機器學習及多媒體資訊檢索用於大資料分析的產學合作案例分享

張智星
台灣大學資訊系
jang@mirlab.org
http://mirlab.org/jang

Outline

- Wafer failure pattern classification
- Singing voice separation and audio melody extraction

Wafer Failure Pattern Classification

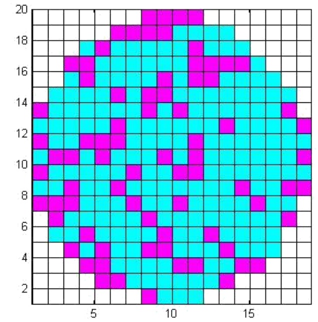
張智星 (Roger Jang)

台灣大學 資工系

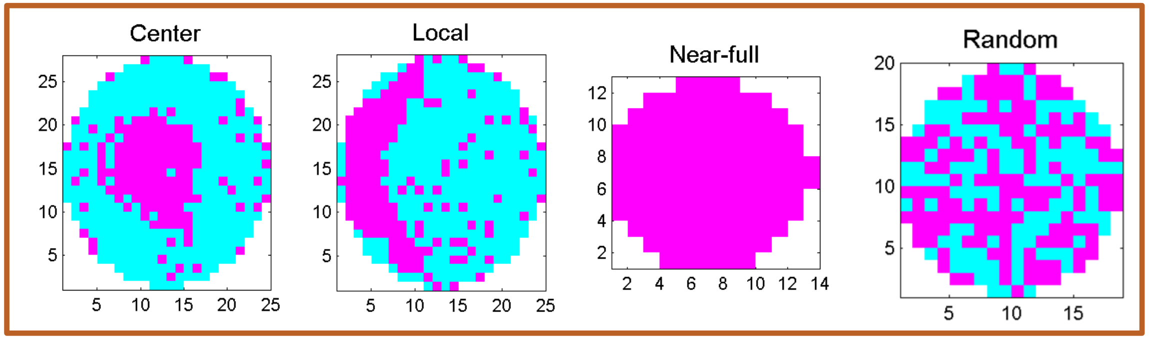
研究背景

- 晶圓製造業重點在於良率的提升，透過晶圓針測 (chip probing) 可以得到晶圓圖(wafer map)。
- 過去工程師需透過肉眼觀察晶圓圖找出特定的錯誤樣式，人工的觀察與標記勢必過於耗時。
- 研究目的
  - 將晶圓的錯誤樣式自動分類
  - 協助工程師判別錯誤樣式

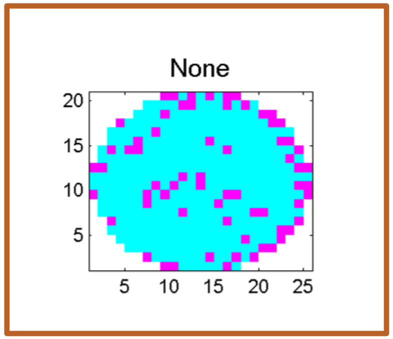
Wafer map



錯誤樣式(Pattern)



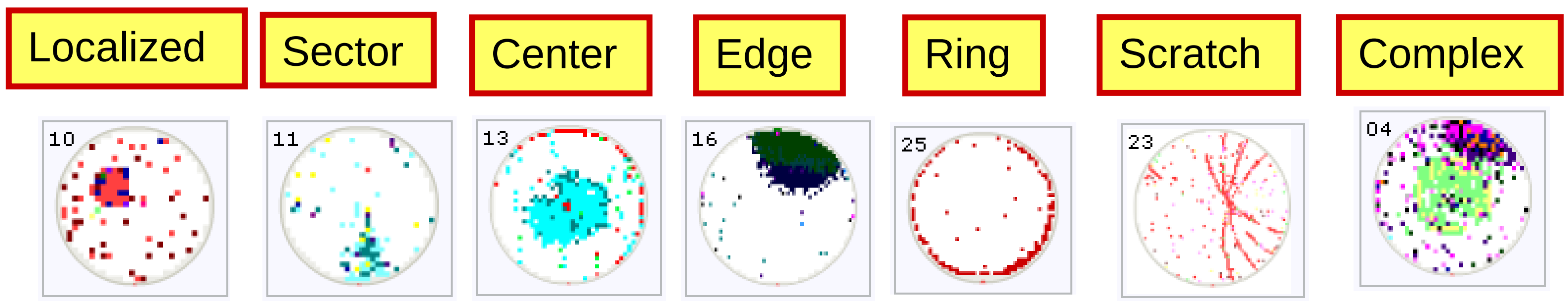
非錯誤樣式(None)



4

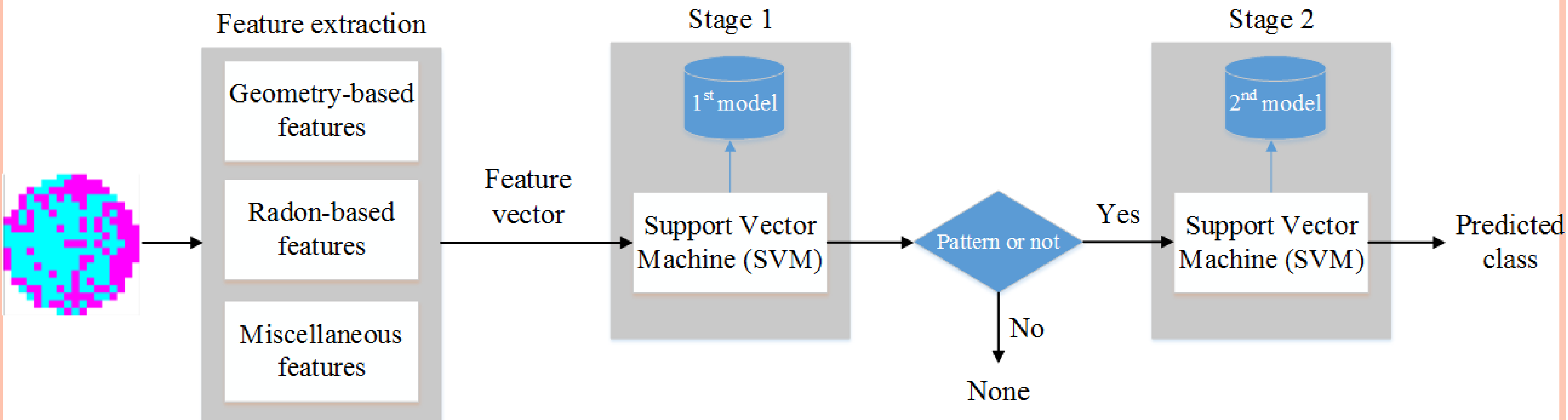
Typical Failure Patterns for Small-die Wafers

- Typical failure patterns for small-die wafers



- Note that
  - Colors represent different bad states
  - Wafers may have different numbers of dies
  - A specific failer pattern is highly correlated with a specific type of a machine’s mal-function.

研究方法
系統架構



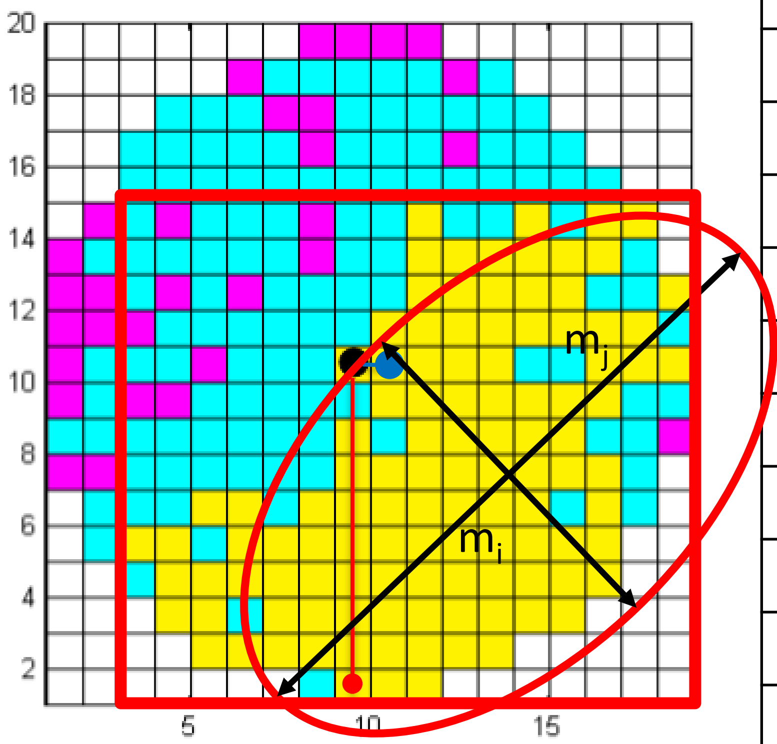
Feature extraction
Geometry-based features
Radon-based features
Miscellaneous features
Feature vector
Stage 1
1st model
Support Vector Machine (SVM)
Pattern or not
Yes
No
None
Stage 2
2nd model
Support Vector Machine (SVM)
Predicted class

Features

- Features are the most important and valuable asset of a classification system!
- Our features
  - Regions' features
    - Geometry
    - Statistics
    - Lines and curves (Hough transform)
  - Radon transform
  - Others: Histogram of 2×2 binary patterns and number of corners

