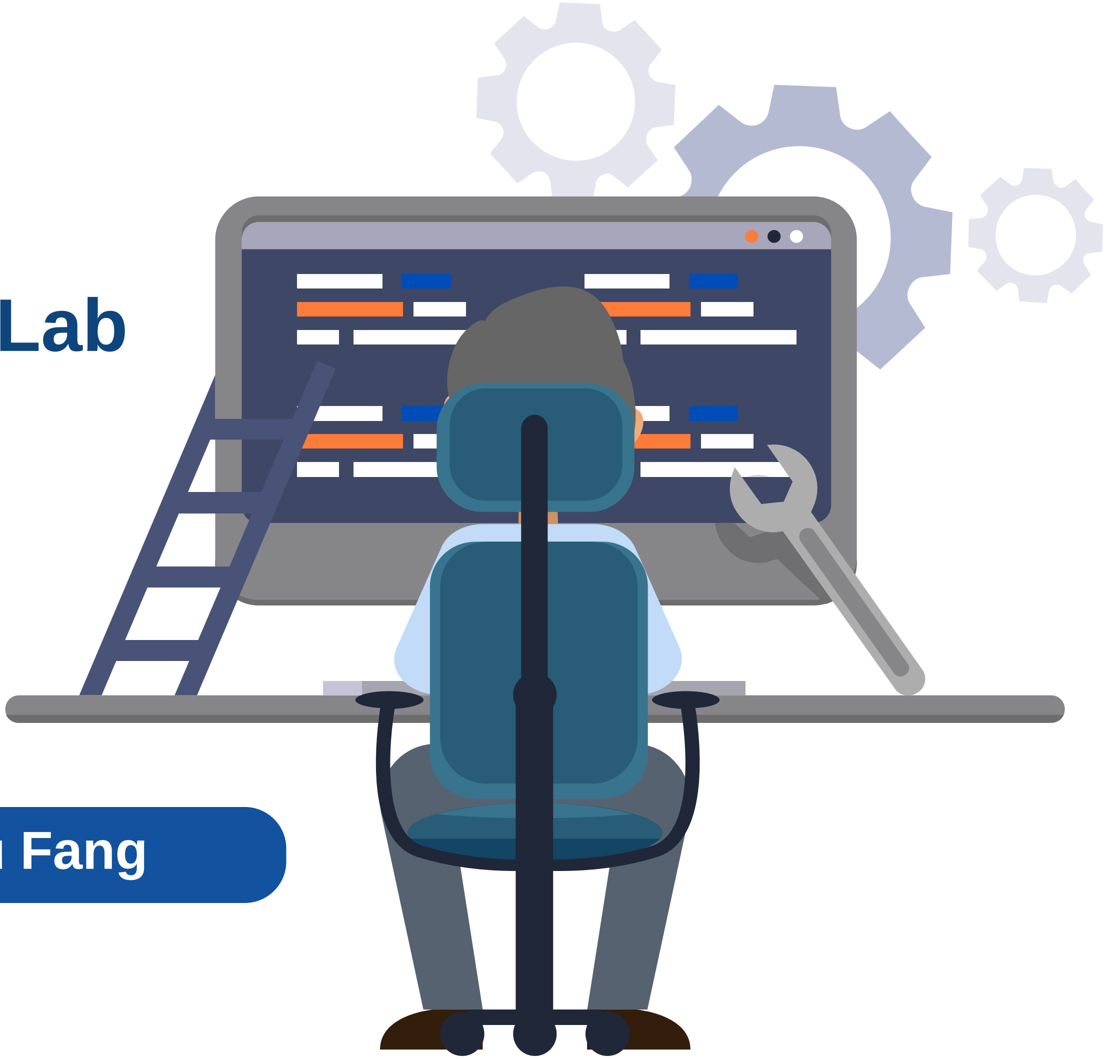
Digital Signal Processing Lab (DSP)

DSP Project 1

Electrical Department Pro. Shih-Hau Fang



Acoustic Sensing Apps_Scoring Standard

- 3 Reports need uploaded to moodle in .pdf (80%)
  - (30%) Labs Practice (8-th week)(Including Lab0-Lab6 in one pdf file)
  - (25%) Project 1 (11-th week)
  - (25%) Project 2 (14-th week)
- ~~1 Midterm Exam (20%)~~
  - ~~(20%) For Labs Practice (8-th week): On-computer-test~~
- 1 Final Presentation (20%)
  - (20%) Paper reading Assignment (the 15-16 week): Each person finds one paper about acoustic sensing.(Any journal or conference are acceptable). Presenting in mandarin is fine, using English will receive bonus score.

Course calendar

Month	Date	Notes
September	Week 1 9/4	• Outline and overview, DSPLab0
	Week 2 9/11	• DSPLab 1 (Basic operation and use)
	Week 3 9/18	• DSPLab 2 (Basic Graphics and Signal Processing)
	Week 4 9/25	• DSPLab 3 (Application of FFT conversion method and image processing)
October	Week 5 10/2	• Typhoon Bye Week
	Week 6 10/9	• DSPLab 4 (LTI and Convolution)
	Week 7 10/16	• DSPLab 5 (basic audio processing)
	Week 8 10/23	• DSPLab 6 (Convolution and Filter) **(Turn in the Lab0-6 in one pdf)**
	Week 9 10/30	• Project 1(Pathological Voice)
November	Week 10 11/6	• Project 1(Pathological Voice)
	Week 11 11/13	• Project 1(Pathological Voice) (Turn in project 1)
	Week 12 11/20	• Project 2 (Noise/Echo)
	Week 13 11/27	• Project 2 (Noise/Echo)
December	Week 14 12/4	• Project 2 (Noise/Echo) (Turn in project 2)
	Week 15 12/11	• Final Presentation
	Week 16 12/18	• Final Presentation

