音訊處理與其應用

王坤卿 副教授兼教學與發展二中心主任
實踐大學 資訊科技與通訊學系

日期： **9/14, 2015**
地點：屏東大學 資工所

What is this?

Siri

Future come true!

- iCAR

Outline

1. 音訊的簡介
2. 語音辨識系統
3. **嵌入式語音系統**
4. Demo
5. 結論

Outline

1. 音訊的簡介
2. 語音辨識系統
3. 嵌入式語音系統
4. Demo
5. 結論

音訊分類

- 音訊可以有很多不同的分類方式，例如，若以**發音的來源**，可以大概分類如下：
  - 生物音: 人聲 **(語音, human voice)**、狗聲、貓聲等。
  - 非生物音：引擎聲、關門聲、打雷聲、樂器聲等。

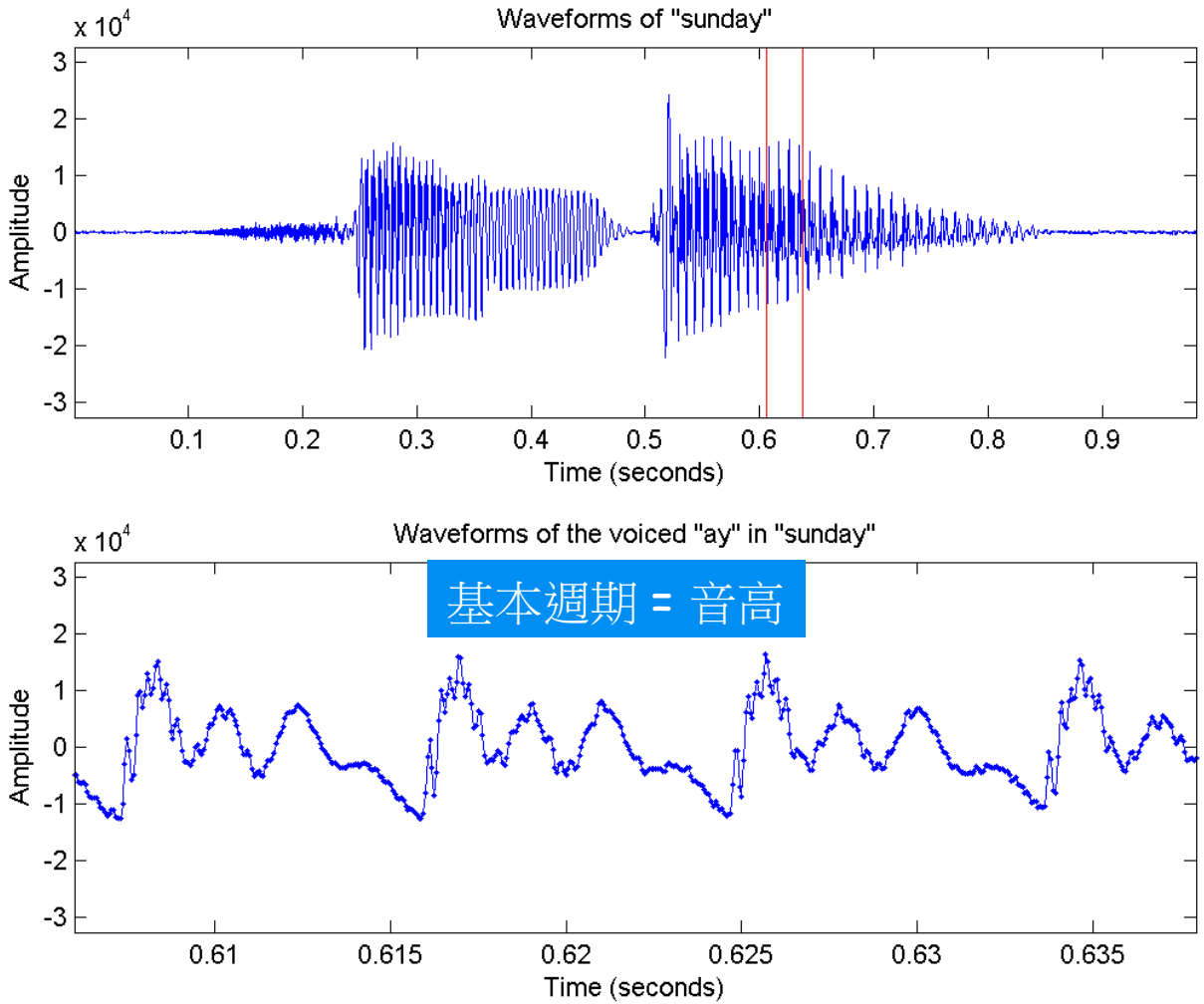
訊號的規律性

- 若以**訊號的規律性**，又可以分類如下：
  - **準週期音**：波形具有規律性，可以看出週期的重複性，人耳可以感覺其穩定**音高**的存在，例如單音絃樂器、人聲清唱等。
  - **非週期音**：波形不具規律性，看不出明顯的週期，人耳無法感覺出穩定音高的存在，例如打雷聲、拍手聲、敲鑼打鼓聲、人聲中的氣音等。

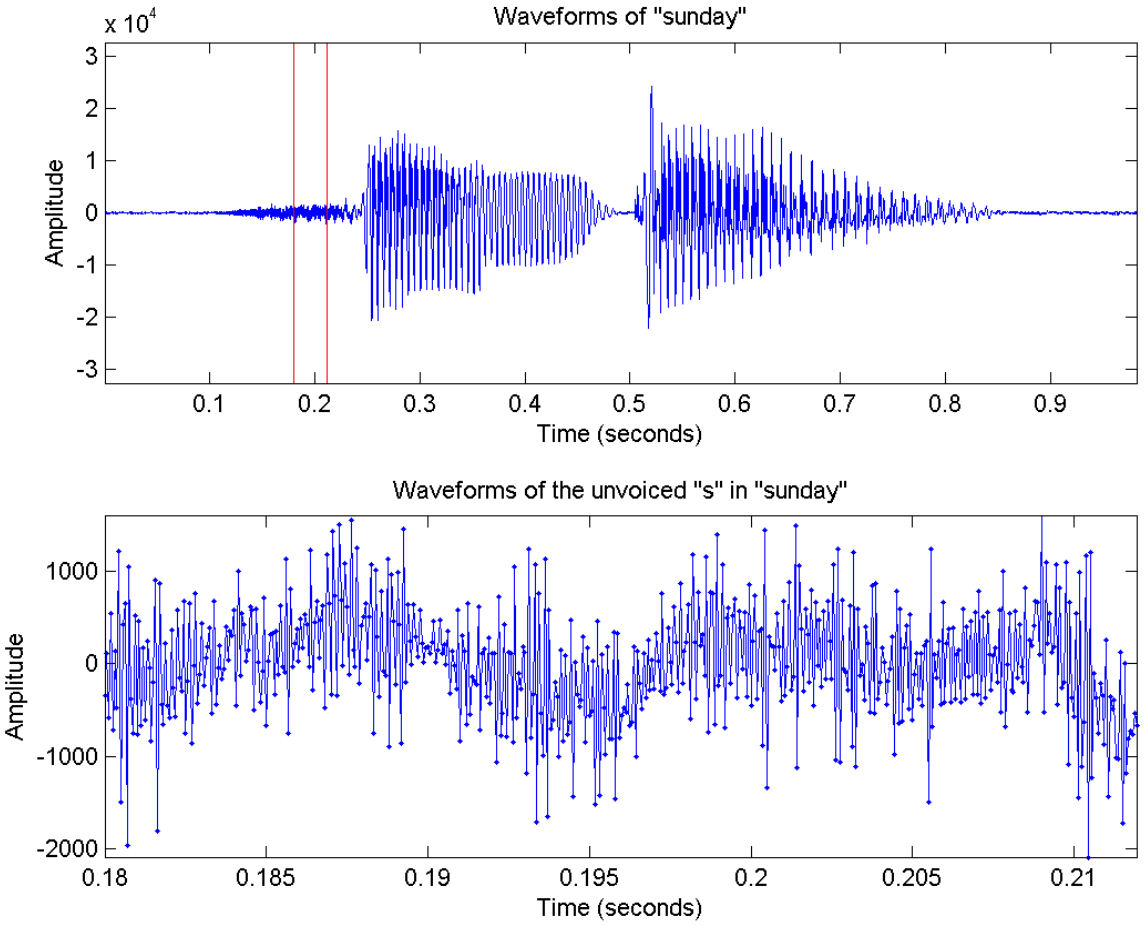
Voiced sound vs. Unvoiced sound

- 以人聲而言，我們可以根據其是否具有**音高(pitch)**而分為兩類，如下：
  - Voiced sound: 由**聲帶振動**所發出的聲音，例如一般的母音等。由於聲帶振動，造成規律性的變化，所以我們可以感覺到音高的存在。
  - Unvoiced sound: 由**嘴唇所發出**的氣音，並不牽涉聲帶的震動。由於波形沒有規律性，所以我們通常無法感受到穩定音高的存在。

Voiced sound



Unvoiced sound



語音特徵介紹

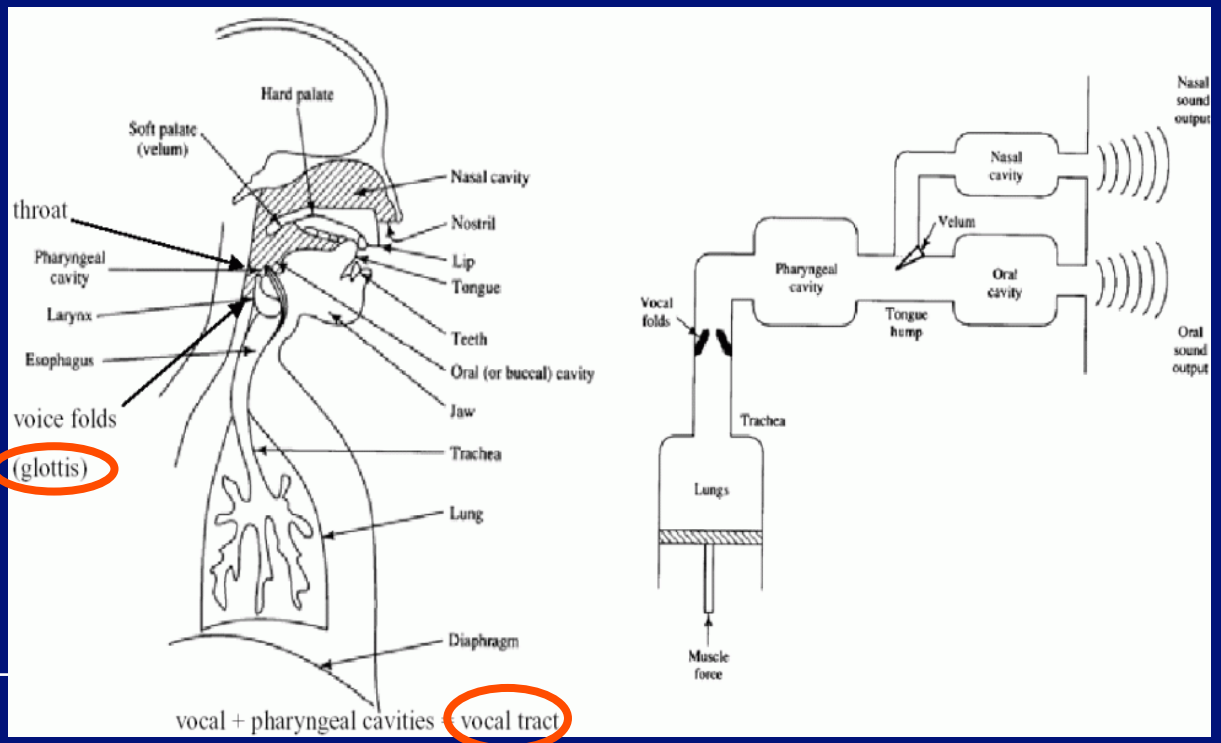
- 在一個特定音框內，我們可以觀察到的三個主要聲音特徵可說明如下：
  - **音量（Volume）**：代表聲音的大小，可由聲音訊號的震幅來類比，又稱為能量（Energy）或強度（Intensity）等。
  - **音高（Pitch）**：代表聲音的高低，可由基本頻率（Fundamental Frequency）來類比，這是**基本週期（Fundamental Period）的倒數**。
  - **音色（Timbre）**：代表聲音的內容（例如英文的母音），可由每一個波形在一個基本週期的變化來類比。