研究方法
幾何特徵 (1/3)

- 區域性屬性 (Properties of region)

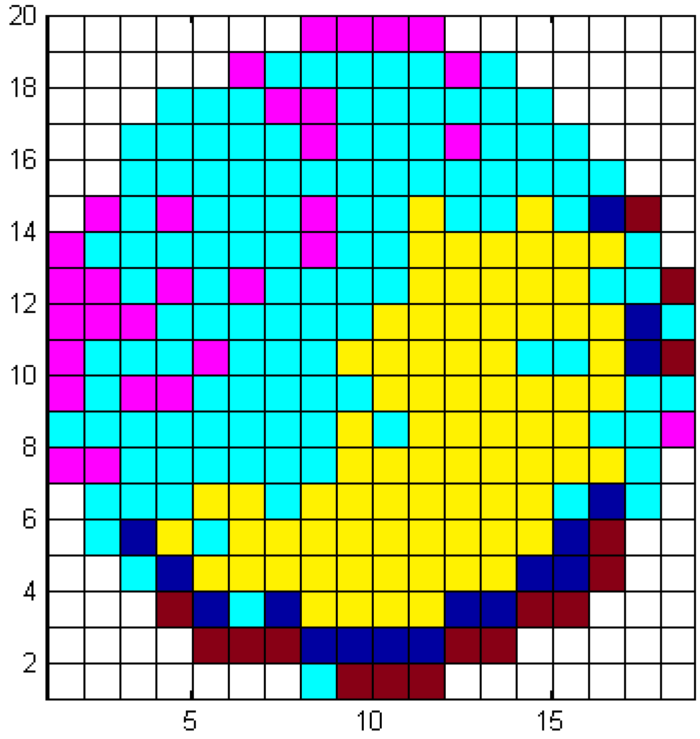


特徵值描述
S到晶圓圖中心的最長距離
S到晶圓圖中心的最短距離
S到晶圓圖中心的最長最短距離差值
S到晶圓圖中心的平均距離
S的實心度 (Solidity)
S的離心率 (Eccentricity)
S區塊橢圓之長軸與半徑比例
S區塊橢圓之短軸與半徑比例
S形成的bounding box有多少比例屬於原本S
S的周長與其bounding box比例

8

研究方法
幾何特徵 (2/3)

- 統計性屬性

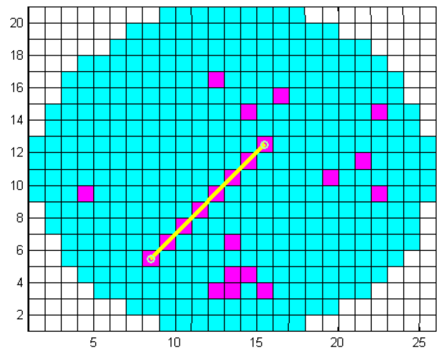


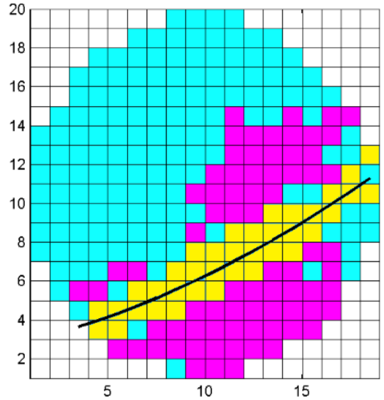
特徵值描述
S與最外圈晶圓圖的交集， 佔最外圈的比例
S與最外圈晶圓圖的交集， 佔S區塊的比例
S與次外圈晶圓圖的交集， 佔次外圈的比例
S與次外圈晶圓圖的交集， 佔S區塊的比例
S的面積佔整片晶圓圖的比例
S面積及次大區塊面積差佔整片晶圓圖的比例
S周長與晶圓圖半徑比例
S周長及次大周長區塊周長差和半徑比例
最外圈晶圓圖壞晶粒比例
次外圈晶圓圖壞晶粒比例
晶圓圖壞晶粒的比例

9

研究方法
幾何特徵 (3/3)

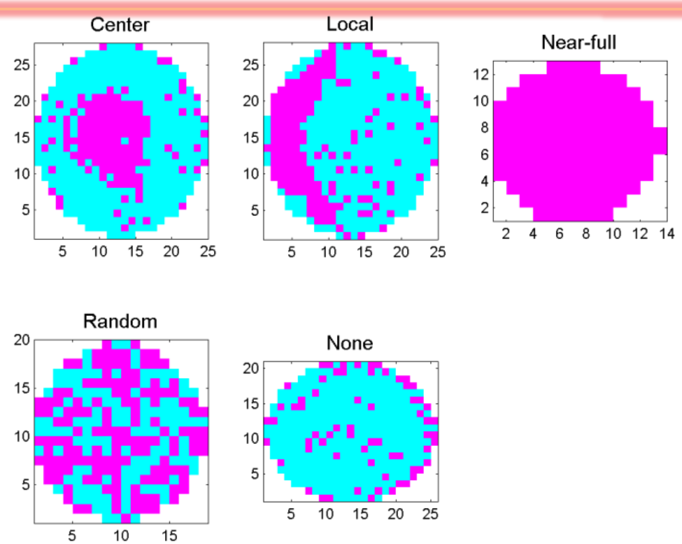
- 直線及曲線屬性

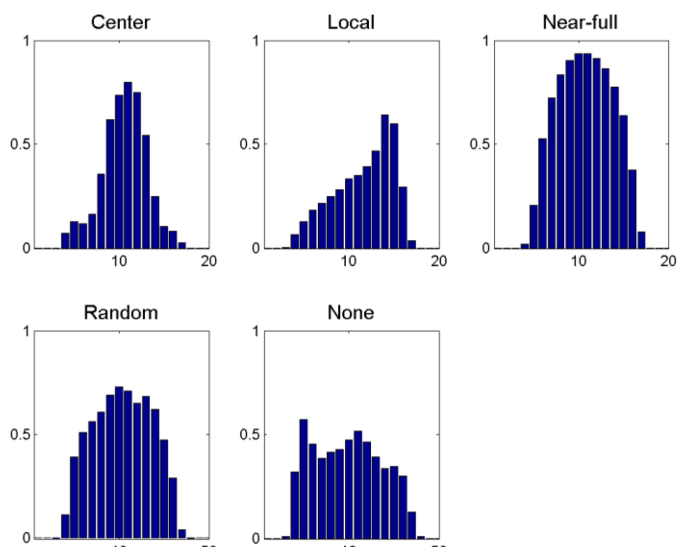


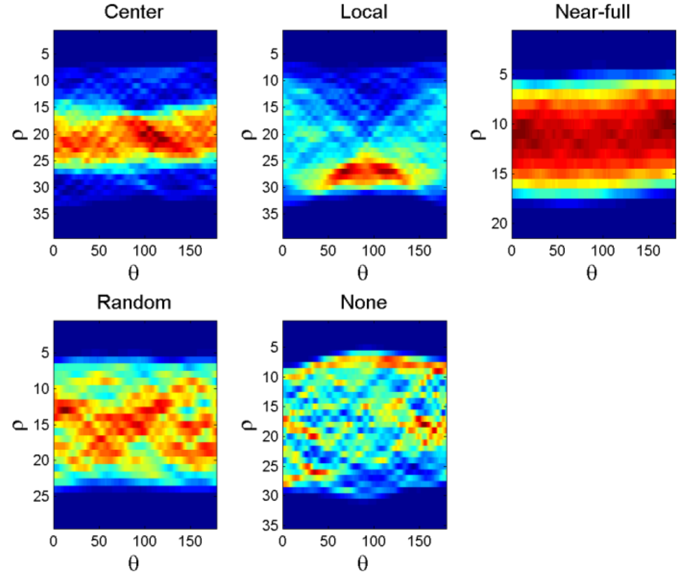


特徵值描述
確認晶圓圖上是否有線段
晶圓圖上線段長度與半徑比例
S上有多少塊壞晶粒接近拋物線，佔所有S的比例

研究方法
Radon特徵 (1/2)







$$g(\rho,\theta)=\sum_{x=1}^{m}\sum_{y=1}^{n}\mathbf{M}(x,y)\delta\boxed{(x\cos\theta+y\sin\theta-\rho)}$$

$$\delta(k)=\begin{cases}1 & \text{if k=0}\\ 0 & \text{otherwise}\end{cases}$$

g(ρ,θ) is the response of projection

$$\mathbf{G}=\begin{bmatrix} g(1,1) & g(1,2) & \cdots & g(1,\theta_{\max}) \\ g(2,1) & g(2,2) & \cdots & g(2,\theta_{\max}) \\ \vdots & \vdots & \ddots & \vdots \\ g(\rho_{\max},1) & g(\rho_{\max},2) & \cdots & g(\rho_{\max},\theta_{\max}) \end{bmatrix}$$