Tutorial Videos

- AI cup：病理嗓音檢測(病史與音檔)
- DSP專題一：以KNN實現病理嗓音偵測實作

Signal Processing

1 **Facebook of IEEE signal processing society**

- https://www.facebook.com/ieeeSPS

2 **What is signal processing**

- https://youtu.be/EErkgr1MWw0

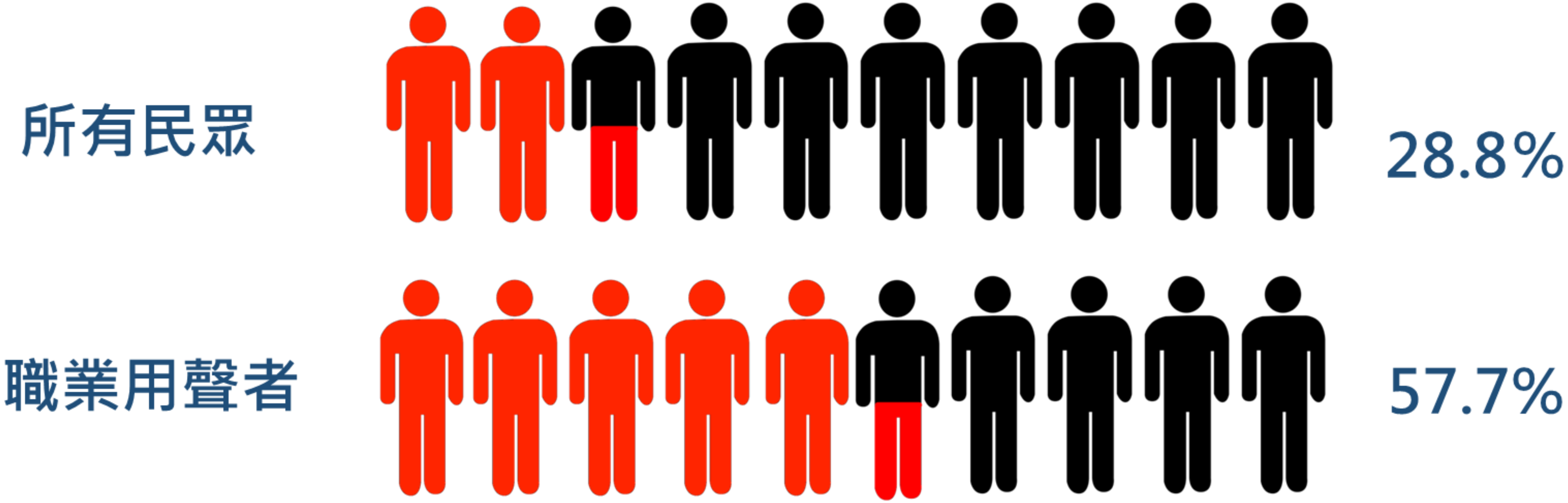
3 **Review**

- Digital audio/Speech signal processing (a typical case)
- Matlab (6 weeks)

Some cool products/projects

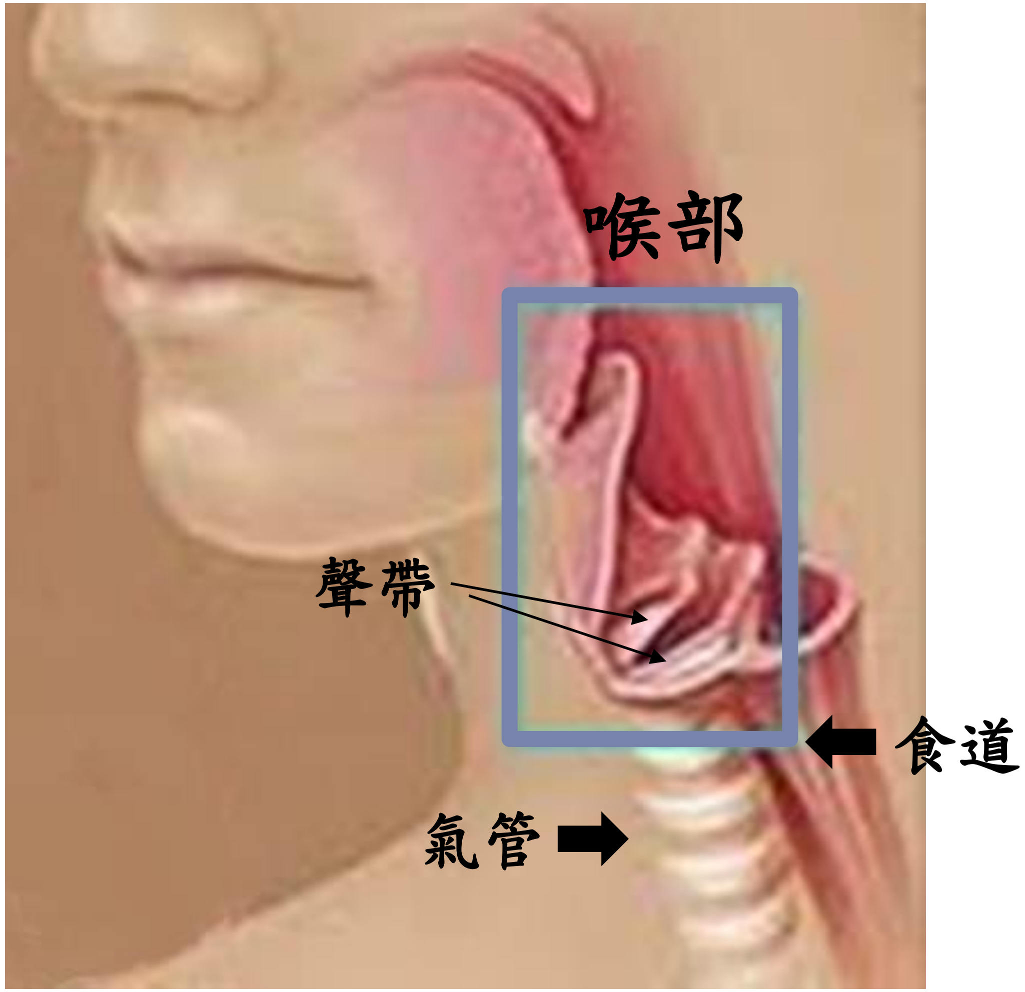
1. 藍芽無線頭巾式耳機
2. 微型無線樂器麥克風
3. "Man vs Robot" in playing Angry Birds
4. SILENT PARTNER 消除打呼聲的眼罩
5. Spear X 聲特耳機
6. MEATER 無線智能溫度計
7. Adobe Project VoCo - 音檔不用重新錄製就能增加新句子
8. Speech Translation (翻譯米糕)

