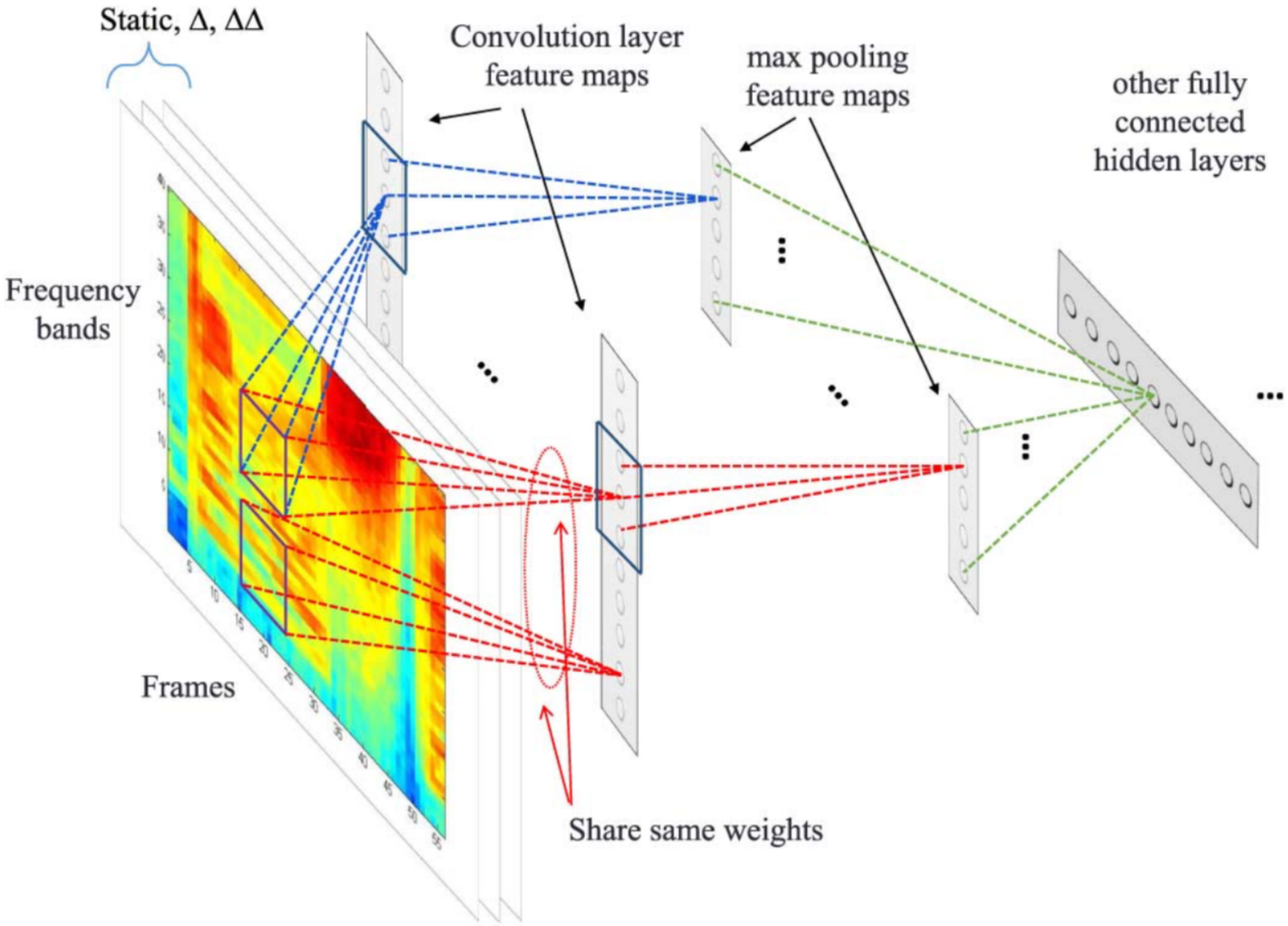
語音辨識 (2/3)