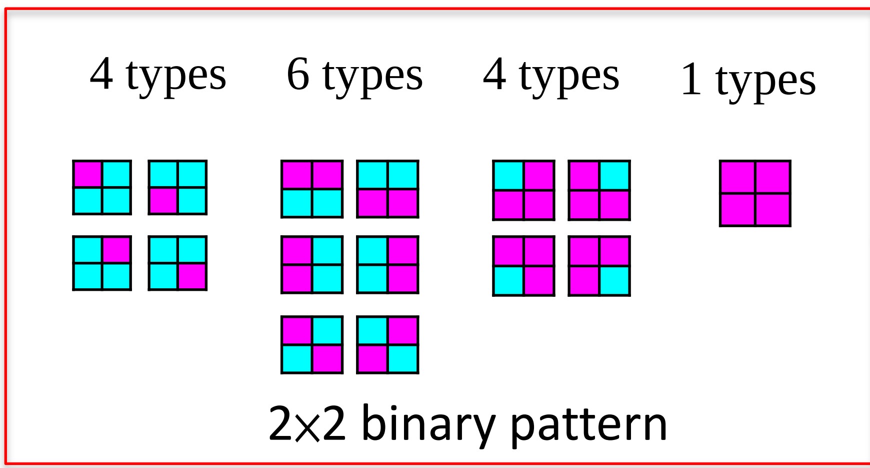
研究方法
Radon特徵 (2/2)

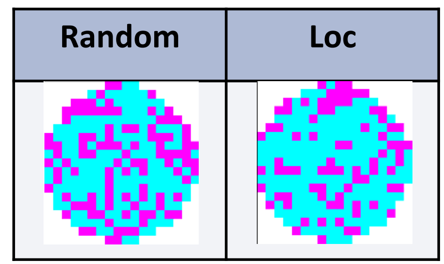
$$\mathbf{G} = \begin{bmatrix} g(1,1) & g(1,2) & \cdots & g(1,\theta_{max}) \\ g(2,1) & g(2,2) & \cdots & g(2,\theta_{max}) \\ \vdots & \vdots & \ddots & \vdots \\ g(\rho_{max},1) & g(\rho_{max},2) & \cdots & g(\rho_{max},\theta_{max}) \end{bmatrix}$$

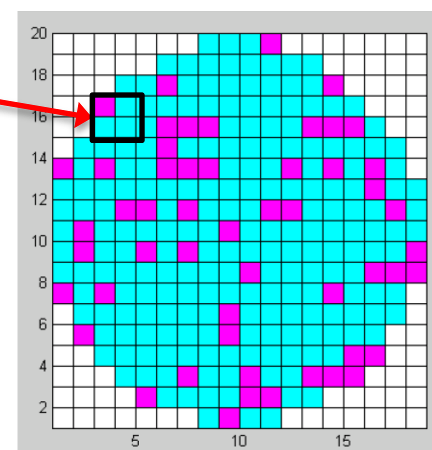
特徵值描述
列向量平均值 (重新取樣至20 dim)
列向量標準差 (重新取樣至20 dim)
列向量平均值的歪度 (Skewness) (1)
列向量平均值的峰度(Kurtosis) (1)
列向量平均值的中位數 (1)
列向量平均值中最大值 (1)
列向量平均值中最大值出現的位置 (1)
列向量平均值的最大減最小值 (1)

研究方法
其他特徵 (1/2)

- 以晶圓圖2x2 binary pattern的數量為特徵值 (Dimension = 15)
  - 計算一片晶圓上，每一種2x2 binary pattern各出現幾次，最後用晶圓圖晶粒數量正規化。

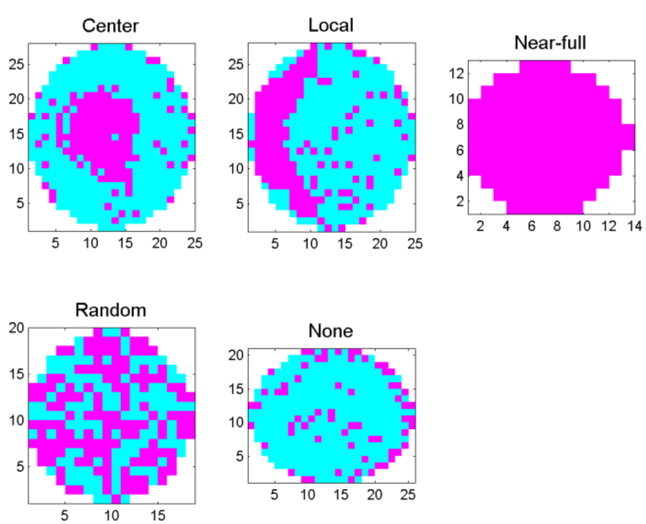


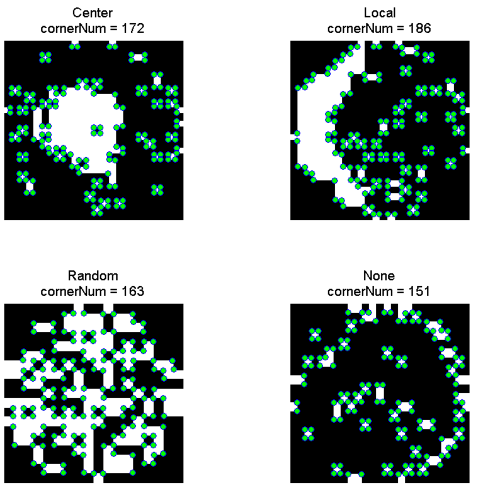




研究方法
其他特徵 (2/2)

- 晶圓圖上的轉角點數量與晶粒數量的比例 (Dimension = 1)
  - 不同錯誤樣式的晶圓圖轉角數量會有明顯差異
  - 計算晶圓圖上共有幾個轉角點，並以晶圓圖晶粒數量正規化，作為特徵值。
    - Harris corner detector





實驗結果與分析
晶圓圖資料集簡介

資料名稱	Big-die dataset		
晶粒數量（Die count）	300至500		
資料集各種類別數量（片）	類別	訓練資料	測試資料
	Center	168	128
	Local	1025	988
	Near-full	34	43
	Random	1262	1259
	None	8881	8922
	Total	11370	11340

（單位：%）	訓練資料	測試資料
錯誤樣式(Pattern)比例	21.9	21.4
非錯誤樣式(None)比例	78.1	78.6

實驗結果與分析
實驗設定說明—效能評估方式 (1/2)

- Test accuracy – 測試資料整體的辨識率結果

	Pattern	none
Pattern	97.73% (2363)	2.27% (55)
none	3.91% (349)	96.09% (8573)

1st stage

	Center	Loc	Near-full	Random
Center	92.8% (116)	1.6% (2)	0	5.6% (7)
Loc	2.9% (28)	91.7% (872)	0	5.4% (51)
Near-full	0	5.1% (2)	89.7% (35)	5.1% (2)
Random	0.8% (10)	6.4% (80)	0	92.8% (1158)
none	5.7% (20)	67.0% (234)	4.6% (16)	22.6% (79)

2nd stage

Predicted result

Ground truth	Center	Loc	Near–full	Random	none
Center	90.6% (116)	1.6% (2)	0	5.5% (7)	2.3% (3)
Loc	2.8% (28)	88.3% (872)	0	5.2% (51)	3.7% (37)
Near–full	0	4.7% (2)	81.4% (35)	4.7% (2)	9.3% (4)
Random	0.8% (10)	6.4% (80)	0	92.0% (1158)	0.9% (11)
none	0.2% (20)	2.6% (234)	0.2% (16)	0.9% (79)	96.1% (8573)

Combine two stages

實驗結果與分析
實驗設定說明—效能評估方式 (2/2)

- Remove “none” test accuracy – 僅計算測試資料中有錯誤樣式 (Pattern)部分的辨識率 (如4x5的藍色框框所示)

Predicted result

Ground truth

	Center	Loc	Near–full	Random	none
Center	90.6% (116)	1.6% (2)	0	5.5% (7)	2.3% (3)
Loc	2.8% (28)	88.3% (872)	0	5.2% (51)	3.7% (37)
Near–full	0	4.7% (2)	81.4% (35)	4.7% (2)	9.3% (4)
Random	0.8% (10)	6.4% (80)	0	92.0% (1158)	0.9% (11)
none	0.2% (20)	2.6% (234)	0.2% (16)	0.9% (79)	96.1% (8573)

類別	測試資料
Center	128
Local	988
Near-full	43
Random	1259
None	8922
Total	11340

實驗一
以晶圓圖2x2 binary pattern的數量為特徵值

Before:
Geometry + Radon + Corner detection (291 dim)
Test accuracy = 94.1 %
Remove “none” test accuracy = 90.7 %

	Center	Loc	Near-full	Random	none
Center	93.0% (119)	1.6% (2)	0	4.7% (6)	0.8% (1)
Loc	2.8% (28)	90.6% (895)	0	4.5% (44)	2.1% (21)
Near-full	0	0	79.1% (34)	11.6% (5)	9.3% (4)
Random	0.9% (11)	7.5% (94)	0.2% (2)	91.0% (1146)	0.5% (6)
none	0.4% (37)	3.2% (287)	0.3% (25)	1.1% (96)	95.0% (8477)

After:
Geometry + Radon + Corner detection
+ 2x2 binary pattern (306 dim)
Test accuracy = 94.2 %
Remove “none” test accuracy = 91.0 %

	Center	Loc	Near-full	Random	none
Center	92.2% (118)	1.6% (2)	0	5.5% (7)	0.8% (1)
Loc	2.9% (29)	89.7% (886)	0	5.3% (52)	2.1% (21)
Near-full	0	4.7% (2)	81.4% (35)	4.7% (2)	9.3% (4)
Random	0.8% (10)	6.5% (82)	0	92.2% (1161)	0.5% (6)
none	0.4% (34)	3.2% (285)	0.3% (25)	1.1% (100)	95.0% (8478)