語音特徵的物理意義

- 如果是用人聲來說明，這些語音特徵的物理意義如下：
  - 音量：代表肺部壓縮力量的大小，力量越大，音量越大。
  - 音高：代表聲帶震動的快慢，震動越快，音高會越高。
  - 音色：代表嘴唇和舌頭的位置和形狀，不同的位置和形狀，就會產生不同的語音內容。

人聲的產生(1)

- 人聲的**發音**流程，可以列出如下：
  1. 聲門的快速打開與關閉
  2. 聲道、口腔、鼻腔的共振
  3. 空氣的波動

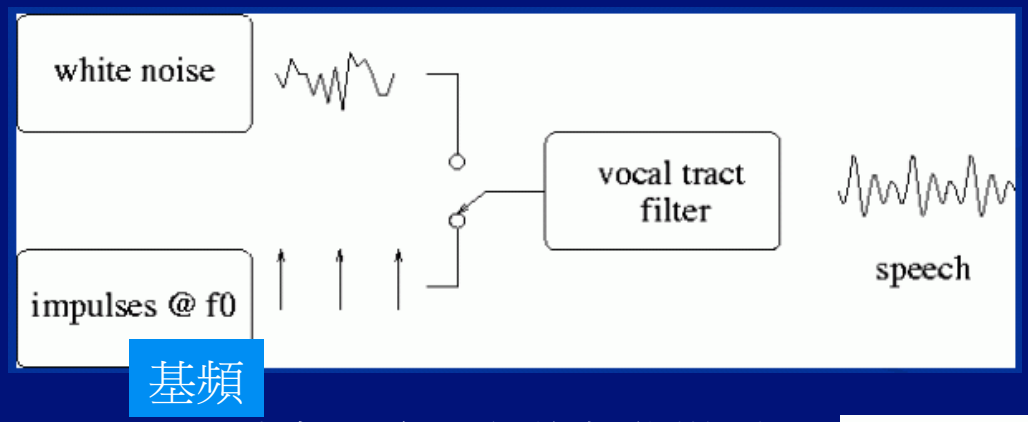


人聲的發音機制

人聲的產生(2)

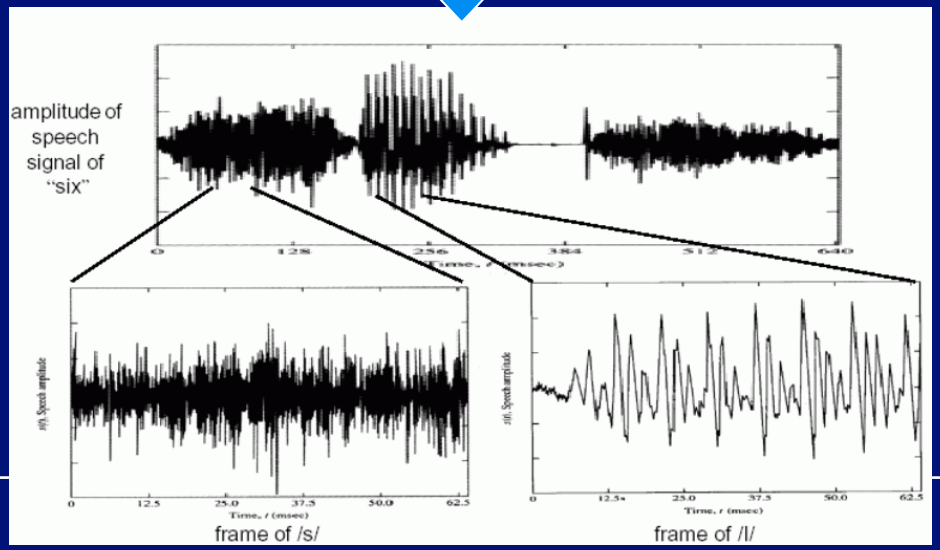
- 由於聲門（Glottis）的肌肉張力，加上由肺部壓迫出來的空氣，就會造成聲門的快速打開與關閉，這個一疏一密的空氣壓力，就是人聲的源頭，在經由**聲道**、**口腔**、**鼻腔**的共振，就會產生不同的聲音（音色）。換句話說：
- 聲門震動的快，決定聲音的基本頻率（即**音高**）。
- 口腔、鼻腔、舌頭的位置、嘴型等，決定聲音的內容（即**音色**）。
- 肺部壓縮空氣的力量大小，決定**音量**大小。

人聲的產生(3)

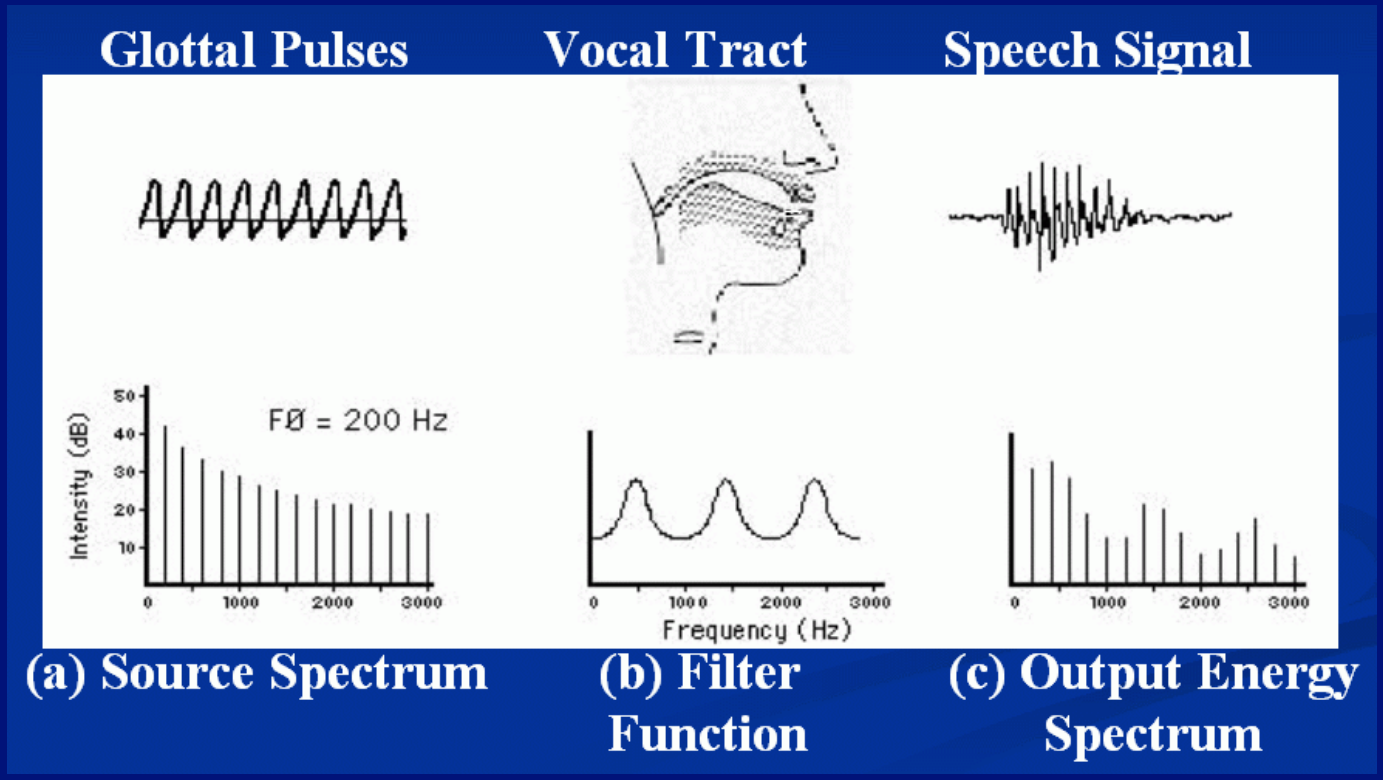


基頻

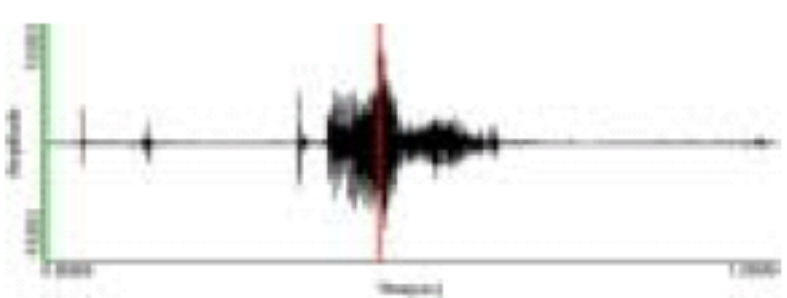
人聲發音過程的數學模型



人聲的產生(4)

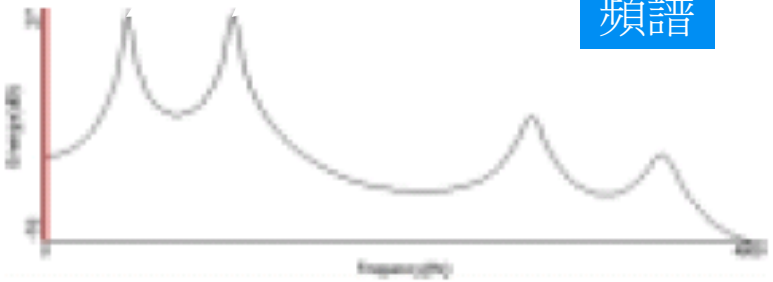

Glottal Pulses
Vocal Tract
Speech Signal
FØ = 200 Hz
Intensity (dB)
50
40
30
20
10
0
1000
2000
3000
Frequency (Hz)
(a) Source Spectrum
(b) Filter Function
(c) Output Energy Spectrum