HMM 可 model 音素→字元、字元→單字、單字→片語



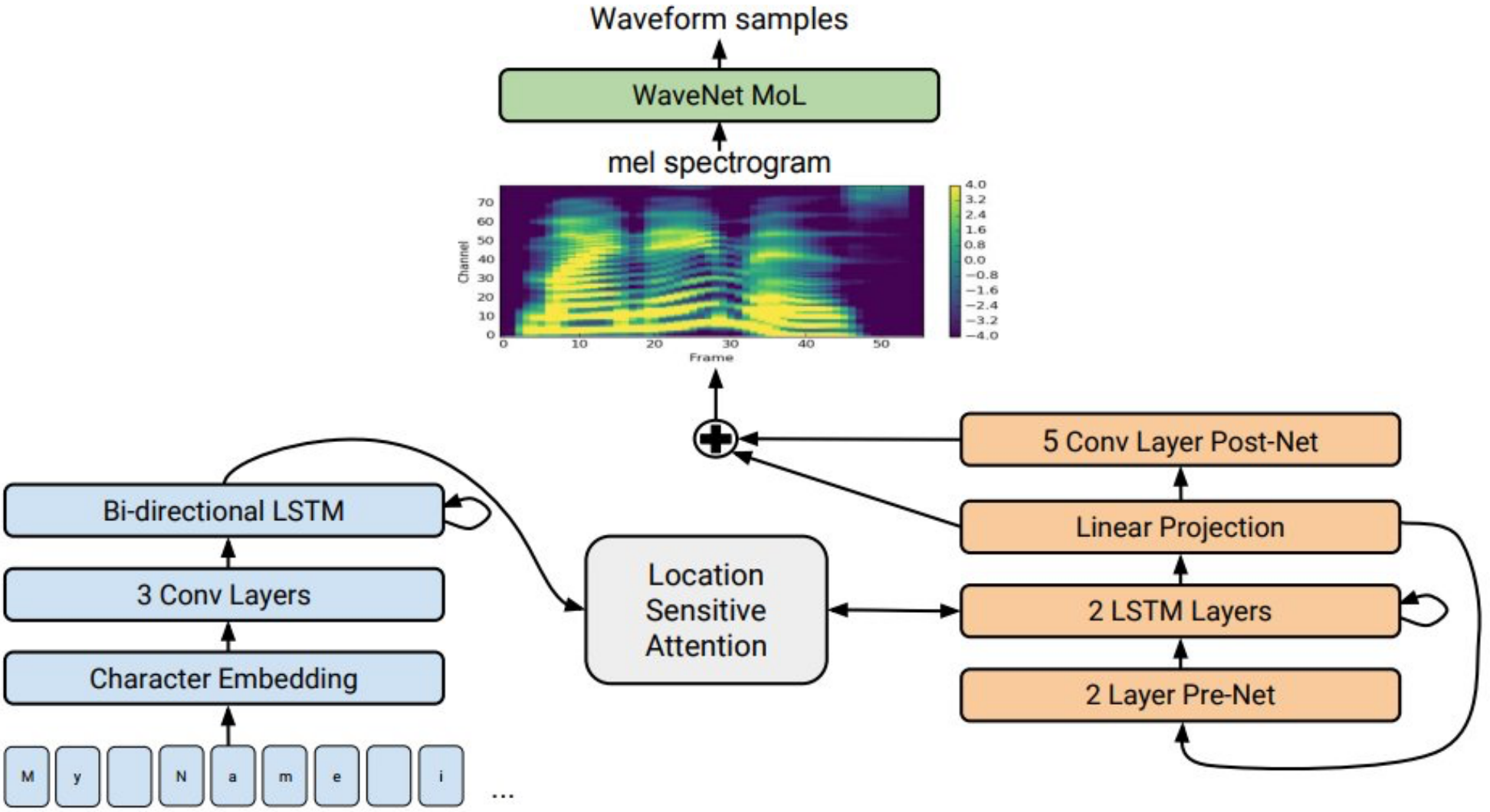
語音辨識 (3/3)