實驗二
以晶圓圖上的轉角點數量為特徵值

Before:
Geometry + Radon + 2x2 binary pattern (305 dim)
Test accuracy = 93.6 %
Remove "none" test accuracy = 88.8 %

	Center	Loc	Near-full	Random	none
Center	91.4% (117)	3.9% (5)	0	3.9% (5)	0.8% (1)
Loc	3.9% (39)	86.2% (852)	0.2% (2)	7.3% (72)	2.3% (23)
Near-full	0	2.3% (1)	83.7% (36)	4.7% (2)	9.3% (4)
Random	2.2% (28)	6.6% (83)	0	90.6% (1141)	0.6% (7)
none	0.6% (55)	3.1% (281)	0.3% (31)	0.9% (82)	95.0% (8473)

After:
Geometry + Radon + Corner detection
+ 2x2 binary pattern (306 dim)
Test accuracy = 94.2 %
Remove "none" test accuracy = 91.0 %

	Center	Loc	Near-full	Random	none
Center	92.2% (118)	1.6% (2)	0	5.5% (7)	0.8% (1)
Loc	2.9% (29)	89.7% (886)	0	5.3% (52)	2.1% (21)
Near-full	0	4.7% (2)	81.4% (35)	4.7% (2)	9.3% (4)
Random	0.8% (10)	6.5% (82)	0	92.2% (1161)	0.5% (6)
none	0.4% (34)	3.2% (285)	0.3% (25)	1.1% (100)	95.0% (8478)

實驗三
透過特徵選取提升辨識結果 (1/3)

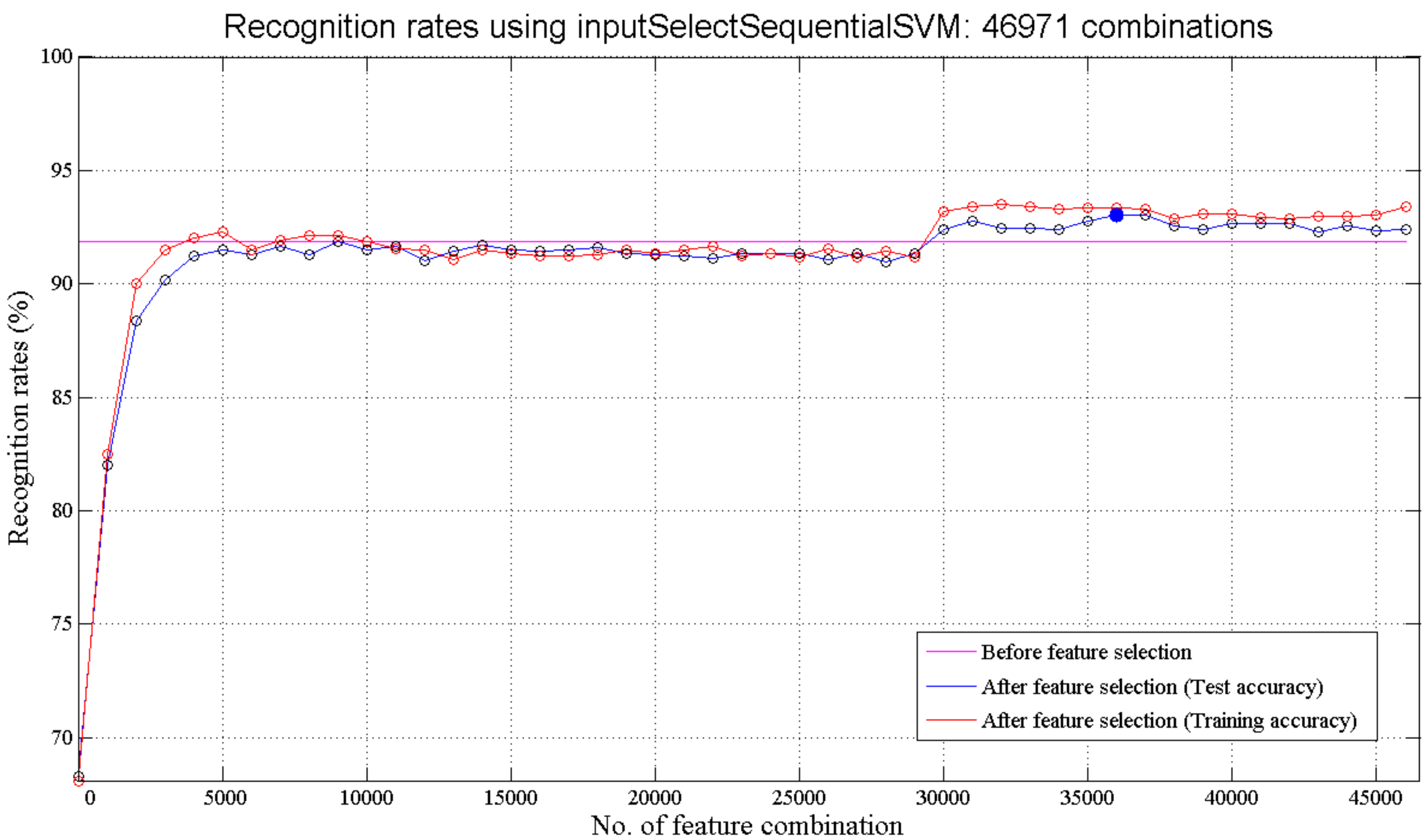
- 特徵選取方法：Sequential forward selection (簡稱SFS)
- 辨識率衡量方式：以本實驗的測試資料集辨識率結果，作為每次特徵選取的依據，辨識率計算如下:

$$辨識率 = \frac{測試資料中錯誤樣式(pattern)晶圓圖辨識正確的數量}{測試資料中錯誤樣式(pattern)晶圓圖總數量}$$

- 分類器：支撐向量機 (SVM)

特徵值	資料量	所需時間
306	Training set：2489片 Pattern資料 Test set：2418片 Pattern資料	35.3hr

實驗三
透過特徵選取提升辨識結果 (2/3)



21

實驗三
透過特徵選取提升辨識結果 (3/3)

Before feature Selection:
Geometry + Radon+ Miscellaneous (306 dim)
Test accuracy = 94.2 %
Remove "none" test accuracy = 91.0 %

	Center	Loc	Near-full	Random	none
Center	92.2% (118)	1.6% (2)	0	5.5% (7)	0.8% (1)
Loc	2.9% (29)	89.7% (886)	0	5.3% (52)	2.1% (21)
Near-full	0	4.7% (2)	81.4% (35)	4.7% (2)	9.3% (4)
Random	0.8% (10)	6.5% (82)	0	92.2% (1161)	0.5% (6)
none	0.4% (34)	3.2% (285)	0.3% (25)	1.1% (100)	95.0% (8478)

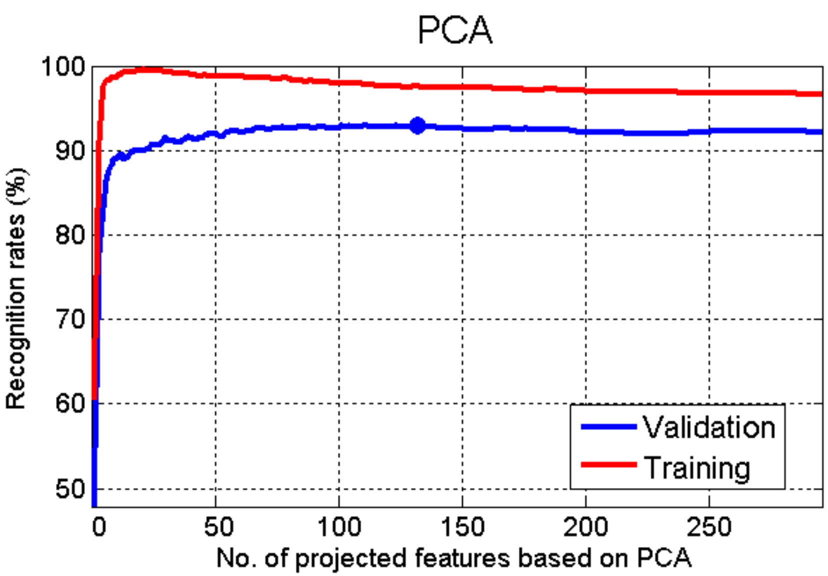
After feature selection:
Selected feature (163 dim)
Test accuracy = 94.4 %
Remove "none" test accuracy = 92.1%