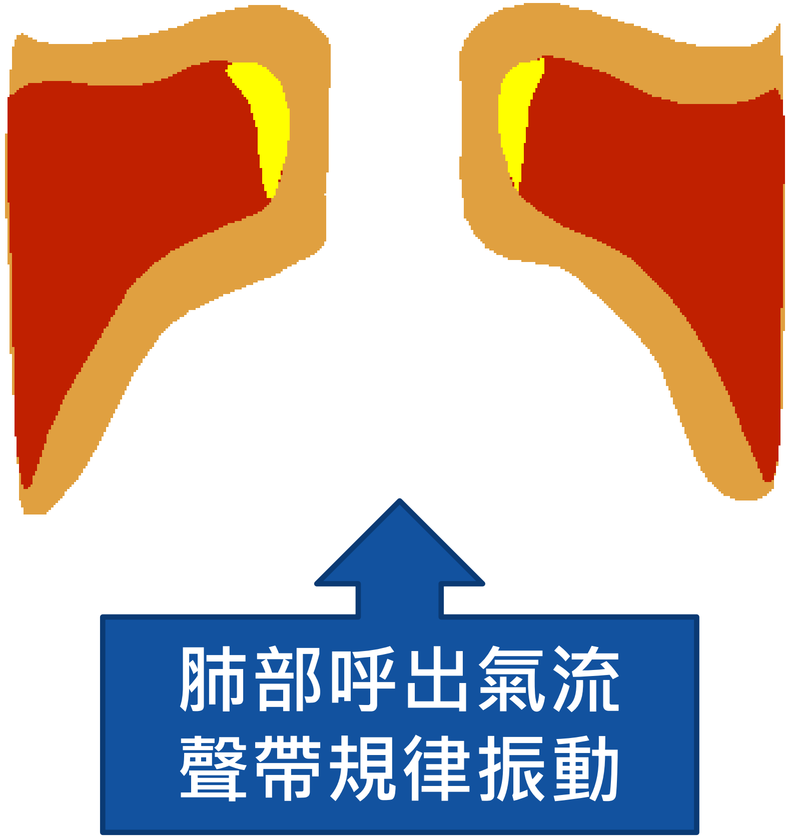
Preview



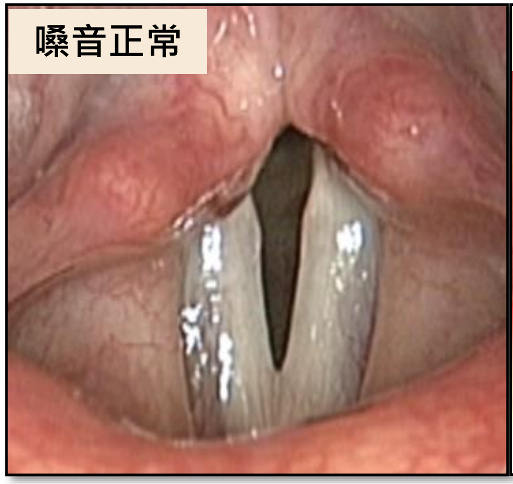
調查研究指出，所有民眾一生中有接近30%曾經經歷過明顯的嗓音障礙，這個比率，在職業用聲者，如老師、業務等，更高達57.7%。