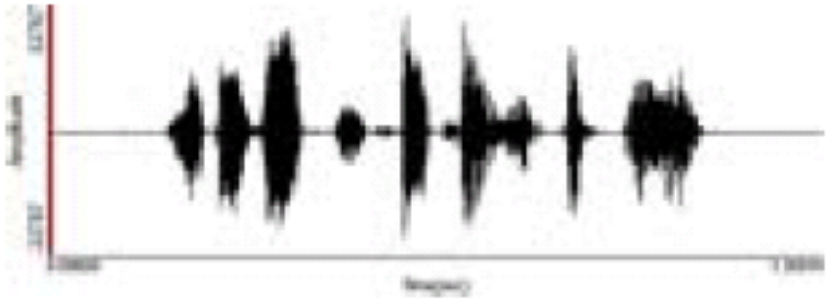
聲音波形、頻譜及聲譜圖



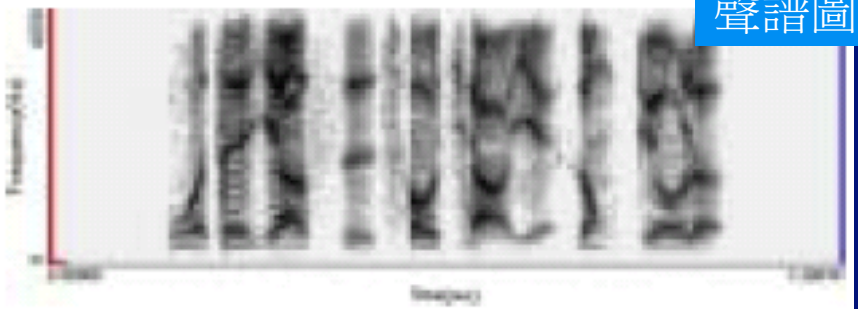
共振峰(formant)

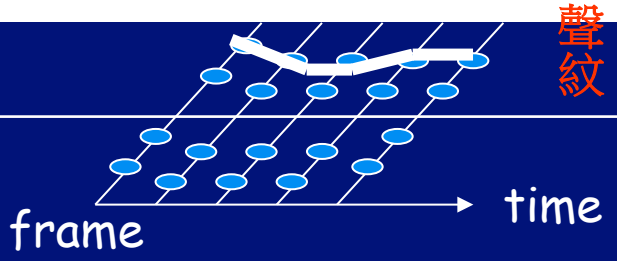
頻譜





聲譜圖





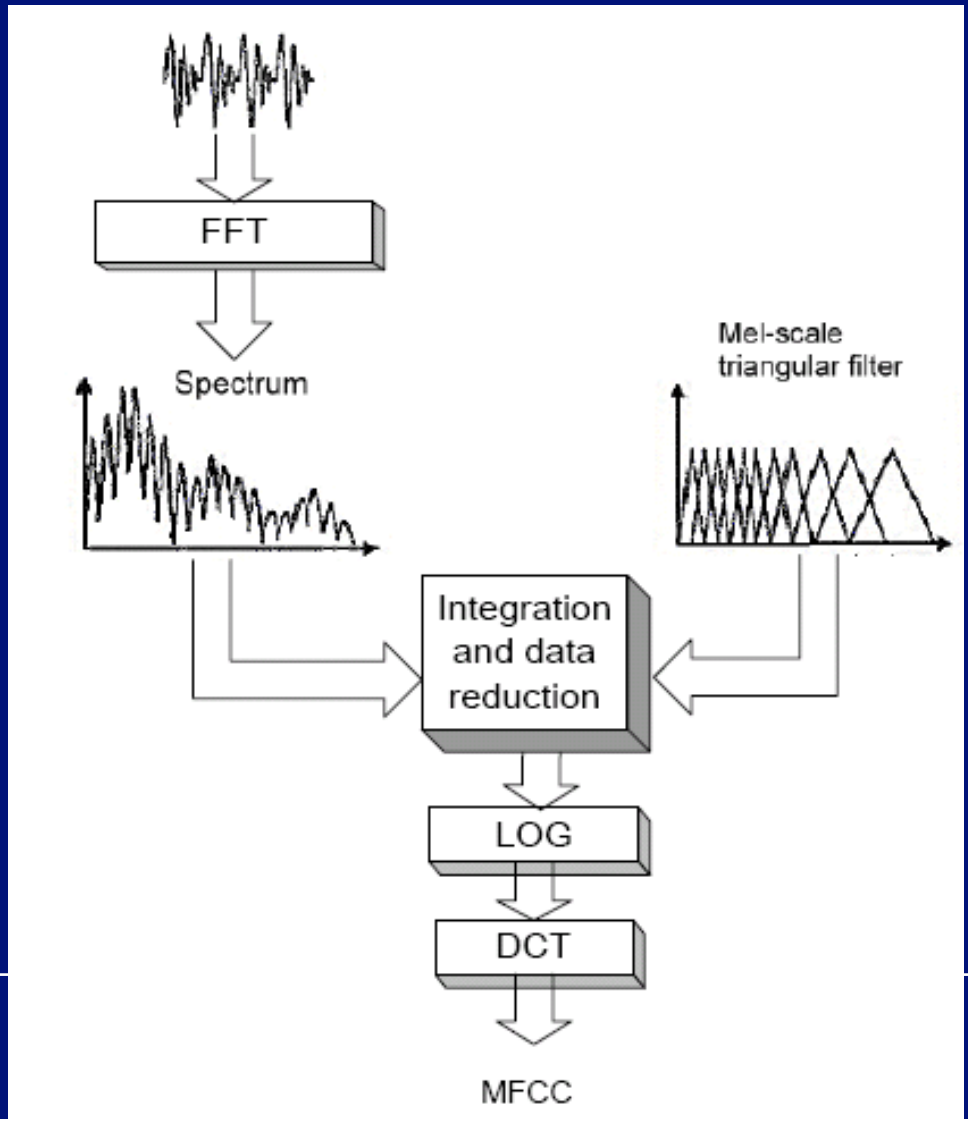
共振峰(formant)

- 在頻域上，能量集中處就是共振峰(formant)之所在，在聲譜圖上就是顏色較深的位置。
- 在發母音(voiced speech)時，音強較大，聲帶振動而呈現出基頻及其諧振頻率，可以明顯看到共振峰，能量集中在低頻。
- 如果是發子音(unvoiced speech)，而且聲帶不振動，就看不到諧振頻率。通常子音的音強小，顏色看來就比較淡，而且能量較集中在高頻。

Mel-frequency cepstral coefficients (MFCCs)

- MFCC特徵的最大優點考慮到人耳聽覺的特性:
  - 梅爾頻率倒頻譜是倒頻譜的一種應用，梅爾頻率倒頻譜常應用在聲音訊號處理，對於聲音訊號處理比倒頻譜更接近人耳對聲音的分析特性。
  - **根據耳朵對低頻有較高解析度**，對頻率的響應並非呈線性關係，而是成**對數**的關係。
  - 對於倒頻譜係數的計算將更**強調低頻**的部分，使所求出的係數更能防止雜訊的干擾。

MFCC 流程


FFT
Spectrum
Mel-scale
triangular filter
Integration
and data
reduction
LOG
DCT
MFCC

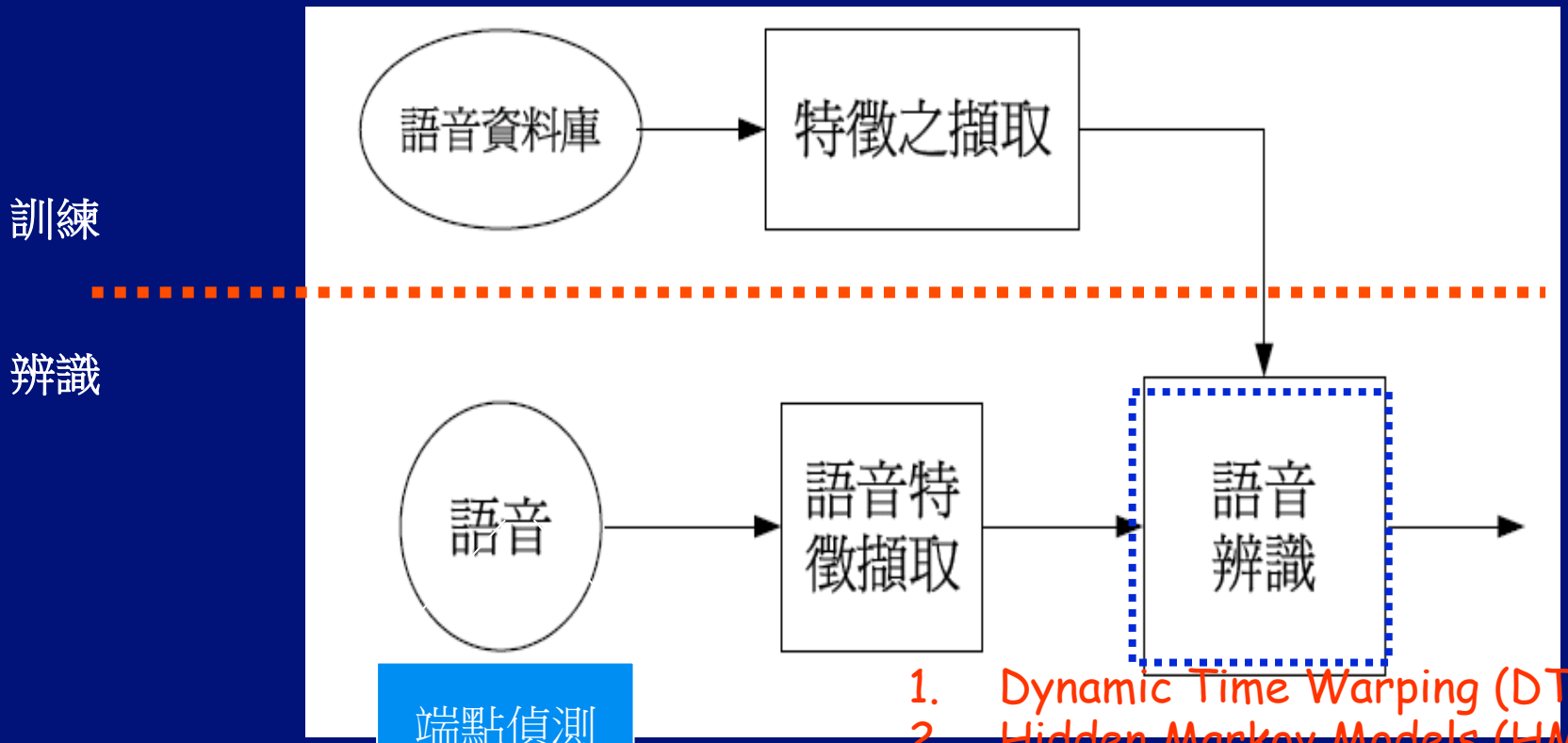
語音處理的幾個技術

- (1) 語音編碼(speech coding)
- (2) 語音合成(speech synthesis)
- (3) 語音辨識(speech recognition)
- (4) 語音增強(speech enhancement)
- (5) 語者辨認(speaker recognition)
- (6) 其他