可把 spectrogram/MFCC 當作圖片用 CNN 處理



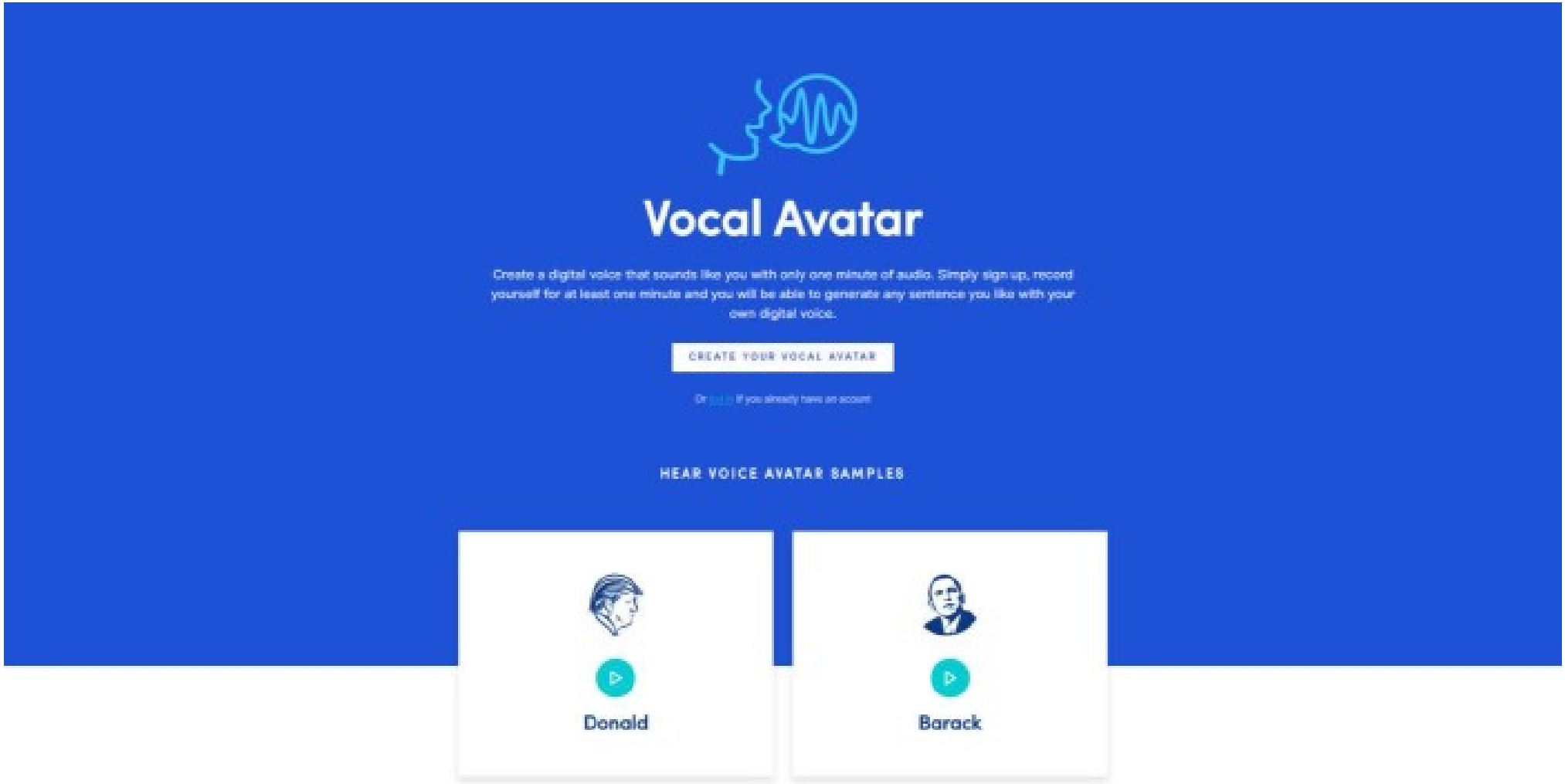
語音合成 (1/2)

就是 speech to text 反過來的 text to speech



語音合成 (2/2)