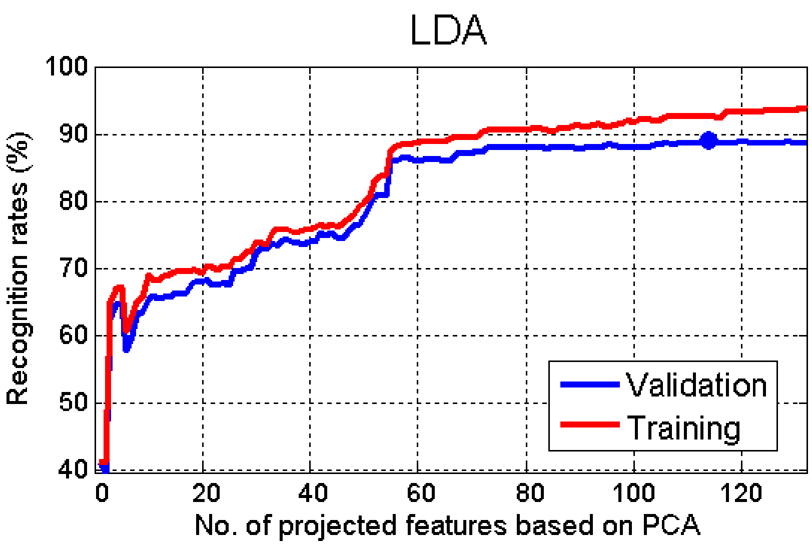
	Center	Loc	Near-full	Random	none
Center	94.5% (121)	1.6% (2)	0	3.1% (4)	0.8% (1)
Loc	2.2% (22)	91.2% (901)	0	4.5% (44)	2.1% (21)
Near-full	0	2.3% (1)	81.4% (35)	7.0% (3)	9.3% (4)
Random	0.7% (9)	5.9% (74)	0	92.9% (1170)	0.5% (6)
none	0.3% (28)	3.1% (281)	0.3% (25)	1.2% (110)	95.0% (8478)

特徵種類	Geometry	Radon	Miscellaneous	Total
特徵選取前(Dim)	198	92	16	306
特徵選取後(Dim)	103	51	9	163

實驗四
降維實驗 (1/2)

- 實驗資料：原先訓練資料(11370)切半，分別為訓練資料、驗證資料(各5685)
- 降維方式與結果：
  - Principal component analysis (PCA) —將資料作壓縮 (306 → 132 dim)
  - Linear discriminant analysis (LDA) — 將不同資料點分開 (132→ 114 dim)





23

實驗四
降維實驗 (2/2)

Before dimensionality reduction
Geometry + Radon+ Miscellaneous (306 dim)
Test accuracy = 94.2 %
Remove “none” test accuracy = 91.0 %

	Center	Loc	Near-full	Random	none
Center	92.2% (118)	1.6% (2)	0	5.5% (7)	0.8% (1)
Loc	2.9% (29)	89.7% (886)	0	5.3% (52)	2.1% (21)
Near-full	0	4.7% (2)	81.4% (35)	4.7% (2)	9.3% (4)
Random	0.8% (10)	6.5% (82)	0	92.2% (1161)	0.5% (6)
none	0.4% (34)	3.2% (285)	0.3% (25)	1.1% (100)	95.0% (8478)

After dimensionality reduction (PCA+LDA)
Projected features (114 dim)
Test accuracy = 95.9 %
Remove “none” test accuracy = 91.1 %

	Center	Loc	Near-full	Random	none
Center	85.2% (109)	9.4% (12)	0	4.7% (6)	0.8% (1)
Loc	1.9% (19)	91.7% (906)	0	4.4% (43)	2.0% (20)
Near-full	0	4.7% (2)	79.1% (34)	4.7% (2)	11.6% (5)
Random	0.5% (6)	7.3% (92)	0	91.7% (1155)	0.5% (6)
none	0.2% (18)	3.0% (268)	0.0% (2)	1.1% (101)	95.6% (8533)

(單位：秒)	降維實驗前 (306 dim)	降維實驗後 (114 dim)
訓練時間	36.00	10.20
測試時間	26.21	5.04

實驗結果分析
將所有實驗特徵加入後

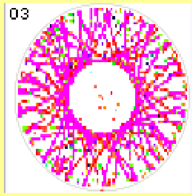
特徵值種類	維度	辨識率結果（%）（Remove “none” test accuracy）
2x2 binary pattern(15)	15	67.2
Corner detection (1)	1	42.0
2x2 binary pattern(15) + Corner detection (1)	16	69.5
Geometry (198) + Radon (92) + 2x2 binary pattern (15)	305	88.8
Geometry (198) + Radon (92) + Corner detection (1)	291	90.7
Geometry (198) + Radon (92) + 2x2 binary pattern (15) + Corner detection (1)	306	91.0
After feature selection	163	92.1
After dimensionality reduction	114	91.1

	所需時間 (s)	平均時間 (s/per wafer)
特徵擷取時間	4860.4	0.2140
訓練時間	34.1	0.0030
測試時間	24.94	0.0022

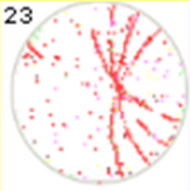
Other Prediction Error for Small-die Wafers

- **Diversity of appearances**

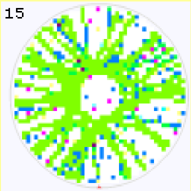
Scratch



Scratch

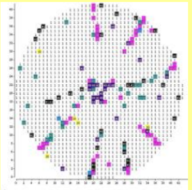


Scratch

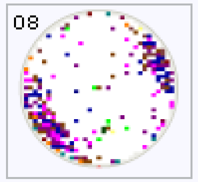


- **Lack of samples**

Radiation

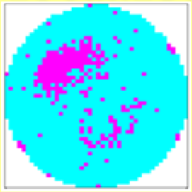


Top bottom



- **Confusing wafers**

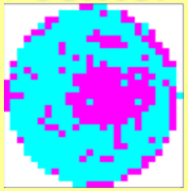
Localized



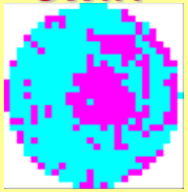
Donut

- **Inconsistent human labels**

Center



Swirl



本研究對TSMC的實際效益

- 以台積電為例，2013年第一季生產的晶圓高達388萬片，平均每日製造的晶圓約為4.3萬片，人工觀察晶圓圖勢必過於耗時。我們的系統也相當有效率，以一台個人電腦的運算能力而言 ，只需約1.4個小時即可處理一日的產能。
- 經過專家評估，假設本系統每天將替工程師節省一小時工時，假設平均每天有100個工程師使用 ，則一年約250個工作天即可省下25000小時，每年創造的經濟效益約100萬美元。
- 目前此項技術已經在台積電上線(計畫的上線率低於10%)，各廠生產的晶圓都會透過本系統來進行檢查，證明了本系統的有效性與可靠性。

Current Status

- 獲得「2014 IPPR技術創新暨產業應用獎」優等獎
- 論文發表
  - Ming-Ju Wu, Jyh-Shing Roger Jang, and Jui-Long Chen, "Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-scale Datasets", IEEE Trans. on Semiconductor Manufacturing, 2015.
  - 晶圓錯誤樣式辨識的設計與改進 (Design And Improvement of Wafer Failure Pattern Recognition) ，林坤優，清華大學碩士論文，2014.

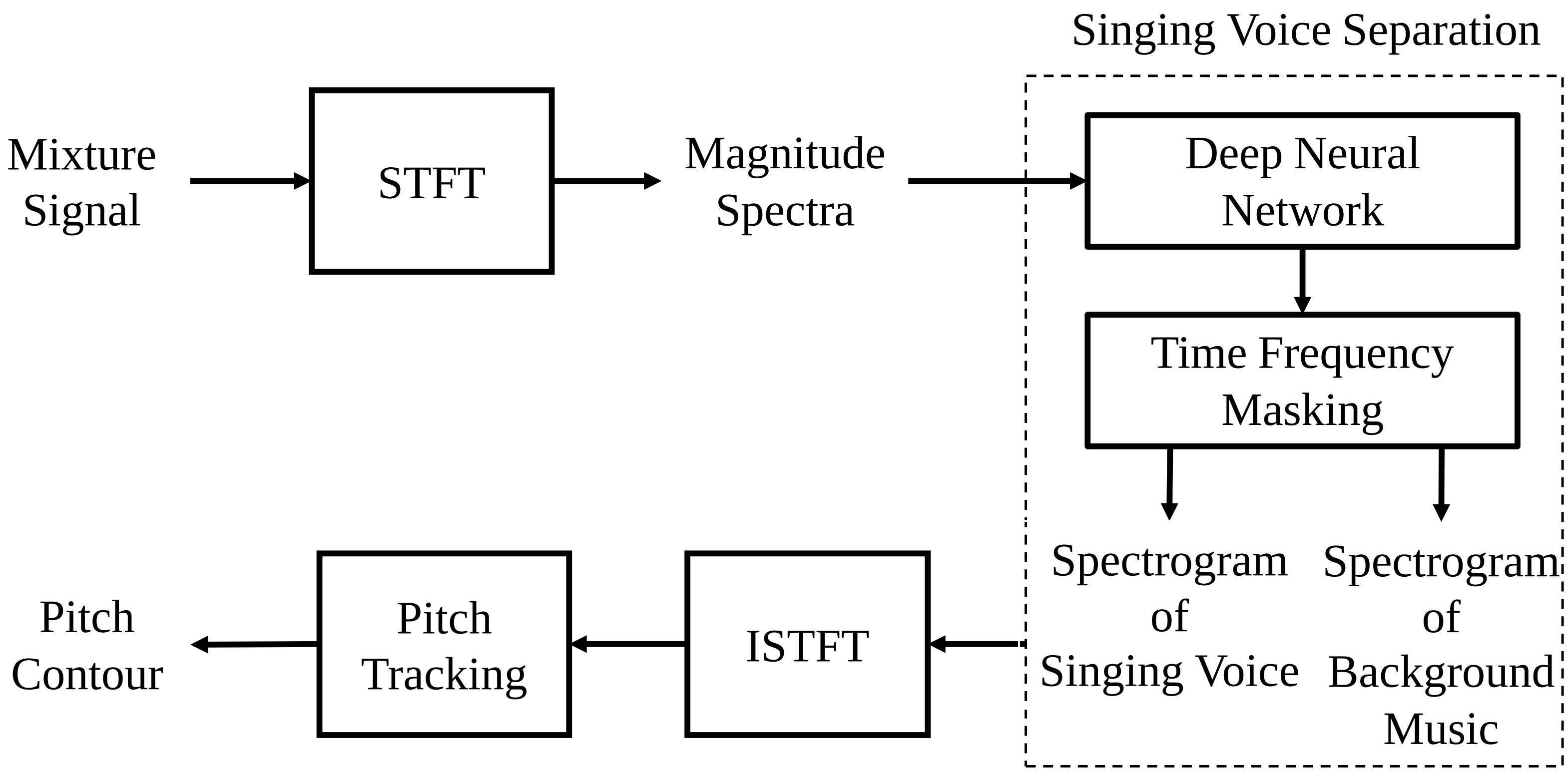
Singing Voice Separation and Pitch Extraction from Monaural Polyphonic Audio Music Via DNN and Adaptive Pitch Tracking