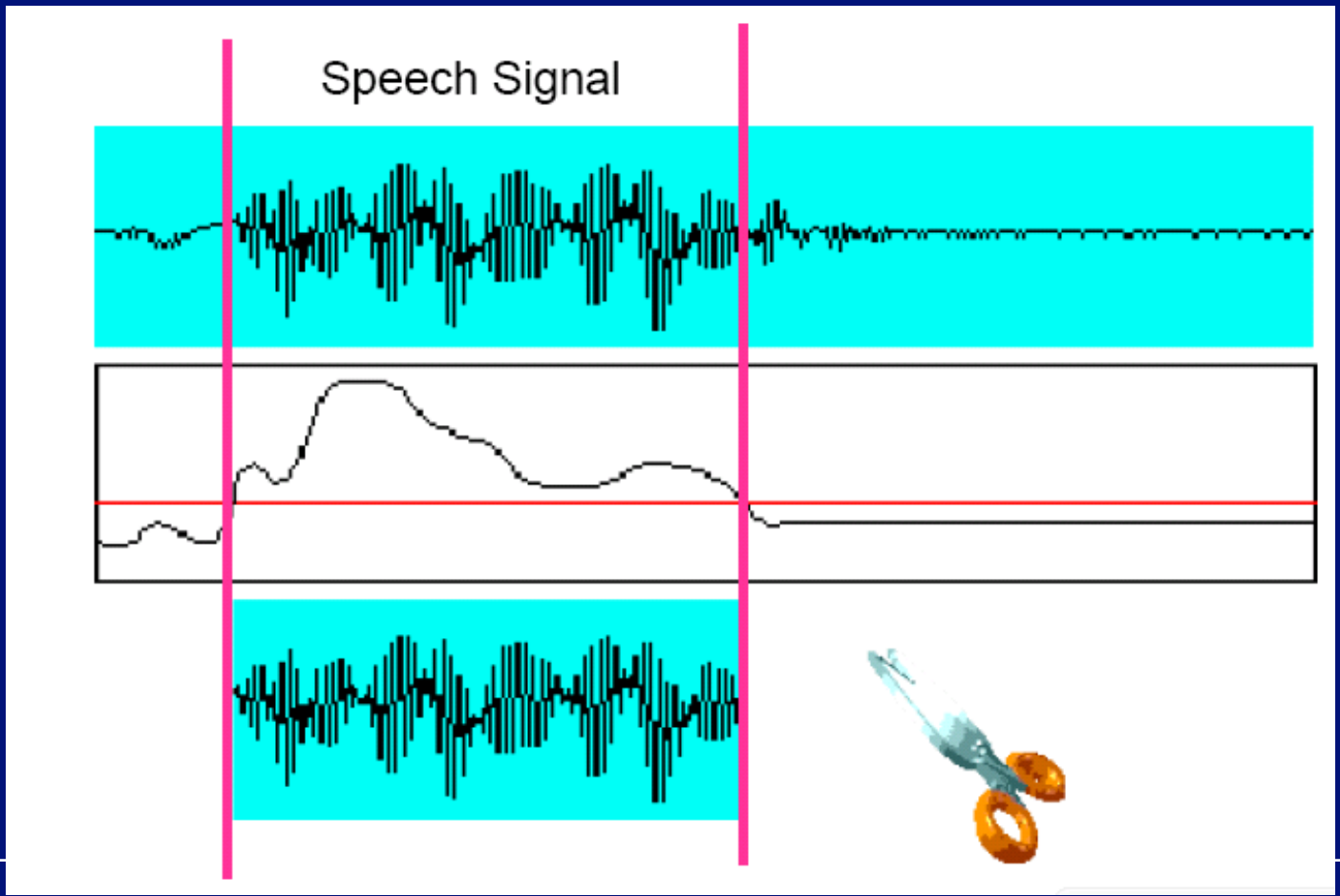
Outline

1. 音訊的簡介
2. 語音辨識系統
3. 嵌入式語音系統
4. Demo
5. 結論

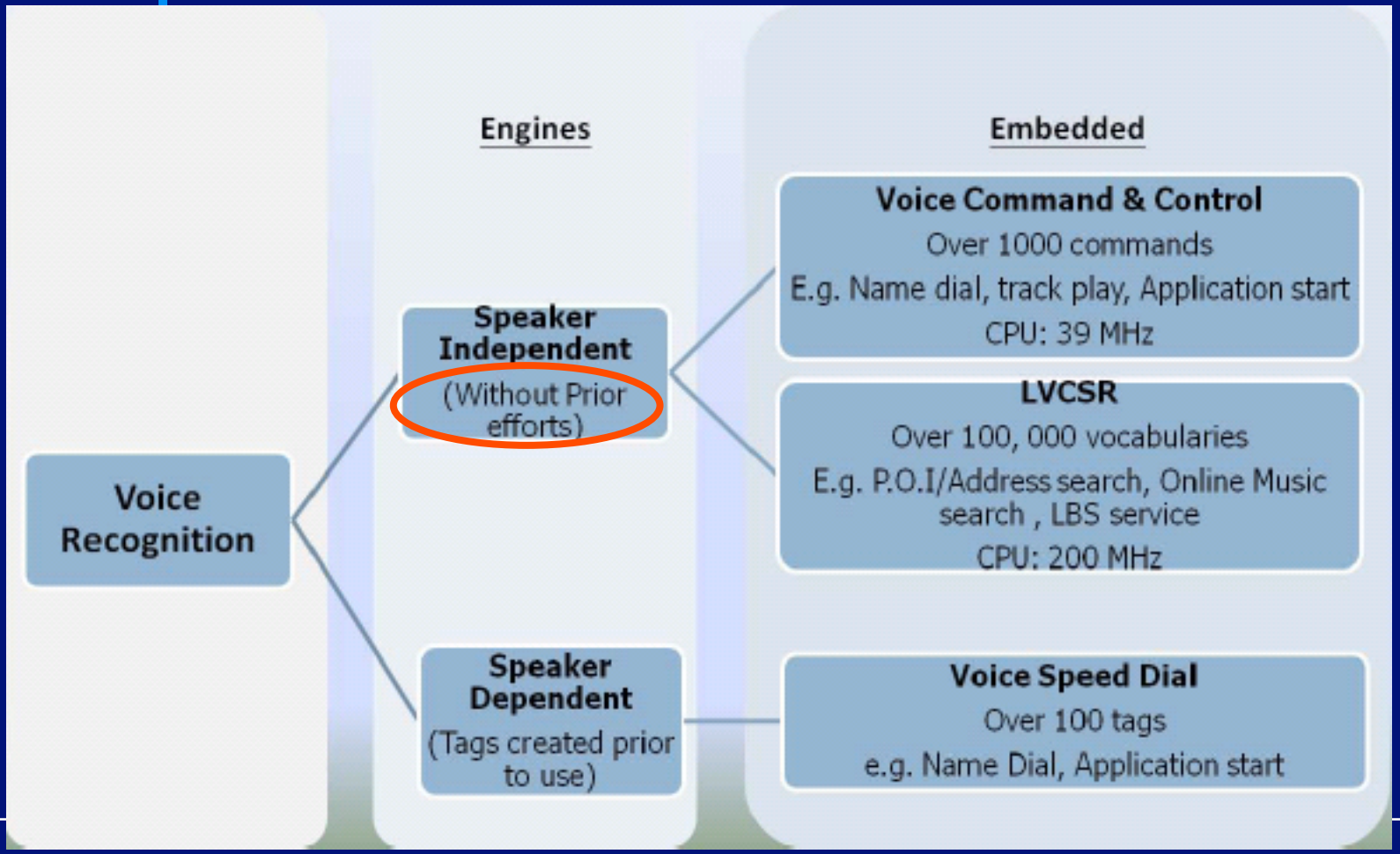
語音辨識系統


訓練
辨識
語音資料庫
特徵之擷取
語音
語音特徵擷取
語音辨識
端點偵測 (Endpoint Detection)
1. Dynamic Time Warping (DTW)
2. Hidden Markov Models (HMM)

端點偵測
(Endpoint Detection)



Speaker Independent (SI) vs. Speaker Dependent (SD)

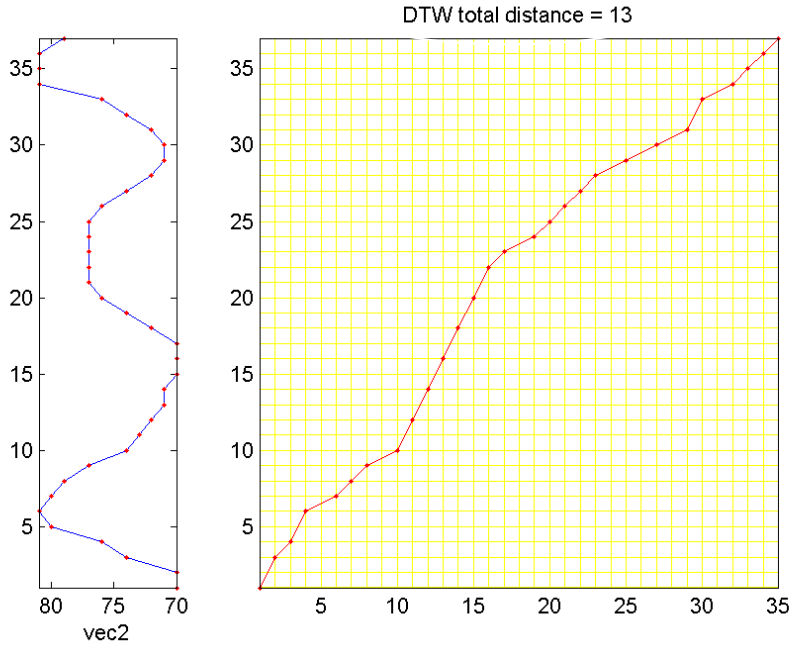


LVCSR: Large vocabulary continuous speech recognition

DTW (1)