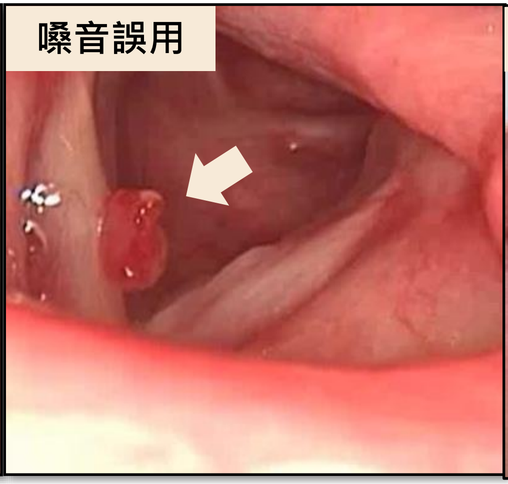
How does sound come from

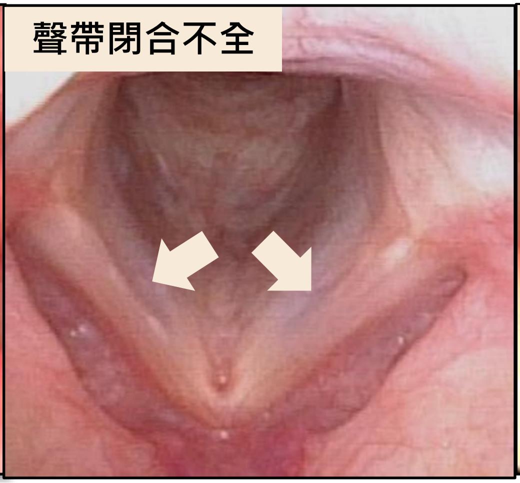


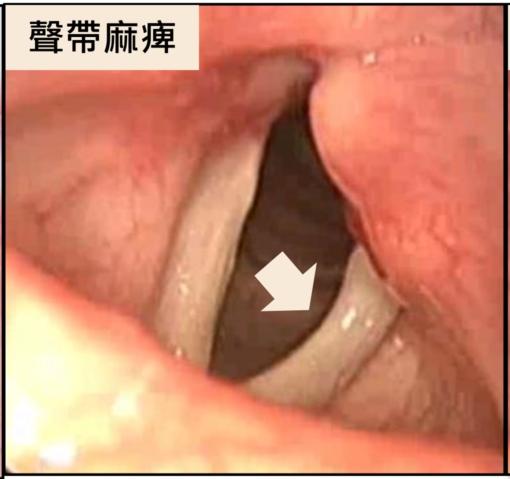

肺部呼出氣流
聲帶規律振動

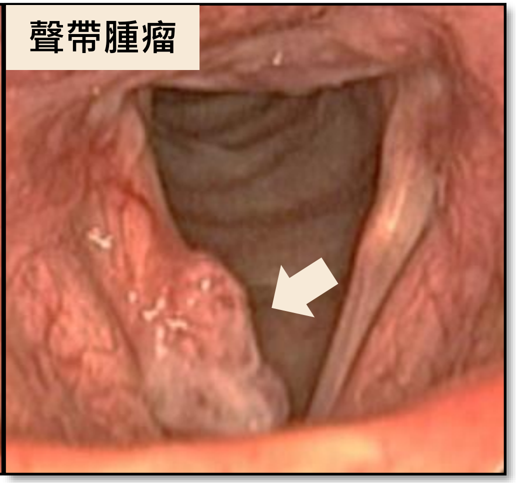
Acoustic illness


嗓音正常


嗓音誤用


聲帶閉合不全


聲帶麻痺


聲帶腫瘤

正常的聲帶，則可能因為使用不當，產生息肉，或是長期抽菸造成癌症病變，又或者是因為腫瘤或手術傷及聲帶的神經損，造成聲帶麻痺萎縮，進而影響了聲音的品質。

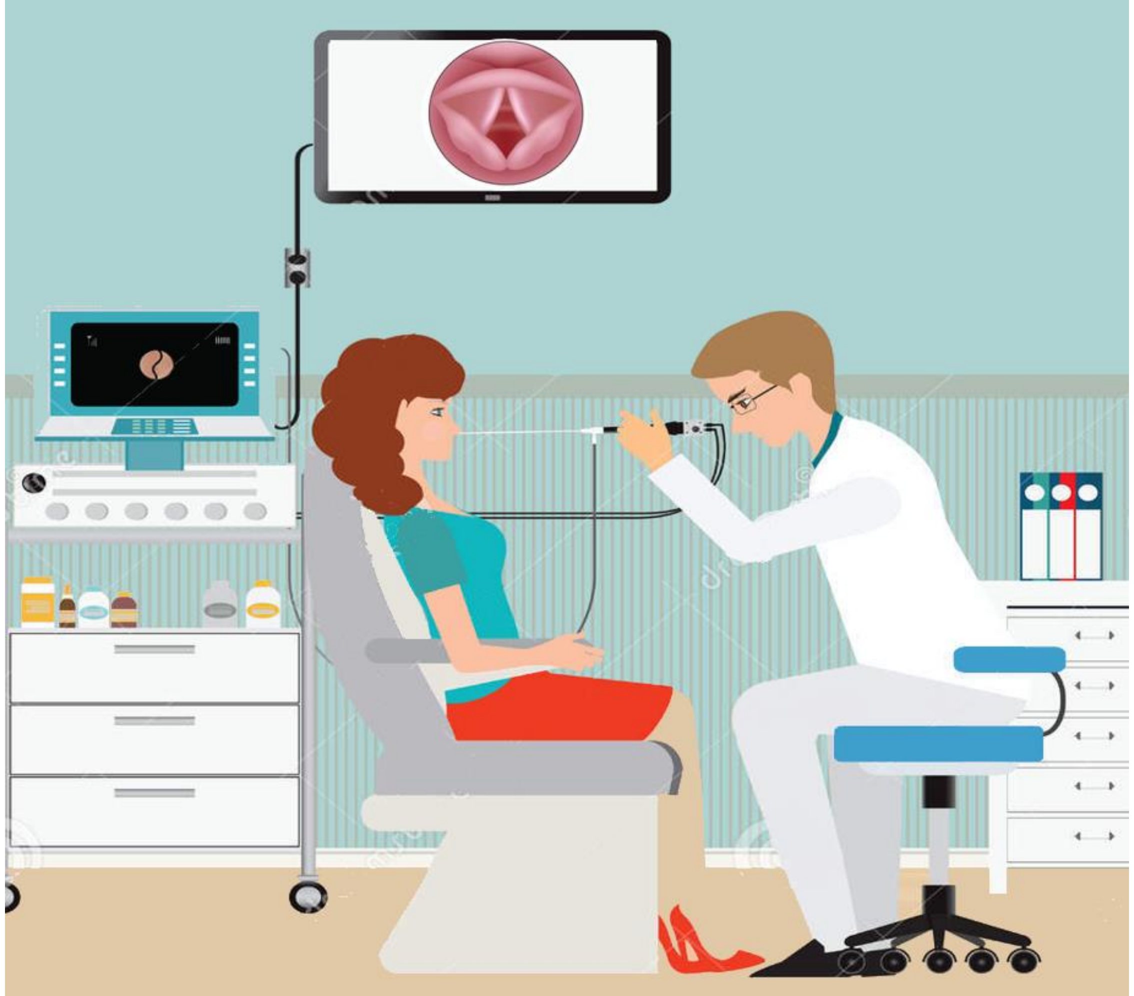
Consultation procedure

問診
病史紀錄

聽聲
錄音紀錄

內視鏡
診斷標準

Practical use



Purpose

- ★ **<u>Use speech analysis technology to detect pathological characteristics in the voice to examine whether acoustic analysis technology can be applied to patients who are experiencing discomfort</u>**