張智星
Multimedia Information Retrieval Lab
Dept. of CSIE, National Taiwan University

Introduction

- Singing voice separation (SVS)
  - Extract singing voice from polyphonic audio music
  - Applications
    - Singing identification
    - Lyrics recognition & alignment
    - Singing assistance for karaoke
- Audio melody extraction (AME)
  - Extract major pitch from polyphonic audio music
  - Applications
    - Cover song identification
    - Database construction for query by singing/humming
    - Singing scoring
- Proposed approaches
  - DNN for SVS, adaptive-UPDUDP for AME

Overall System Flowchart

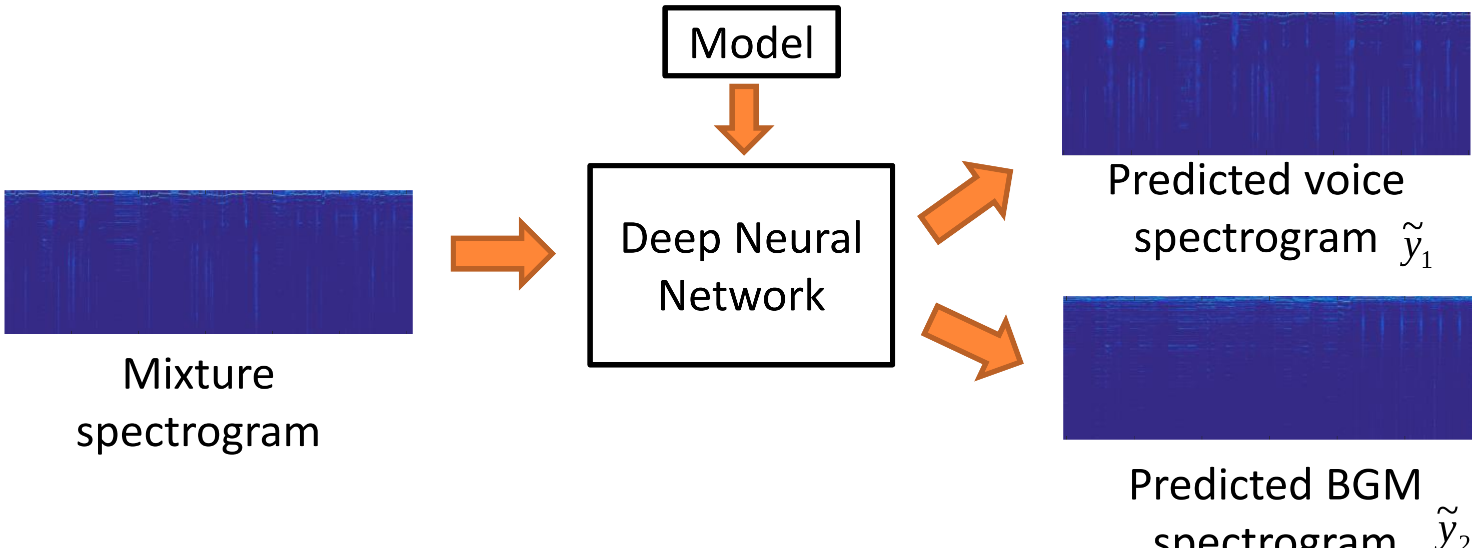


Steps in Singing Voice Separation

- Convert music into spectrogram
  - Short-time Fourier transform (STFT)
  - Mixture, voice, background music (b.g.m.)
- Deep neural network (DNN)
  - Nonlinear mapping from mixture spect. to vocal and b.g.m. spect.
  - Activation function : Sigmoid
  - Objective function : Square error
  - Gradient Optimization : RMSProp
  - Dropout rate : 0.5
  - GPU used : true
- Convert voice spectrogram into music
  - Inverse short-time Fourier transform (ISTFT)

Details of Singing Voice Separation

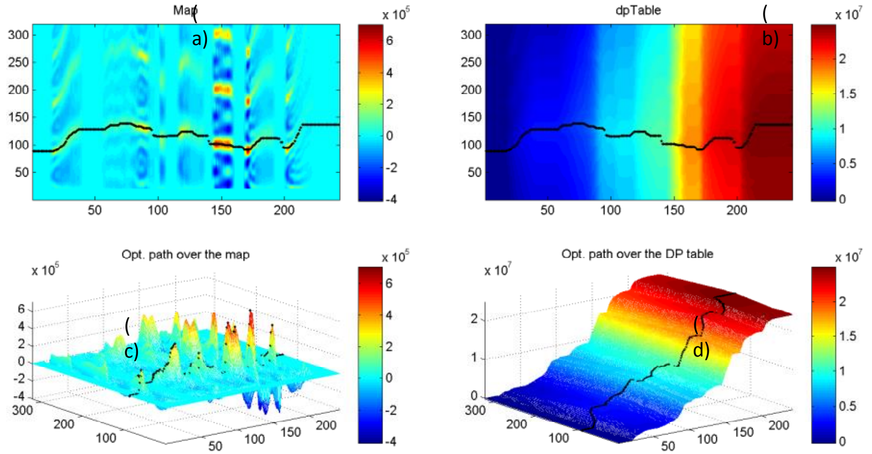
- Vocal voice reconstruction
  - $m(f) = |\tilde{y}_1(f)| / (|\tilde{y}_1(f)| + |\tilde{y}_2(f)|)$ f =1, 2, … F (frequency bins)
  - $\tilde{s}_1(f) = m(f)z(f)$
    $\tilde{s}_2(f) = (1 - m(f))z(f)$, where *z(f)* is mixture spectrogram
  - $\tilde{s}_1$: estimated voice spectrogram
  - $\tilde{s}_2$: estimated background music spectrogram



Method for Singing Pitch Extraction

- Unbroken Pitch Determination Using Dynamic Programming (UPDUDP)
  - Average magnitude difference function (AMDF)
    - Calculate each frame's AMDF vector
  - Objective function

$$\mathrm{cost}\left(\mathbf{p},\theta,m\right)=\sum_{i=1}^{n} amdf_i\left(p_i\right)+\theta\times\sum_{i=1}^{n-1}\left|p_i-p_{i+1}\right|^m \quad i=1..n \quad \text{frames}$$

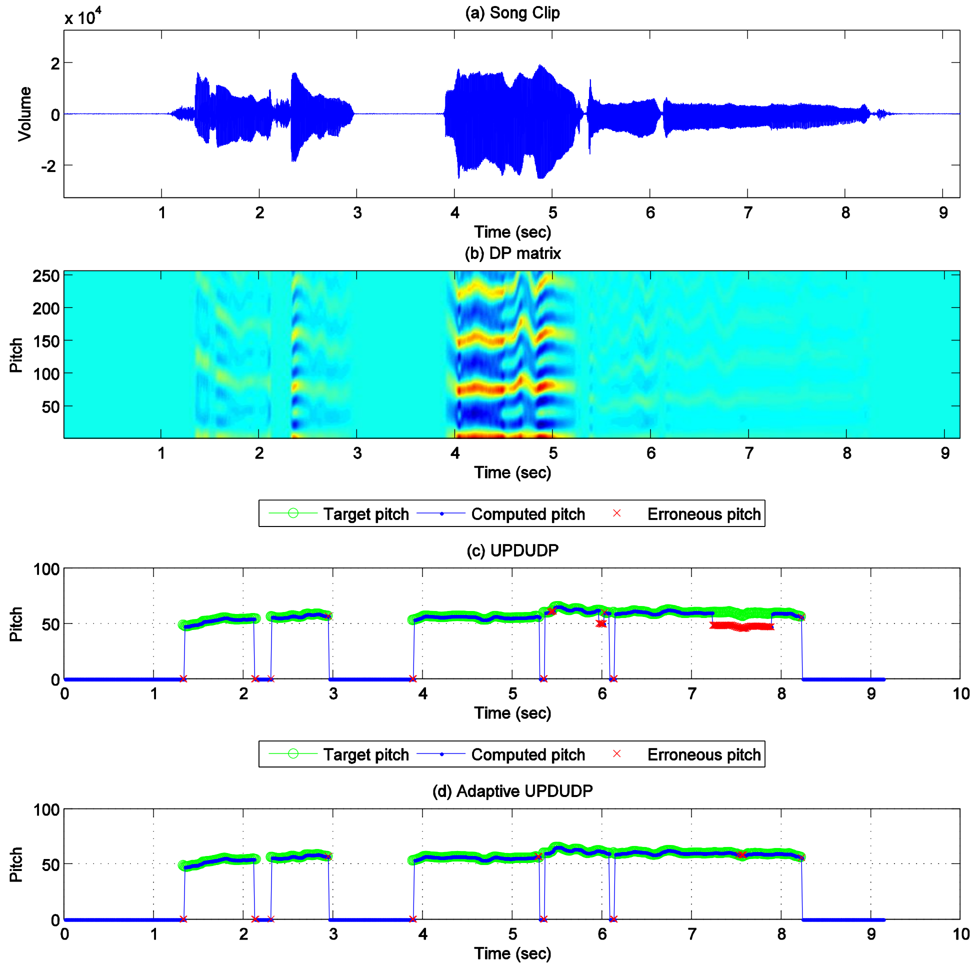


J. C. Chen and J.S. Jang, "TRUES: Tone Recognition Using Extended Segments," in ACM Transactions on Asian Language Information Processing, No. 10, Vol. 7, Aug 2008.