min

- 中文可以翻譯成「動態時間扭曲」或是「動態時間校正」，這是一套根基於「動態規劃」（Dynamic Programming，簡稱DP）的方法。

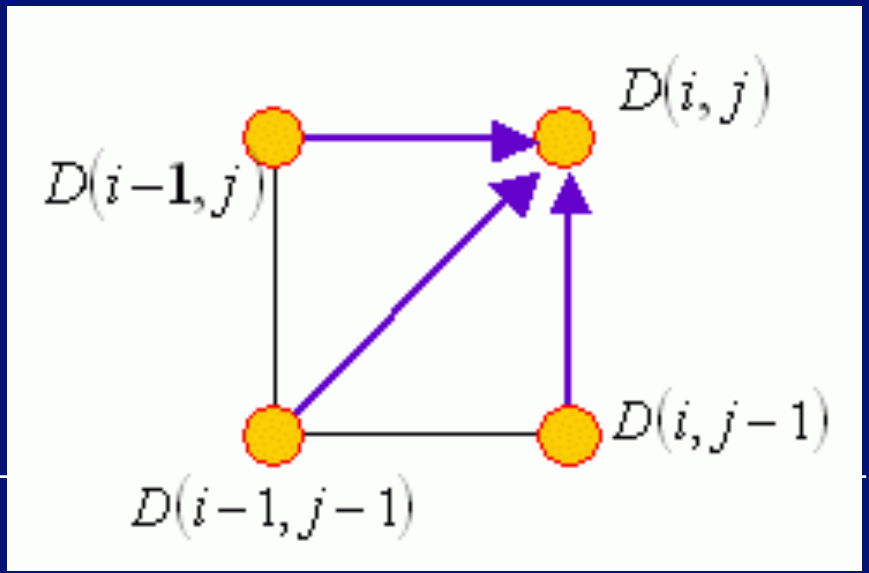


特徵係數_r2
特徵係數_r3
特徵係數_r4
特徵係數_r5

特徵係數_t

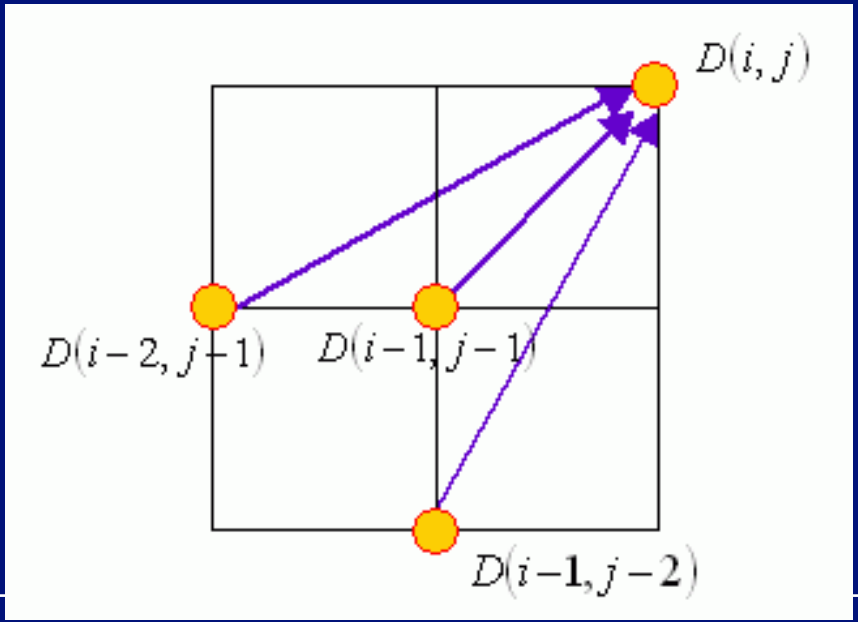
DTW (2)

- 我們要如何很快地找到這條最佳路徑呢？我們可以根據 DP 的原理，來將 DTW 描述成下列四大步驟：
  - 目標函數之定義：定義 D(i, j) 是 t(1:i) 和 r(1:j) 之間的 DTW 距離，對應的最佳路徑是由 (1, 1) 走到 (i, j)。
  - 目標函數之遞迴關係：D(i, j) = ｜t(i) - r(j)｜ + min{D(i-1, j), D(i-1, j-1), D(i, j-1)}
  - 端點條件：D(1, 1) = ｜t(1) - r(1)｜
  - 最後答案：D(m, n)



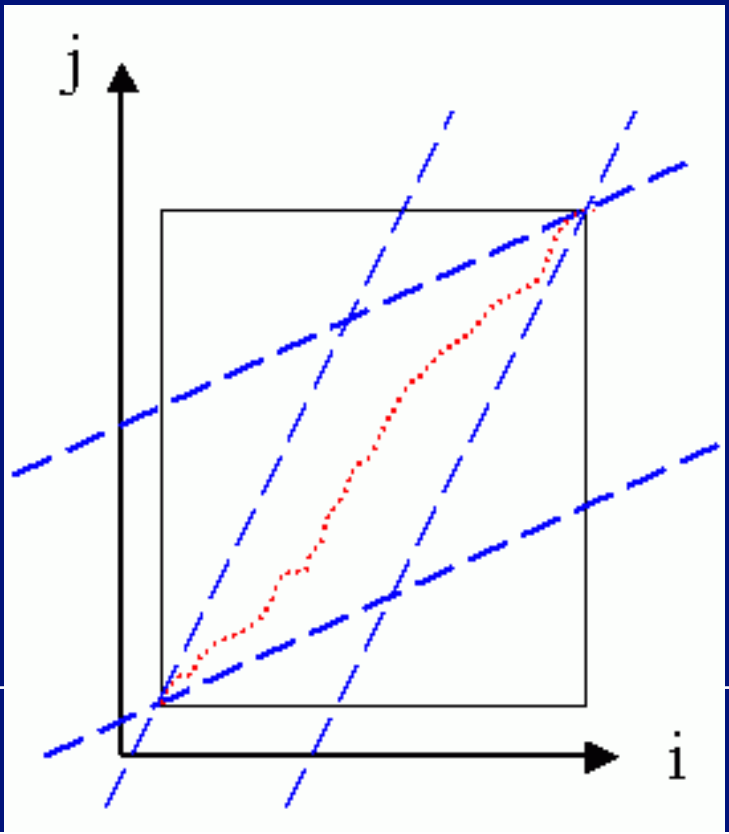
DTW (3)

- 另一個常用到的 local path constraint，是 27°-45°-63°



DTW(4)

- global path constraint
  - 可以有效地將搜尋比對的時間大幅降低



DTW(5)

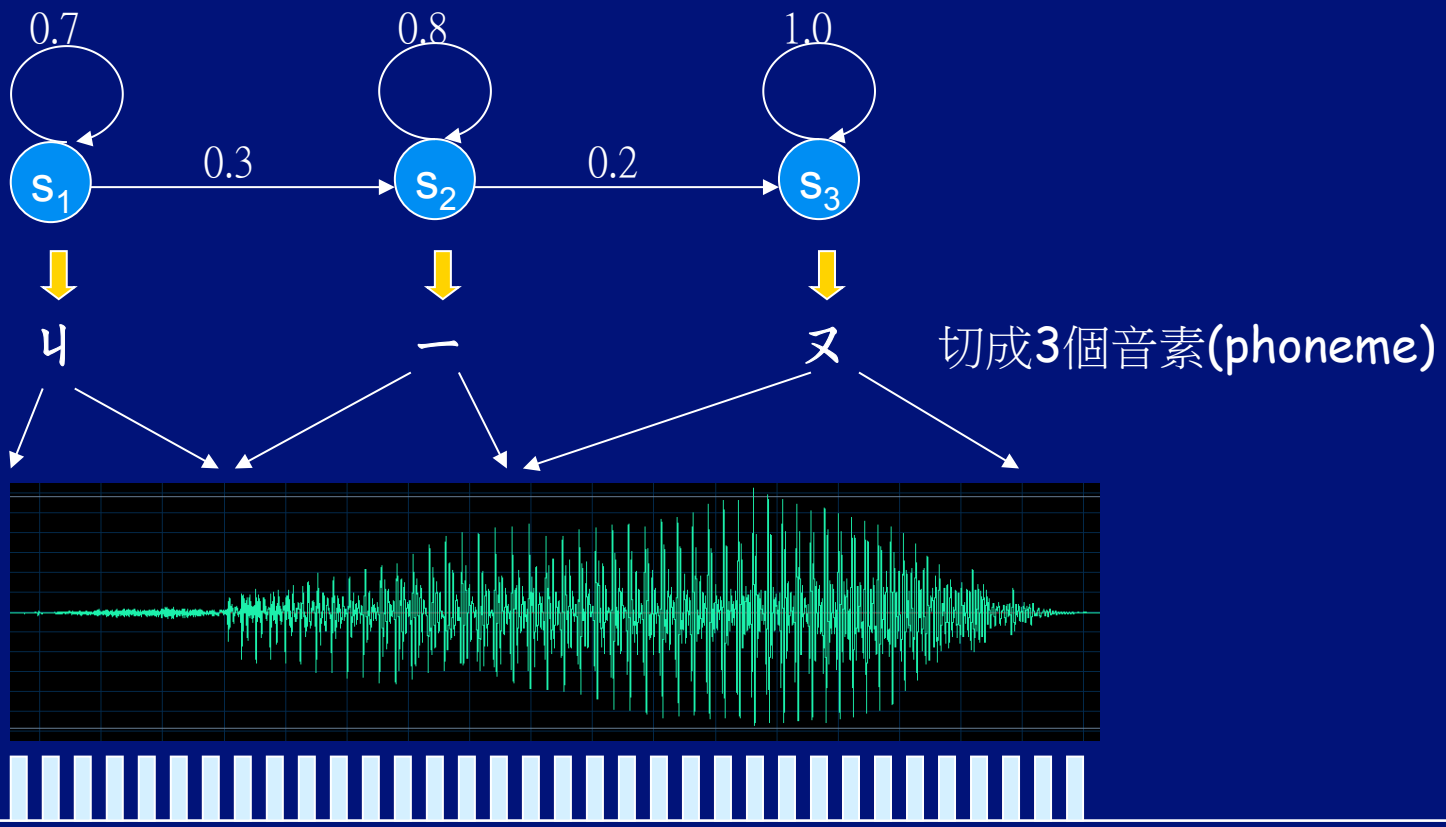
- 大部分是用於語者相關（Speaker Dependent）的語音辨識
- 這一類的應用大部分需要使用者自行錄音，然後再以自己的聲音來比對之前錄好的語音資料
- 應用範圍比較狹隘，譬如目前手機 Name Dialing 等等

HMM (1)

- 如果我們要做到語者無關（Speaker Independent）的語音辨識，最常見的方法，就是「隱藏式馬可夫模型」（Hidden Markov Models），簡稱 HMM。
- HMM 是根基於統計的機率模型，特別適用於具有大量訓練資料的語音辨識系統。

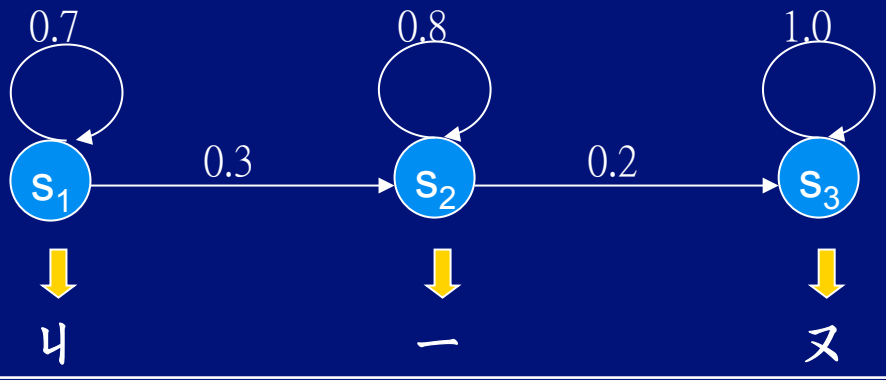
HMM (2)

- 數字「九」的模型：



HMM (3)

- 數字「九」的 HMM 模型的相關參數
  - Transition Probability A: A(i,j)是從state i 跳到 state j 的機率
  - State Probability B: B(i,j)是 frame i 隸屬於 state j 的機率



語音辨識的應用

安全

醫療照顧

娛樂

語音辨識應用範圍

	聽/寫	控制
電腦系統	教學及訓練 聽寫及文章撰寫	以語音操作取代滑鼠/鍵盤操作 如：個人電腦、伺服器、Kiosk、 針對殘障/視障使用之電腦系統
行動通訊/PDA	語音自動答錄、搜尋留言/郵件 個人數位語音助理(電子記事本)	語音搜尋電話號碼/自動撥號 自動化總機系統
消費性電子	音樂旋律辨識及搜尋	家電控制 聲控玩具
汽車		汽車音響/導航/通訊系統
工業/自動化/機器人		自動化生產 倉儲管理
安全/軍事	聲音(槍砲聲)辨識/搜尋/監控	安全監控 門禁管理