1. This project learns the K-Nearest Neighbor (KNN) algorithm
2. Use of Mel-scale Frequency Cepstral Coefficients (MFCC) tool
3. Conduct simulation and discussion results of speech detection on Matlab

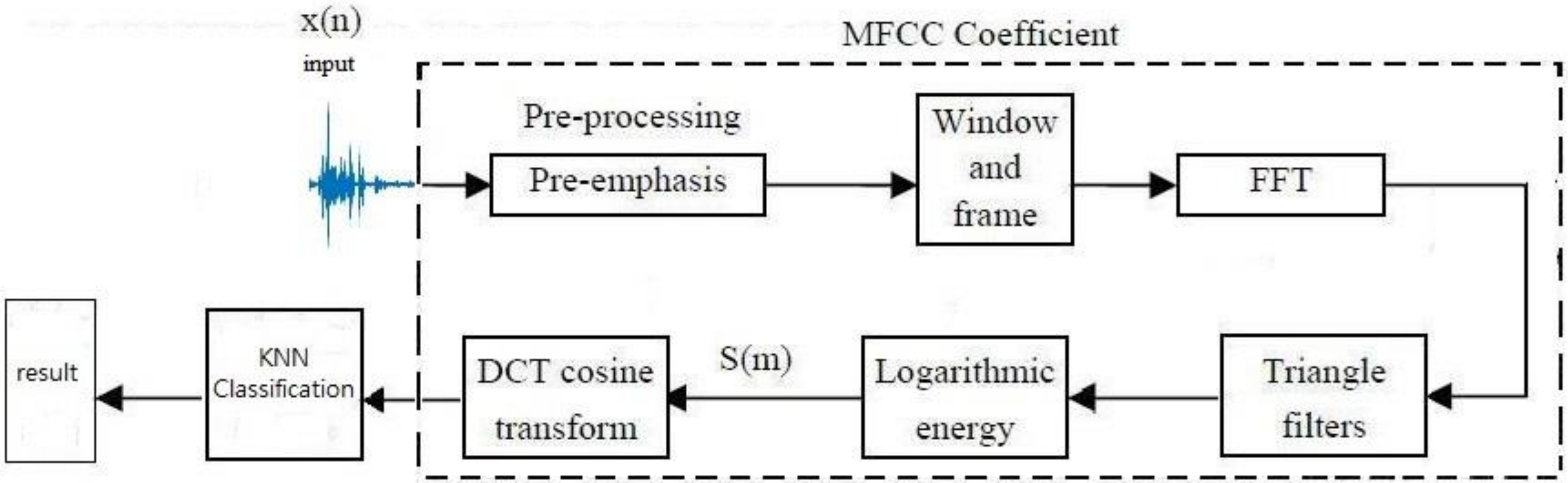
Data Description

Provides 40 male and female mixed sound wav files

1. **Voice content: Single sound "Ah"**
2. **Voice time range: 4~30 seconds**
3. **20 male and female mixed sick wav files**
   - Two types of lesions related to the throat: cancer and CYST
4. **20 male and female mixed normal wav files**