Adaptive UPDUDP

- Adaptive UPDUDP for finding a good ϑ
  - Goal : Find the proper ϑ
  - Target equation: $d(\theta) = \max\limits_{i=1\sim n-1} |s_i - s_{i+1}| < \tau$
    - Pitch curve : $s = [s_1, \cdots s_i, \cdots s_n]$ (semitones)
    - τ : 7 semitones
  - Algorithm : identify the interval $\hat{I} = [\theta_l, \theta_u]$ to satisfy the condition $d(\theta_l) \geq \tau$ and $d(\theta_u) \leq \tau$
    - if d(ϑ) == 0 , we are done.
    - (a) Set $I_0 = [\theta_0, \theta_1] = [0,1]$. If the bracket condition is fulfilled, then we are done with $\hat{I} = I_0$. Otherwise set i=1 go to the next step.
      (b) Set $I_i = [\theta_i, \theta_{i+1}]$ with $\theta_{i+1} = 2\theta_i$.
      (c) If the bracket condition for I_i is fulfilled, then we are done with $\hat{I} = I_i$. Otherwise increment i and go back to step b.

Example of Adaptive UPDUDP



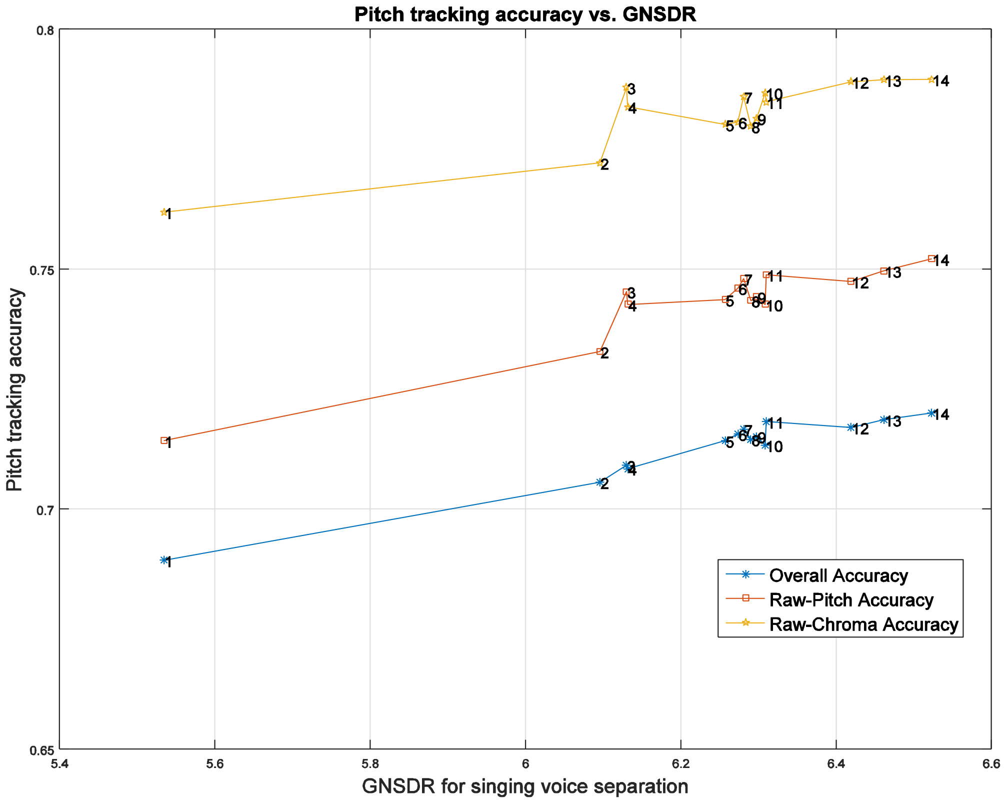
Corpora for Experiments

- MIR-1K dataset
  - 1000 song clips
  - Durations : 4-13 secs
  - Sample rate : 16K Hz
  - 110 Chinese popular karaoke songs
  - Sung by male and female amateurs

- iKala dataset
  - 252 song clips (public set)
  - Durations : 30 secs
  - Sample rate : 44.1K Hz
  - Sung by male and female professionals

AME vs. SVS

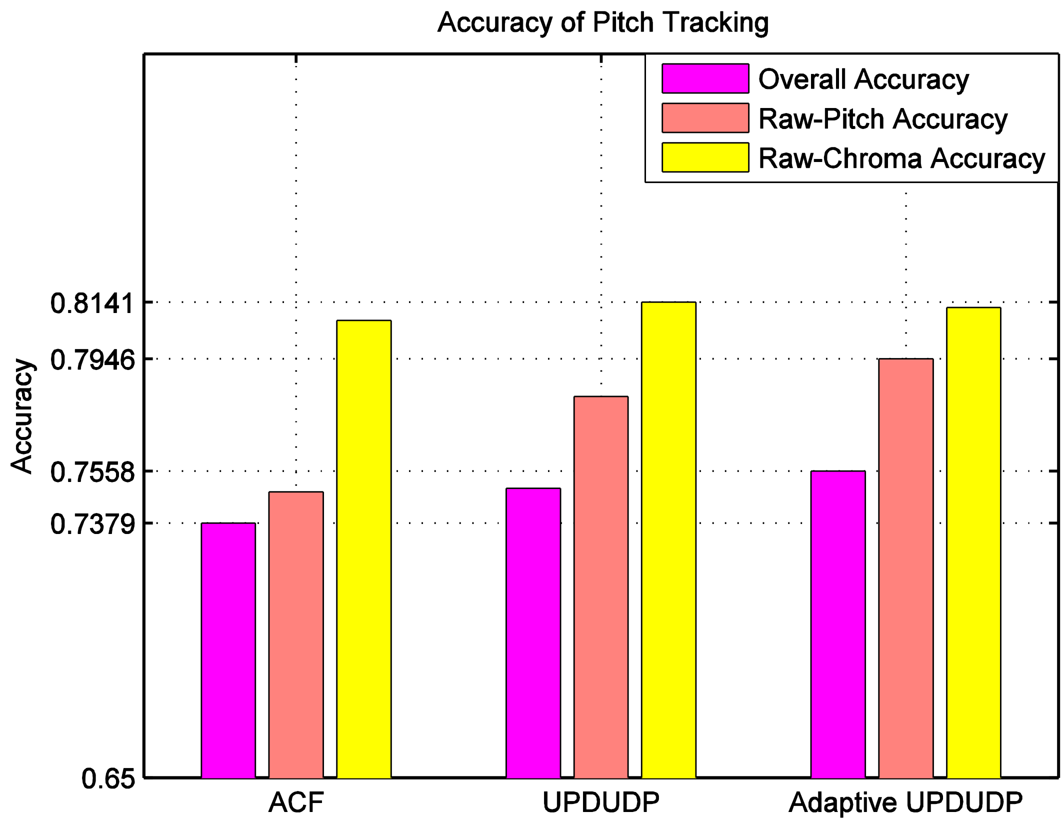
- Dataset : MIR-1K
  - 175 clips for training, 825 clips for testing
  - Method : DNN+ UPDUDP



Indices of DNN architecture	Numbers of nodes in each hidden layer	Numbers of hidden layers
1	64	3
2	128	1
3	128	2
4	256	2
5	1024	1
6	2048	1
7	512	2
8	768	1
9	512	1
10	256	3
11	768	2
12	512	3
13	768	3
14	1024	3