Outline

1. 音訊的簡介
2. 語音辨識系統
3. 嵌入式語音系統
4. Demo
5. 結論

嵌入式聲控系統簡介(1)

- 市面上大多數的消費性多媒體產品缺少語音聲控介面
  - **MP3**隨身聽, 翻譯機, **DVD**播放器, **LCD**電視

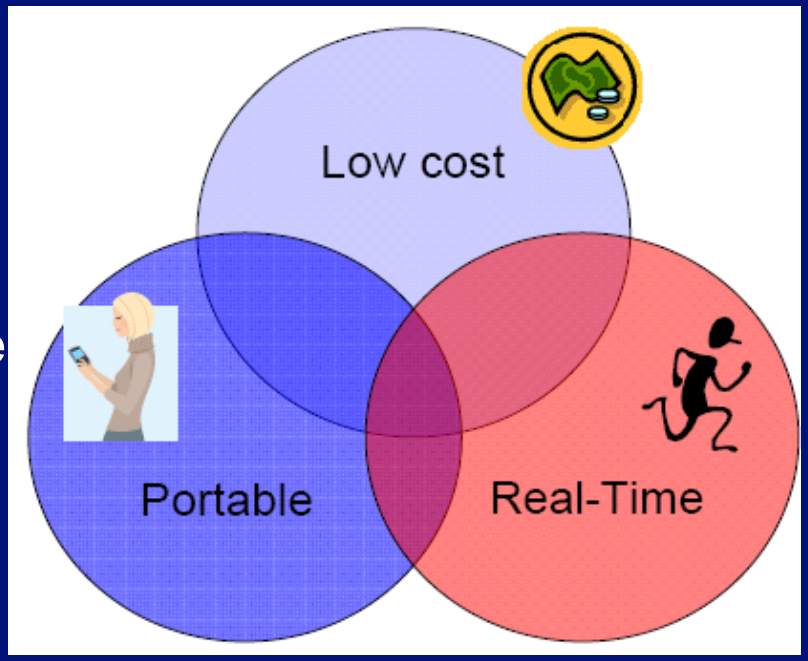
嵌入式聲控系統簡介(2)

- 某些產品具有聲控介面, 但是要搭配電腦方可運作
  - **EZ Talk, Talk to Me, MyET, LiveABC, and etc.**

嵌入式聲控系統??

Design Criteria for Embedded System

- CPU Capability
- Real time consideration
- Floating or fixed point computation
- Platform Selection
- Performance assu
- Memory Requirement
- Cost issue



嵌入式聲控系統的三大要求:
成本低，體積小，可即時執行語音辨識

Embedded Speech Recognition (嵌入式語音系統)

- Cyberon (賽微) speech recognition for Mobil Phone
- Sunplus (凌陽) Embedded SR
- HOTECH (創意先進) Voice Me

Cyberon 賽微
--Voice Commander

整合賽微技術之產品

ELIYA Easy Life, In Style	http://www.eliya.com.tw s856, F608	FAR EASTONE 遠傳	http://www.fetnet.net Commtiva Z1 SHOW All
FUJITSU SIEMENS	http://www.fujitsu-siemens.com Loox T810, T831	garmin asus	http://www.garminasus.com M20 SHOW All
glofiish	http://www.glofiish.com X500, M700, X500+, X800, X600, DX900	Gsmart	http://www.gigabytecm.com S1200, MW700, MS800, g-Smart t600, g-Smart i350, g-Smart i300, g-Smart i120 SHOW All
Haier	http://www.haier.com V9000(Asus J101)	HITACHI Inspire the Next	http://www.hitachi.com HTG-668
htc smart mobility	http://www.htc.com P3600, P3300, P4350, S310, MteoR, Touch, TyTN II, Touch Dual SHOW All	hp invent	http://www.hp.com iPAQ 612,iPAQ rw6828, iPAQ rw6818, iPAQ 510, iPAQ Glisten

	http://www.motorola.com MPx200, V690, V878, V872, A668, A732, A3100	NOKIA Connecting People	http://www.nokia.com 6708, 330 (PND)
O2	http://www.o2.co.uk Atom, Xda Stealth, Flame, Graphite, Zinc	OKWAP	http://www.okwap.com A236, A268, A272, A375, i519, A363+, A323, 英語霸 SHOW All
PANTECH	http://www.pantech.com Matrix Pro. SHOW All	PAPAGO!	http://www.papago.com.tw R15 (語音導航), R6600 (語音導航)
PHILIPS	http://www.philips.com 960, 768	大眾電信	http://www.phs.com.tw PG2000
SIEMENS	http://www.siemens.com T55, ST60	Skyworth	http://www.skyworth.com E760, E780
SAMSUNG	http://www.samsung.com/tw/index.htm SGH-i608, OMNIA2 i8000, Lite B7300	Sony Ericsson	http://www.sonyericsson.com X2

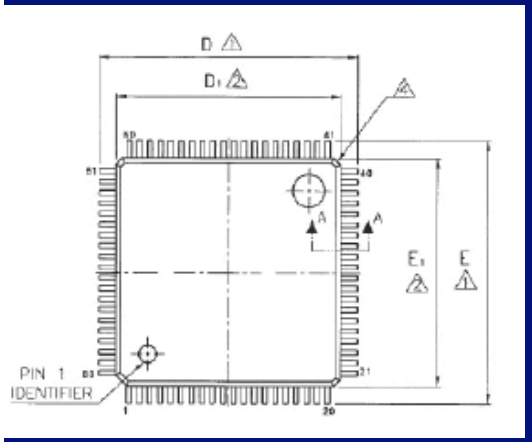
acer	http://www.acer.com.tw DX900, F900, M900, X960, SHOW All	ASUS	http://www.asus.com P835, P526, P735, P535, Z801, P305, M310, J501, J502, R600(PND), P527, P735 SHOW All
AnexTEK Global Inc.	http://www.anextek.com.tw moboDA 3360	BenQ	http://www.benq.com/ P30, P50, EF51, P51, E72
Chunghwa Telecom	http://www.emome.com.tw CHT9000, CHT9100, CHT9110	CHANGHONG 长虹	http://www.changhong.com.cn/ A320, V828
Coolpad 酷派	http://www.coolpad.cn/ 768	dopod 多普达	http://www.dopod.com D600
dopod Enjoy Mobility	http://www.dopod.com.tw M700, C800, P800W, S300, D600, C500, C730, U1000 SHOW All	ETEN	http://www.eten.com.tw M800,X650,P300, M500, M600, G500, M600+, G500+

iDo Commitment for Life	http://www.ido-mobile.com S600,S601	i-mate	http://www.imate.com SP3, JAM, K-JAM, JAMin, Smartflip, SPJAS, JASJAM, JAQ3.Ultimate 6150,8150,8502,9502
LUXGEN	http://www.luxgen-motor.com.tw LUXGEN7 MPV SHOW All	lenovo	http://www.lenovo.com V800, i807, i908, S9
malata	http://www.malata.com M696, M699	mio Digi Walker	http://www.mio-tech.com A702, C720, C520, C517, A701, A501, C210, P350 SHOW All
	http://www.motorola.com MPx200, V690, V878, V872, A668, A732, A3100	NOKIA Connecting People	http://www.nokia.com 6708, 330 (PND)
O2	http://www.o2.co.uk Atom, Xda Stealth, Flame, Graphite, Zinc	OKWAP	http://www.okwap.com A236, A268, A272, A375, i519, A363+, A323, 英語霸 SHOW All
PANTECH	http://www.pantech.com Matrix Pro. SHOW All	PAPAGO!	http://www.papago.com.tw R15 (語音導航), R6600 (語音導航)

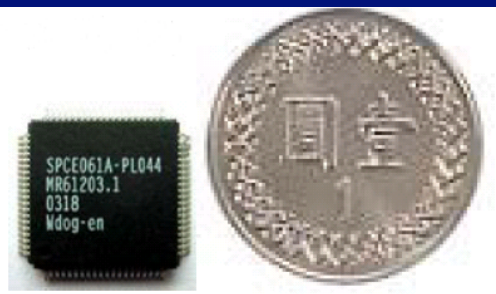
Sunplus 凌陽 --SPCE061A (1)

Core	16-bit MCU
Crystal	32768Hz
Internal PLL	Built-In
CPU Clock (MHz)	0.32 ~ 49.512
MIPS	24
Internal RAM	2K Words
Internal ROM	32K Words
Operating voltage	2.6V ~ 3.6V
Operating current	26mA @ 3.3V
Sleep/Idle current	2μA @ 3.3V
Operation Temperature	0°C ~ 60°C
General I/O pins	32

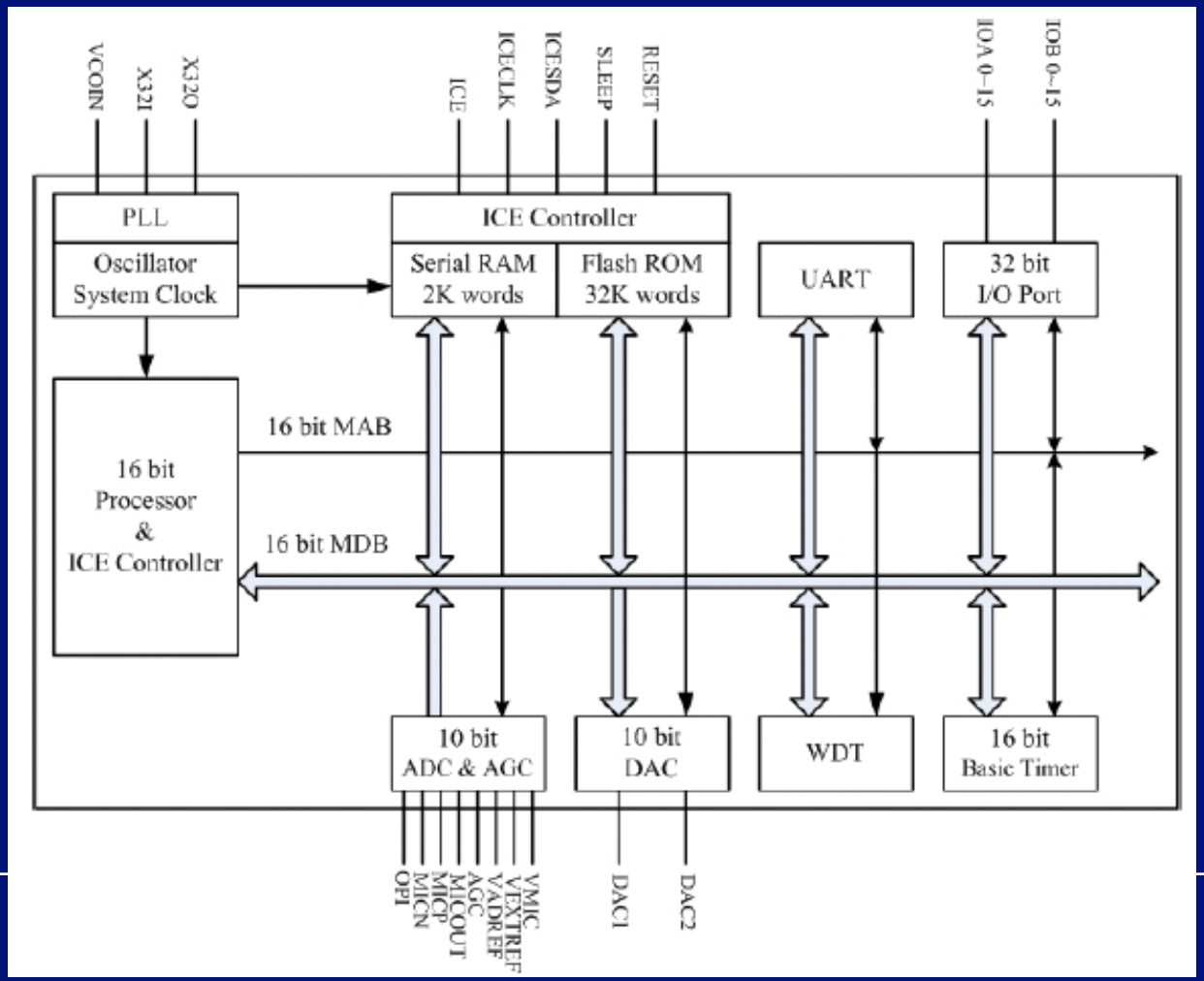
ADC	10 bit
DCA	10 bit (×2)
Timer/Counter	2
Interrupt source	14
Watch Dog	1
Microphone Pre-Amp	Built-In
Digital Filter	-
Die Price	US $2 (2.5K)