MFCC

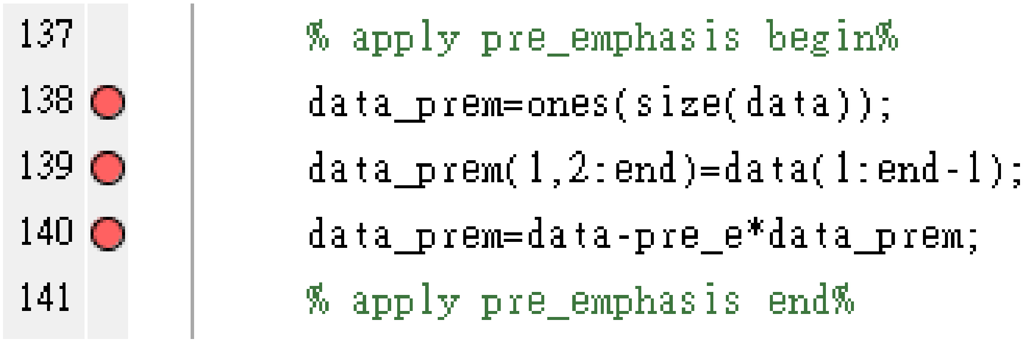
MFCC Flow Chart



Understand the Details of MFCC Execution

1 Before executing mfcc, place breakpoints at each step in mfcc_v2

Step 1	Lines 138~140
Step 2	Lines 186~189
Step 3	Line 194
Step 4	Lines 198~201
Step 5	Line 205
Step 6	Lines 212~214



```
137     % apply pre_emphasis begin%
138 ●   data_prem=ones(size(data));
139 ●   data_prem(1,2:end)=data(1:end-1);
140 ●   data_prem=data-pre_e*data_prem;
141     % apply pre_emphasis end%
```

2 Observe the Workspace and the program code to understand the new values and details when the execution reaches this step.

Step 1、Pre-emphasis

1 Send a piece of sound signal x(n) to a high-pass filter, the formula is:

$$x_2(n) = x(n) - a * s(n-1)$$

2 Where a is a constant between 0.9 and 1.0, filtered by the z transform to get:

$$\mathrm{H(z)} = 1 - \mathrm{a} * z^{-1}$$

3 Its purpose is to remove the effects of the vocal cords and lips during phonation to compensate for the high-frequency part of the speech signal that is suppressed by articulation.

Step 2、Window and Frame

1 The input voice signal is divided into sound frames of 20ms to 30ms, and there is an overlapping area of 1/3 to 1/2 the size of the sound frame between two sound frames. In order to avoid excessive changes in adjacent sound frames and facilitate the use of subsequent FFT, generally speaking, the audio sampling used in speech recognition the frequency (sampling frequency) is 8 kHz or 16 kHz.

2 Taking 8 kHz as an example, assuming the sound frame length is 200 sampling points, the corresponding time length is:

$$\frac{200}{8000} = 0.025\text{s} = 25\text{ ms}$$

3 In addition, if there are 100 points in the overlapping area, the frame rate will be

$$\frac{8000}{200-100} = 80\ {}^{\text{frames}}\!/_{\text{s}}$$

Step 3、Fast Fourier Transform

- Since it is difficult to see changes in the signal in the time domain,
  it is necessary to observe the distribution in the frequency domain instead.
  , and different energy distributions can represent the characteristics of different speech sounds.

Step 4、Triangular Filters

1 Multiply the spectral energy by a set of 20 triangular filters to find the logarithmic energy of each filter output.

2 It must be noted that these 20 triangular filters are evenly distributed at the "Mel Frequency". The relationship between Mel frequency and general frequency f is as follows:

$$\mathrm{Mel}(f) = 2595 * \log_{10}(1+\frac{f}{700}) \quad \text{or} \quad \mathrm{Mel}(f) = 1125 * \ln(1+\frac{f}{700})$$

3 Mel frequency represents the human ear's perception of frequency.
From this, it can be observed that the human ear's perception of frequency f shows
a logarithmic change. For the low-frequency part, the human ear feels more sensitively;
for the high-frequency part, the human ear feels increasingly rough and slow.

Step 5、Logarithmic Energy

- The energy of a sound frame is an important feature of speech,
so it is usually added to the logarithmic energy of a sound frame (defined as a sound frame sum of the squares of the internal signals, then take the logarithmic value with base 10, and then multiply by 10), so that the speech characteristics of each sound frame have
1 term log energy.

Step 6、Discrete Cosine Transform

1 Enter the 20 logarithmic energies Ek from step 4 into the discrete cosine transform to find the 12th order Mel-scale Cepstrum parameters.

2 The discrete cosine transform formula is as follows:

$$C_m = \sum_{k=1}^{N} \cos(m * (k - 5) * \frac{N}{\pi}) * E_k \text{ ， where m = 1,2,3,……,L}$$

3 Among them, Ek is the inner product value of the triangular filter and the spectrum energy calculated in step 4, and N is the number of triangular filters.
Because Mel-Frequency was previously used to convert to Mel frequency coefficients,
it is called Mel-scale Frequency Cepstral Coefficients.