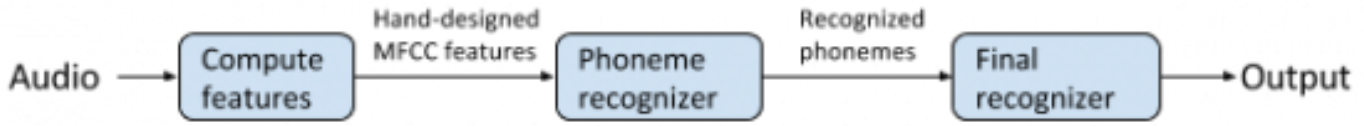
套入個人聲音特徵即 speaker normalization的相反



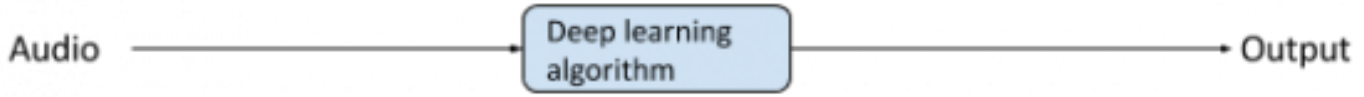
Why big data? (1/2)

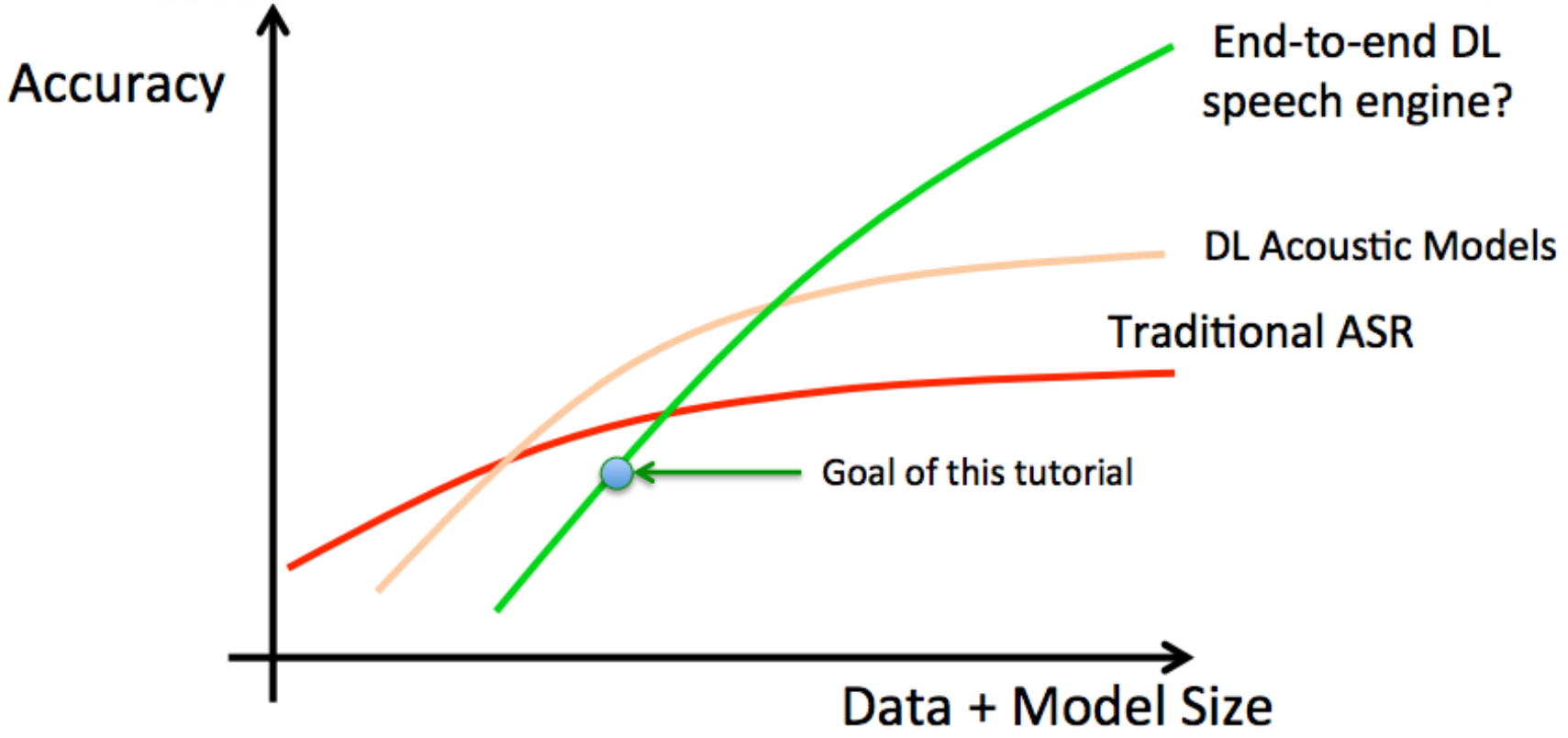