Size:
D1 = E1 = 0.472 in = 1.2cm
Area: 1.44 cm2



Sunplus 凌陽 -SPCE061A (2)



Sunplus 凌陽 -SPCE061A (3)

- 應用領域
  - 語音識別類產品
  - 通用語音合成產品
  - 智慧語音互動式玩具
  - 高級教具類玩具
  - 兒童電子故事書產品

HOTECH 創意先進
VOICE ME聲控遙控器(1)

- e摩控 VOICE ME是由 HOTECH 創意先進 所研發生產

HOTECH 創意先進
VOICE ME聲控遙控器(2)

- e摩控產品主要特點:
  - 輕鬆賦予紅外線遙控器家電聲控功能
  - 不需改裝家電，安裝操作容易
  - 遠距離聲控啟動，免手按鍵
  - 30 組聲控指令
  - 單一組聲控指令可以儲存 3 個紅外線信號
  - 智慧型紅外線學習功能
  - 多方向紅外線發射，接收無死角
  - 聲控指令設定國語、英語、台語、客家話..皆可以

- e摩控產品規格:
  - 特定語者語音辨認技術
  - 紅外線遙控器接收距離 1 公尺
  - 紅外線發射距離最遠 10 公尺
  - 語音說話距離最遠 5 公尺
  - 電源消耗 9V DC 300 mA
  - 重量 130 公克
  - 工作溫度 0 度至 40 度

Outline

1. 音訊的簡介
2. 語音辨識系統
3. 嵌入式語音系統
4. **Demo**
5. 結論

Demo(1)

- Speaker Identification (語者確認)
- Voice Command (聲控系統)
- Sound Source Localization (音源定位系統)
- **Speech emotional recognition (語音情緒辨識技術)**

Demo(2)

- Animal Sounds Recognition (動物聲響辨識)
- Music Recognition System (歌曲辨識系統)
- Instruments Recognition(樂器辨識)
- Automatic Speech Recognition & Melody Recognition (from MIR Lab.)

Outline

1. 音訊的簡介
2. 語音辨識系統
3. 嵌入式語音系統
4. Demo
5. 結論

結論

- 優勢
  - 張嘴比動手容易
  - 語音辨識技術具人性化的操作介面
  - 未來不管在室內的任何地方均可隨時隨地通話而不須手持話筒或是配戴免持耳機

結論

- 缺點
  - 語音訊號的差異性大(說話速度、習慣、生理狀況、性別、年齡、地域等)
  - 語音訊號分段的困難
  - 辨識率易受背景雜訊的影響。目前語音辨識技術仍有許多困難尚待解決，如在吵雜工作環境下辨識率不佳
  - 辨識模型複雜度高，對硬體規格要求過高等問題

結論

- 市場
  - 系統穩定性
    - 外在環境噪音的干擾
    - 內在行為特徵的改變
  - 實用性
    - 及時性
    - 準確度
  - 接受度
    - 價錢
    - 用手快於用嘴

音訊處理未來扮演角色

- 4G
- Cloud
- Mobile
- Big Data

Speech-related APP --行動化語音辨識應用

- 卡拉OK APP
  - 哼歌搜尋適合歌曲
- 校區生物辨聲APP
  - 有聲導覽
- 聽診APP
  - 遠距脈搏心跳判斷
- 失智老人跌倒/異常聲音偵測APP

研究成果分想享

- 科技部計劃
- 主要論文著作

科技部計劃（連續8年）

計畫名稱	起訖年月	補助或委託機構	執行情形	計畫內擔任的工作	經費總額
基於二維度長時距與短頻距之頻譜熵並結合單獨型遞迴模糊網路的車用語音有效偵測系統之研究(MOST 104-2221-E-158-002-)	2015/8/1~2016/7/31	行政院科技部	執行中	主持人	664,000
使用二維度紋理圖像資訊為基礎之語音情緒辨識系統以應用在遠距居家照護的研究與實作(MOST 103-2221-E-158-003-)	2014/8/1~2015/10/31	行政院科技部	執行中	主持人	475,000
基於主成份分析及支撐向量法的MP3音樂物件作自動化分類之研究及ARM嵌入式系統的實現(MOST 102-2221-E-158 -006)	2013/8/1~2014/10/31	行政院科技部	執行中	主持人	760,000

科技部計劃（連續8年）

計畫名稱	起訖年月	補助或委託機構	執行情形	計畫內擔任的工作	經費總額
針對手持式行動裝置建立一個具低運算量及高準確性的人聲有效檢測方法(NSC 101-2221-E-158 -005)	2012/8/1~ 2013/7/31	行政院國家科學委員會	已結案	主持人	433,000
基於特徵空間分析與多樣性時間頻率相關語音特徵權值整合的語音有效偵測之研究(NSC 100-2221-E-158-010-)	2011/8/1~ 2012/7/31	行政院國家科學委員會	已結案	主持人	459,000
以數位信號處理器實現即時語音增強系統及其在助聽器上的應用(99-2221-E-158-006-)	2010/8/1~ 2011/7/31	行政院國家科學委員會	已結案	主持人	573,000
具人聲特性的噪音頻譜預估以完成即時語音增強系統之研究(98-2221-E-158-004-)	2009/8/1~ 2010/7/31	行政院國家科學委員會	已結案	主持人	681,000
應用於可變噪音程度環境下的一個以聽覺遮蔽效應與小波閥值為基礎的單聲道語音增強演算法之研究(97-2218-E-158-003-)	2008/8/1~ 2009/7/31	行政院國家科學委員會	已結案	主持人	550,000

分享代表著作(1)

- B. F. Wu and K. C. Wang (通訊作者),, "A Robust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 762-775, Sep. 2005. **(SCI, Impact Factor =1.008, Rank Factor =12/20)**

A Robust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments

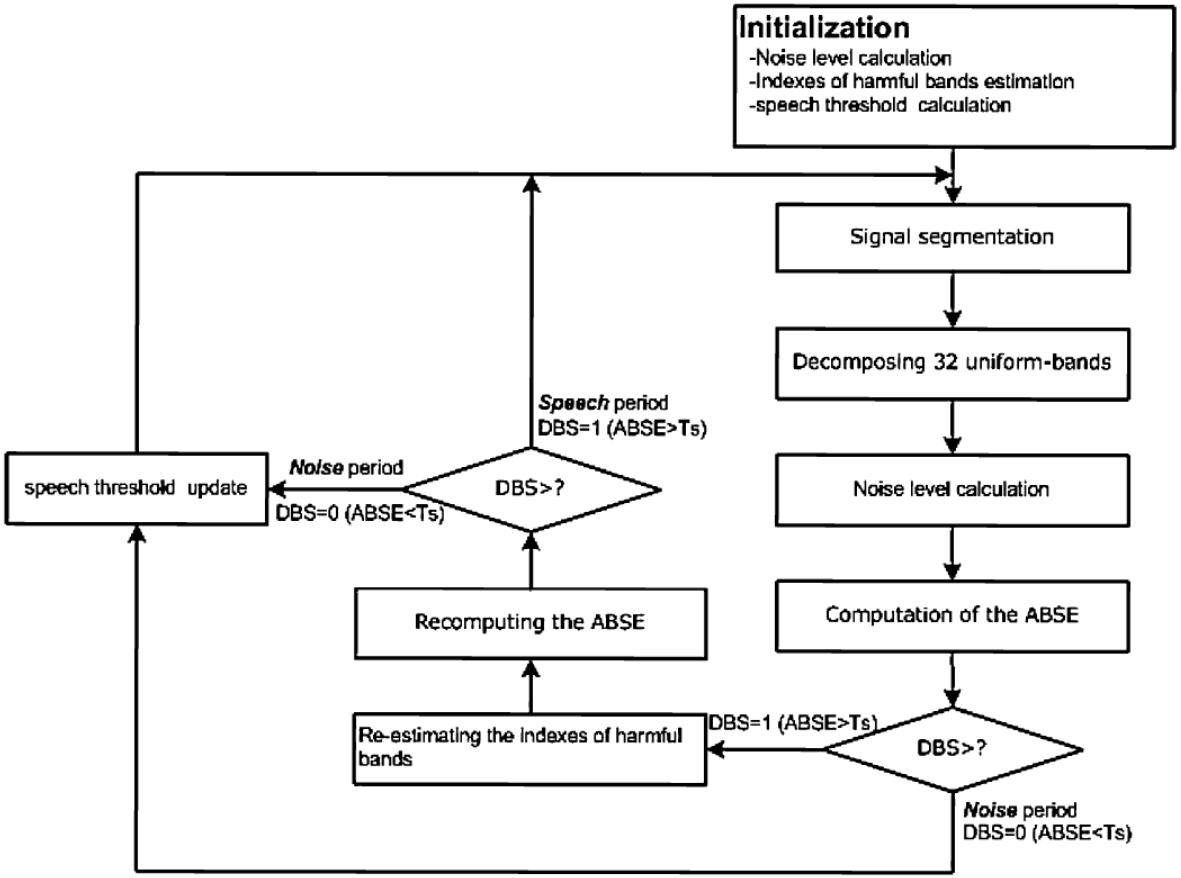


Fig. 10. Flowchart of the proposed ABSE-based endpoint detection algorithm.