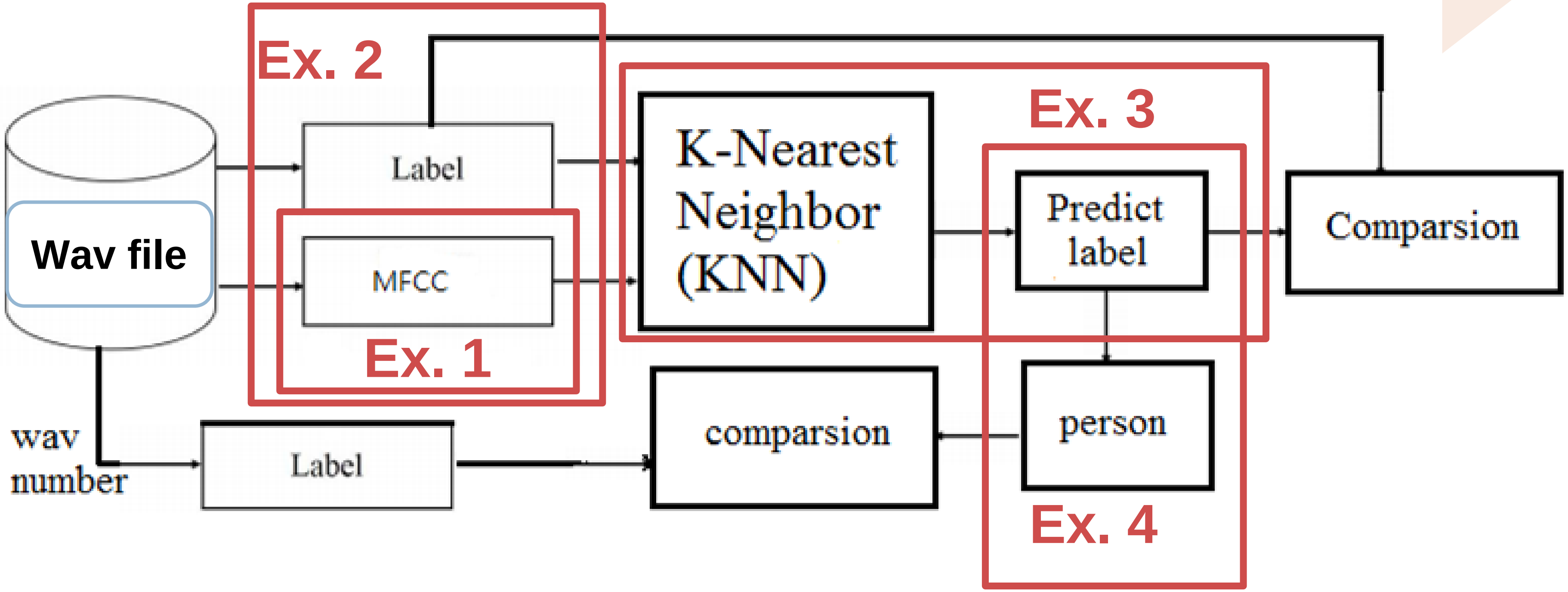
Experimental steps and processes

Experiment 1
Generate 26-dimensional MFCC

Experiment 2
String the data together and give it a label

Experiment 3
Classification using KNN algorithm

Experiment 4
Sound frame feature convert to people

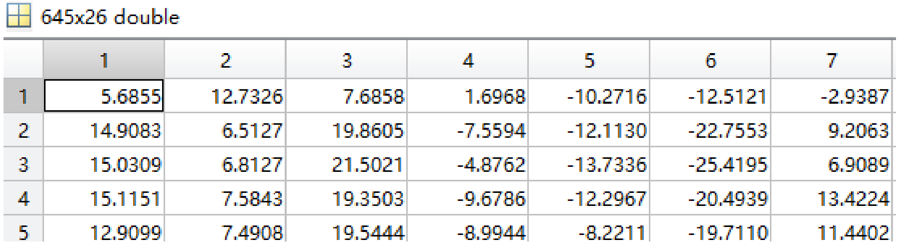


Experiment 1

Generate 26-dimensional (row) MFCC from the provided wav file

- **Use the mfcc_v2.m file parameters in the attachment to adjust**
  - sample_rate = 44100
  - frame_time = 30 (Unit：ms)
  - frame_move_time = 15 (Unit：ms)
  - Parameter frame_move_time = frame_time / 2 of mfcc_v2.m

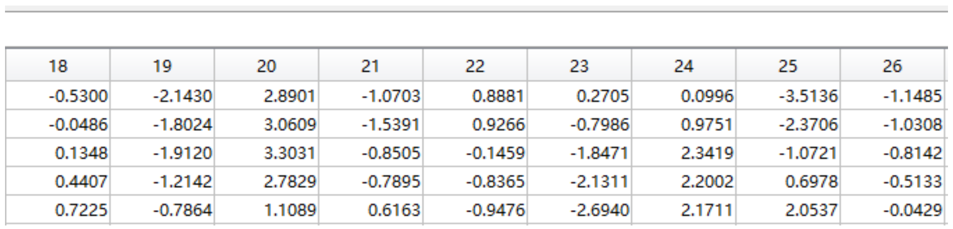
Reference results



645x26 double

	1	2	3	4	5	6	7
1	5.6855	12.7326	7.6858	1.6968	-10.2716	-12.5121	-2.9387
2	14.9083	6.5127	19.8605	-7.5594	-12.1130	-22.7553	9.2063
3	15.0309	6.8127	21.5021	-4.8762	-13.7336	-25.4195	6.9089
4	15.1151	7.5843	19.3503	-9.6786	-12.2967	-20.4939	13.4224
5	12.9099	7.4908	19.5444	-8.9944	-8.2211	-19.7110	11.4402

…



18	19	20	21	22	23	24	25	26
-0.5300	-2.1430	2.8901	-1.0703	0.8881	0.2705	0.0996	-3.5136	-1.1485
-0.0486	-1.8024	3.0609	-1.5391	0.9266	-0.7986	0.9751	-2.3706	-1.0308
0.1348	-1.9120	3.3031	-0.8505	-0.1459	-1.8471	2.3419	-1.0721	-0.8142
0.4407	-1.2142	2.7829	-0.7895	-0.8365	-2.1311	2.2002	0.6978	-0.5133
0.7225	-0.7864	1.1089	0.6163	-0.9476	-2.6940	2.1711	2.0537	-0.0429