Speech recognition

Traditional model:



End-to-end learning:

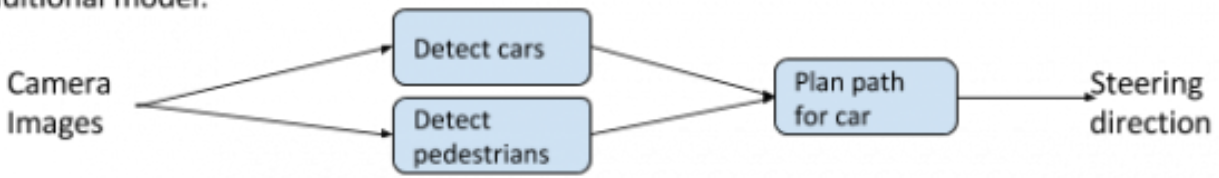




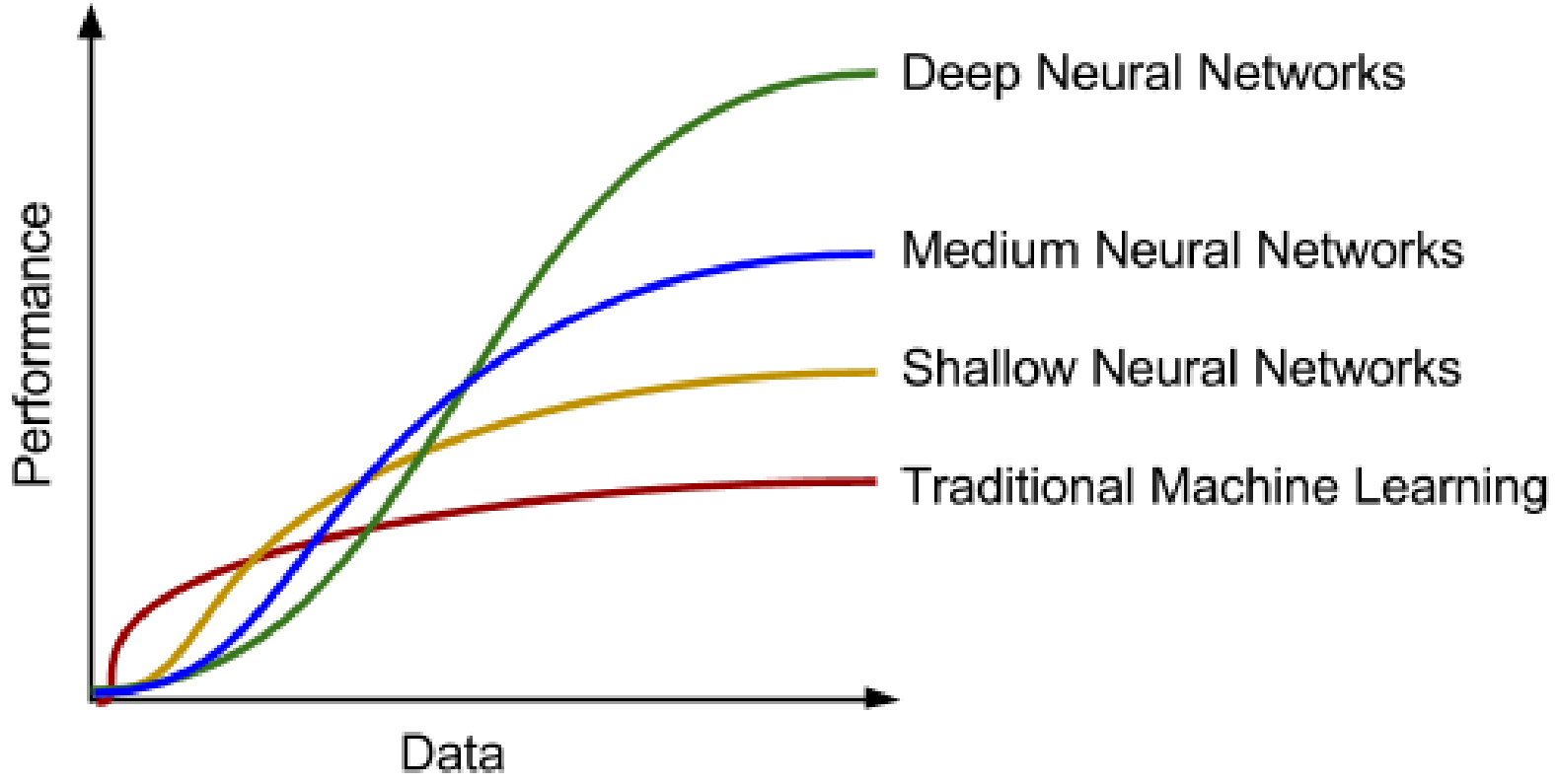
Why big data? (2/2)