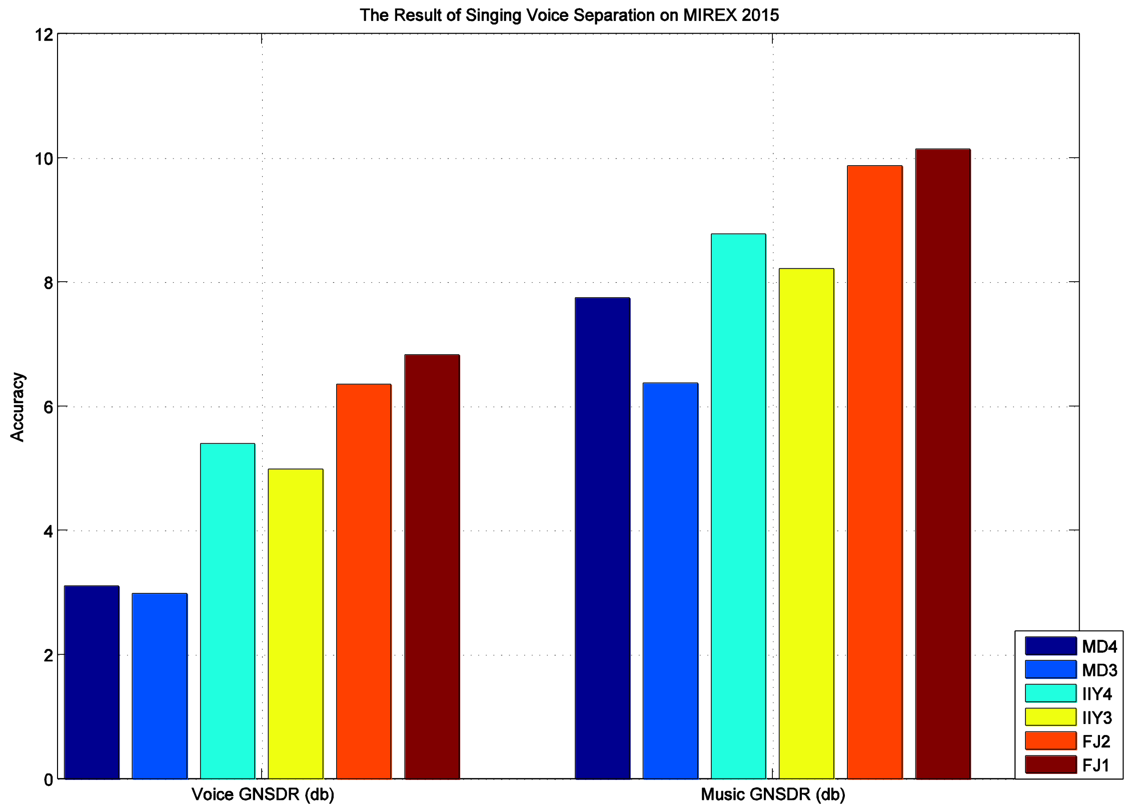
Experimental Results of SVS

- Dataset : MIR-1K
  - 175 clips for training, 825 clips for testing
  - Compare different pitch-tracking method on extracted vocal



SVS Results for MIREX 2015

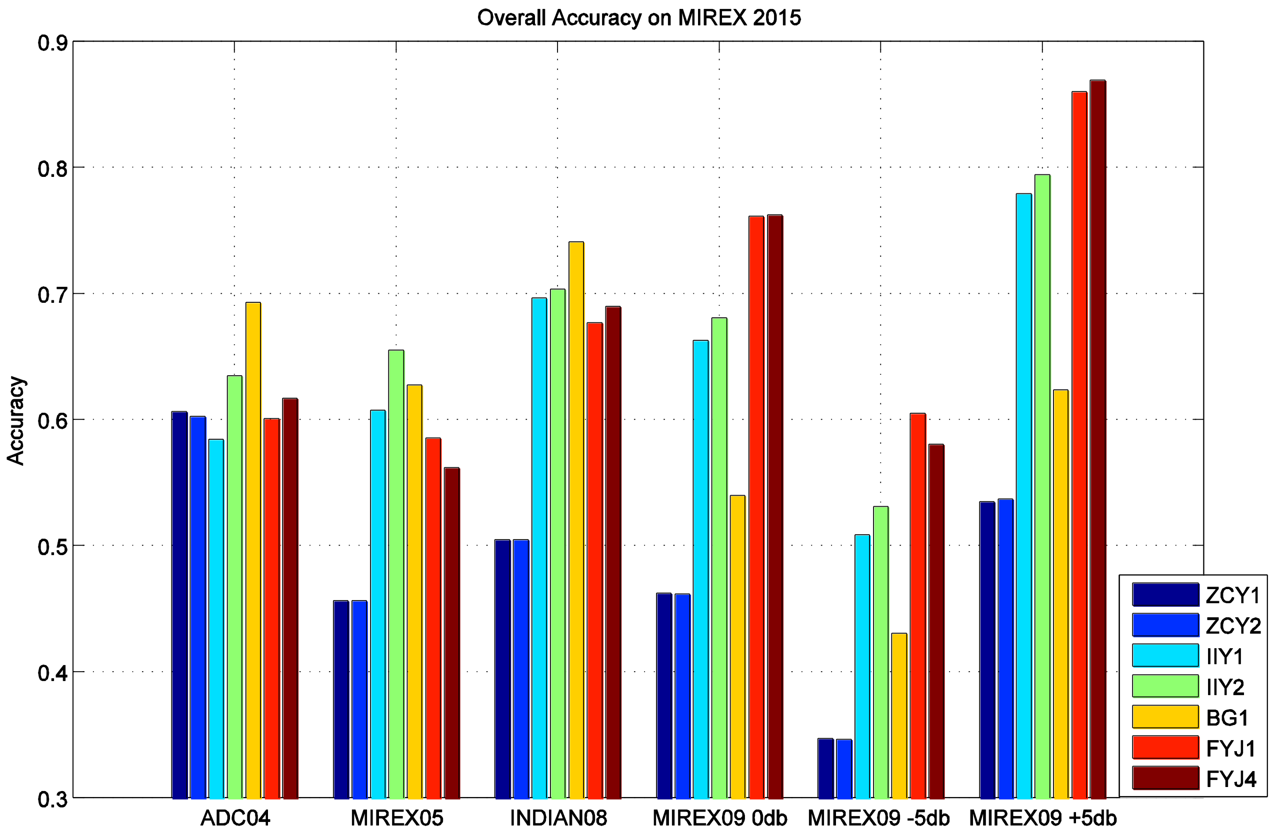
- Singing voice separation on MIREX 2015
  - Choose 126 clips from iKala dataset for training randomly



Settings	FJ1	FJ2
Prog. language	Matlab	
Initial learning rate	0.001	
Dropout fraction	0.5	
Activation function	Sigmoid	
Cost function	Square error	
Gradient optimization	RMSProp	
Num. of nodes / layer	1024	
Num. of hidden layers	3	1
Training data	126 files from the public set of iKala.	
Training time	7.70 hours	1.73 hours

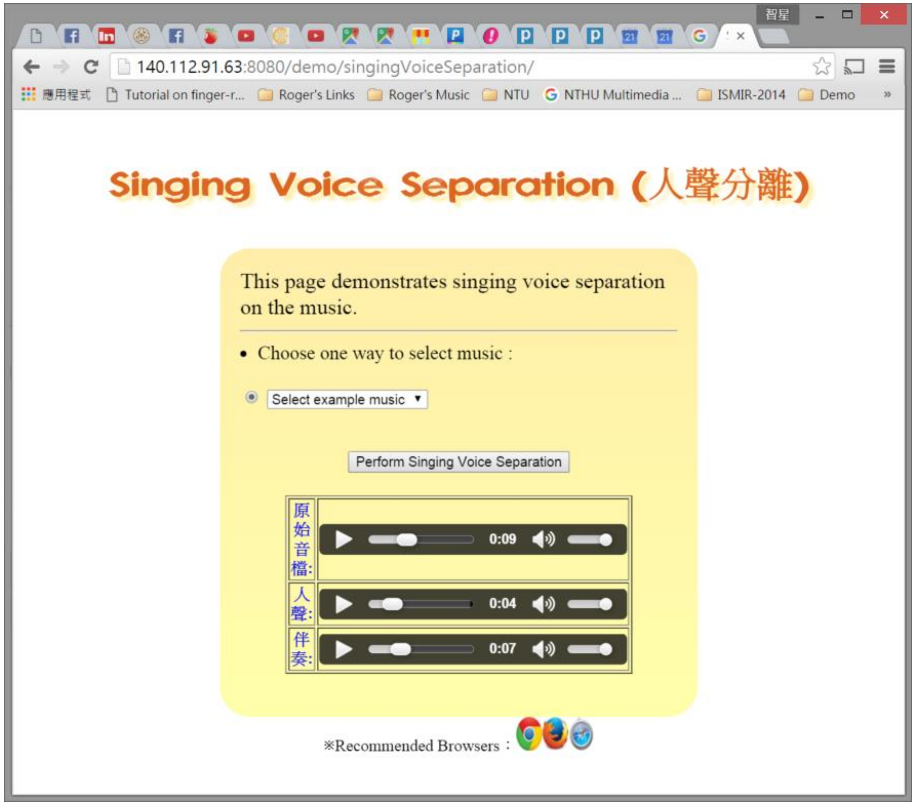
AME Result for MIREX 2015

- Competition result of audio melody extraction
  - FYJ1=DNN+UPDUDP
  - FYJ4=DNN+adaptive UPDUDP



Demo of Singing Voice Separation

- Online demo of our singing voice separation
  - http://mirlab.org/demo/singingVoiceSeparation



Conclusions

- Data science is on the rise!
  - Big data, machine learning, data mining…
- Thanks go to …
  - TSMC
  - Terasoft