分享代表著作(2)

- K. C. Wang (單一作者), "Wavelet-Based Speech Enhancement Using Time-Frequency Adaptation," *EURASIP Journal on Advances in Signal Processing*, pp. 1-8, Oct. 2009. (SCI, Rank Factor=132/246, Impact Factor=0.885)

Wavelet-Based Speech Enhancement Using Time-Frequency Adaptation

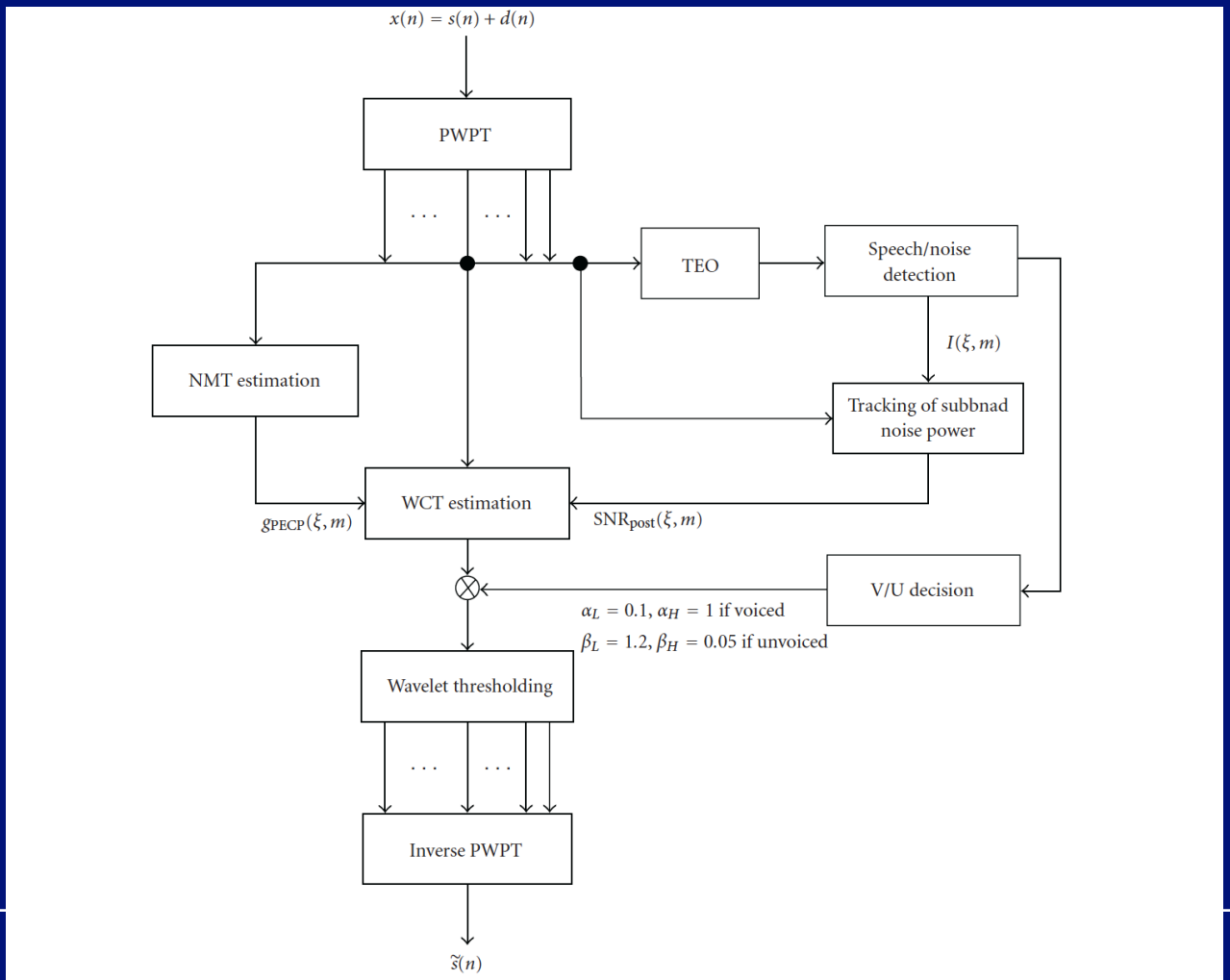


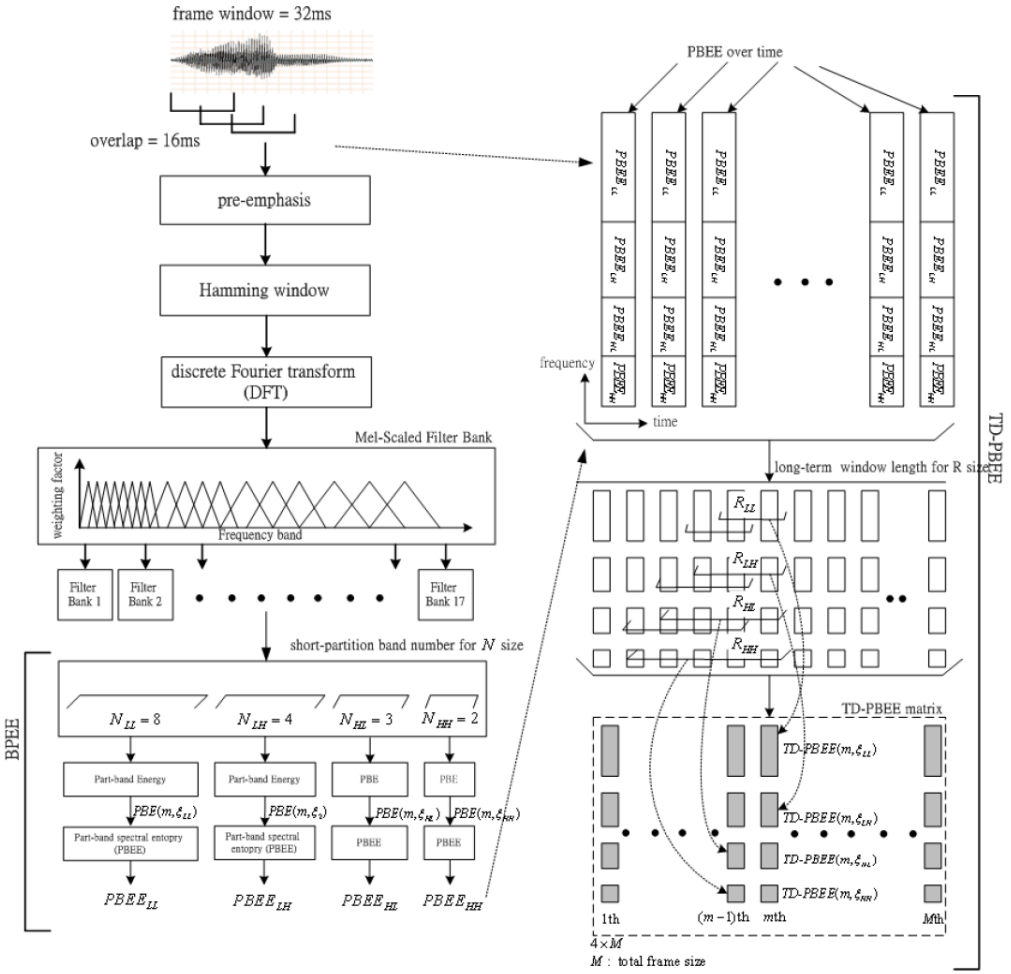
FIGURE 1: The architecture of proposed speech enhancement method based on the time-frequency adaptation of the wavelet threshold.

分享代表著作(3)

- K. C. Wang (單一作者), "A Novel Voice Sensor for the Detection of Speech Signals," *Sensors*, Vo.13, No.12, Dec. 2013. (SCI, Impact Factor=1.953 (2012), Rank Factor=8/57)

A Novel Voice Sensor for the Detection of Speech Signals

Figure 1. The block diagram of the feature extraction for TD-PBEE measurements.

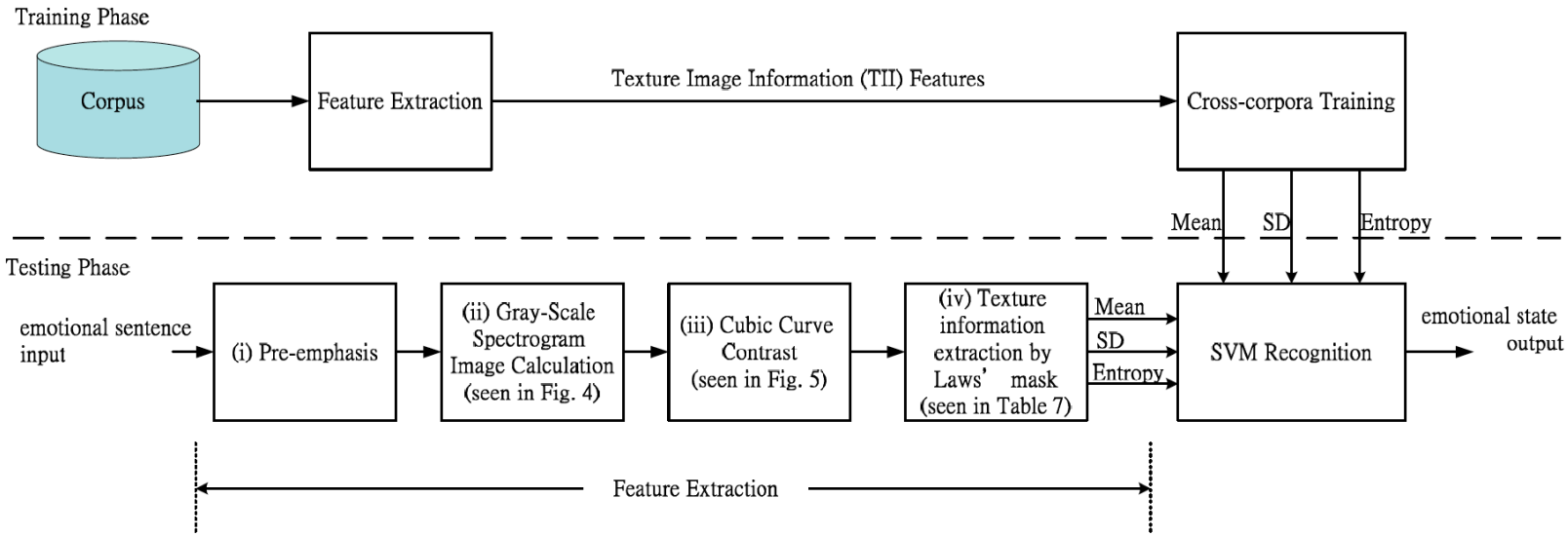


分享代表著作(4)

- **K. C. Wang (單一作者)**, "The Feature Extraction Based on Texture Image Information for Emotion Sensing in Speech," *SENSORS*, Vol. 14, no. 9, pp. 16692-16714, Sept. 2014. **(SCI, Impact Factor: 2.245 (2014); 5-Year Impact Factor: 2.474 (2014), Rank Factor=10/56=Top 17.8%)**

The Feature Extraction Based on Texture Image Information for Emotion Sensing in Speech

Figure 1. Diagram of the proposed TII-based ESS Algorithm.



To研究生的建議

- 研究題目不用太大
- 要有創新
- 多看相關論文, 增加廣度, 避免陷阱

國內資源

- 研究單位
  - ITRI 資通所-前瞻技術中心
  - 中華電信研究所
- 產業單位
  - 賽微科技
  - 聲碩科技
  - 凌陽科技
  - 清蔚科技
- 學術單位
  - 台大　李琳山教授、張智星教授 (MIR Lab.)
  - 交大　陳信宏教授、王逸如教授
  - 清大　王小川教授
  - 成大　王駿發教授

參考資料

1. 台大資工張智星教授-講義
2. 清大電機王小川教授-講義
3. 成大電機王駿發教授-講義
4. 樹德科大陳璽煌教授-講義
5. 新華電腦-講義
6. 華亨-實驗教材

謝謝聆聽~