Autonomous driving

Traditional model:

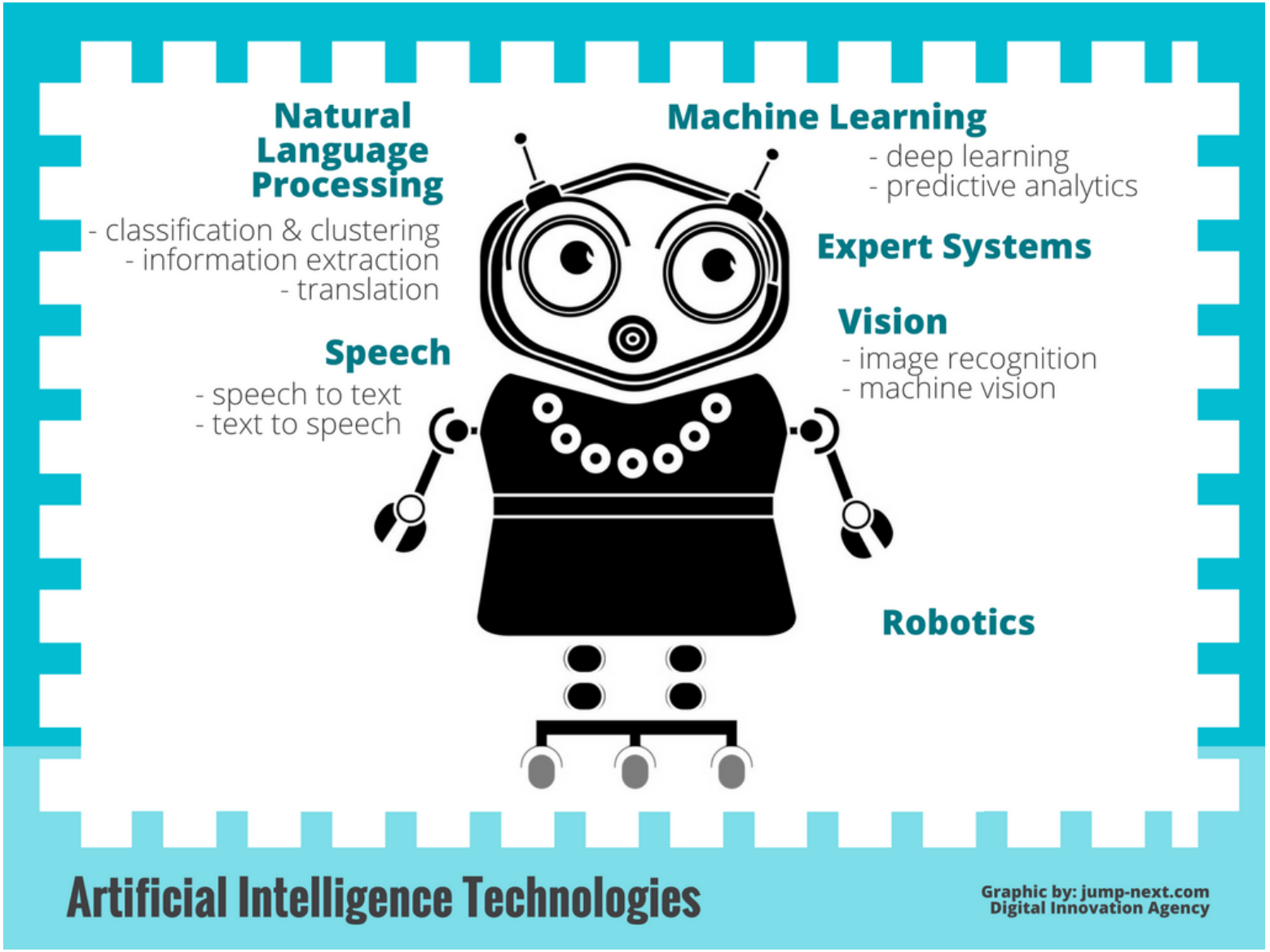


End-to-end learning:



非結構化資訊：影像 + 聲音 + 文字

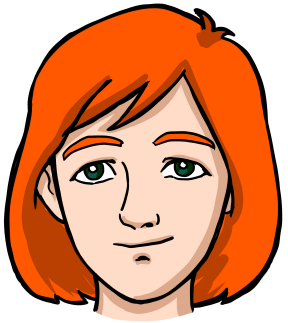
From Data Science to Artificial Intelligence

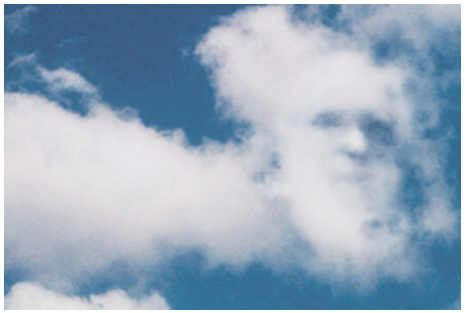


本週作業

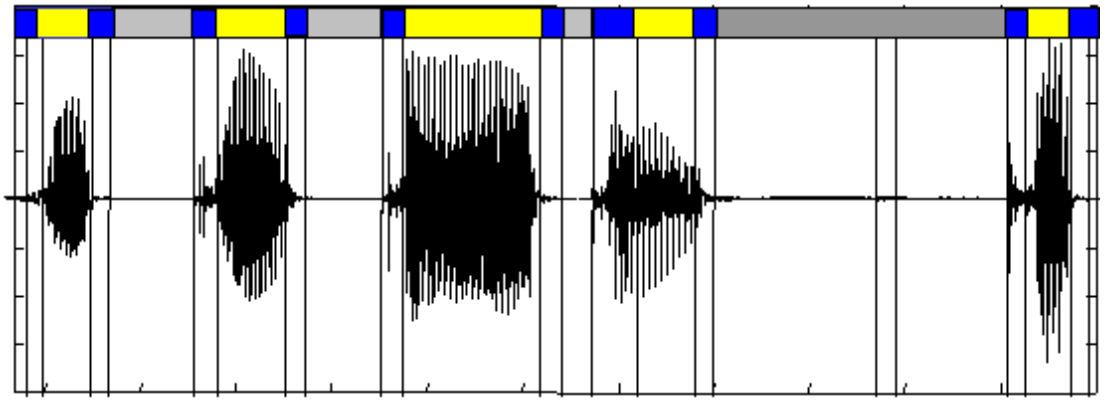
進一步研究影像與音訊處理

1. 能偵測到這兩張臉嗎？Why or why not?





2. 幾個片段？總長多久？



Game Over