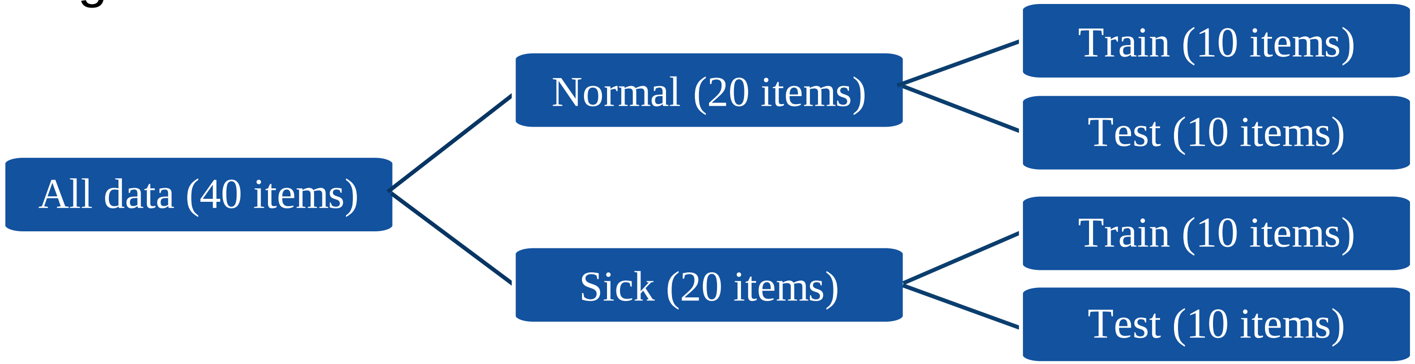
Experiment 2 (1/2)

String training and test data together and label them

1 **Data selection**

PS: This experiment only distinguishes whether there is illness or not, regardless of gender.

- Select the normal data in order of name, the 1st to 10th are training data, and the 11th to 20th are test data.
- Select the 1st to 5th data from each of the two disease data as training data, and the 6th to 10th data as test data.
- The following is the data distribution situation



Experiment 2 (2/2)

String training and testing data together and label them

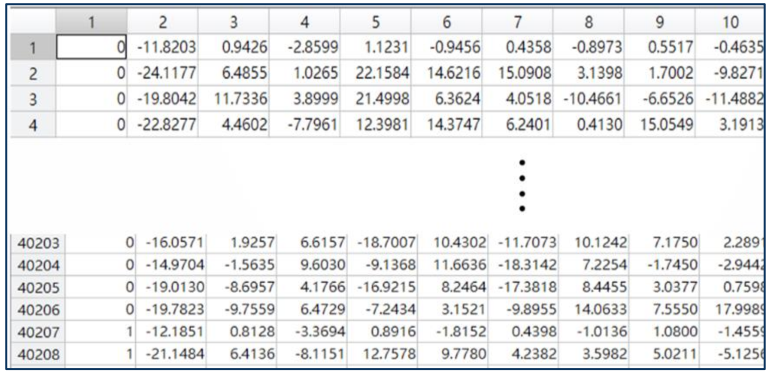
2 The sick and normal MFCCs in the training data are strung together, and the same steps are followed for the test data.

- The order is first normal and then sick. For detailed procedures, please refer to Word

3 Give MFCC labels for normal (0) and sick (1) (help query zeros and ones)

4 Store test data and training data strung together

5 Store the length of each mfc file in normal and sick test data

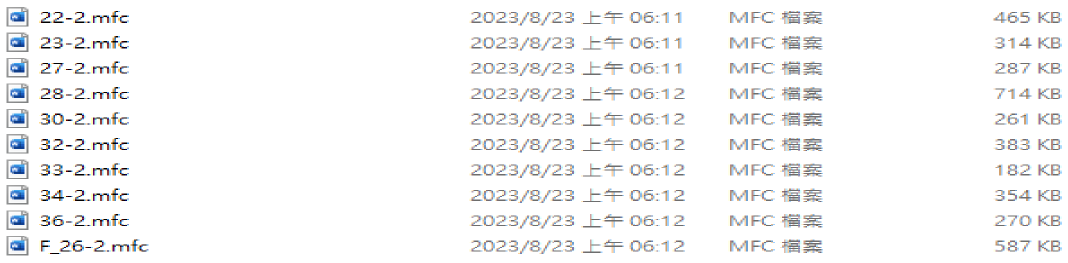


	1	2	3	4	5	6	7	8	9	10
1	0	-11.8203	0.9426	-2.8599	1.1231	-0.9456	0.4358	-0.8973	0.5517	-0.4635
2	0	-24.1177	6.4855	1.0265	22.1584	14.6216	15.0908	3.1398	1.7002	-9.8271
3	0	-19.8042	11.7336	3.8999	21.4998	6.3624	4.0518	-10.4661	-6.6526	-11.4882
4	0	-22.8277	4.4602	-7.7961	12.3981	14.3747	6.2401	0.4130	15.0549	3.1913
							⋮			
40203	0	-16.0571	1.9257	6.6157	-18.7007	10.4302	-11.7073	10.1242	7.1750	2.289
40204	0	-14.9704	-1.5635	9.6030	-9.1368	11.6636	-18.3142	7.2254	-1.7450	-2.944
40205	0	-19.0130	-8.6957	4.1766	-16.9215	8.2464	-17.3818	8.4455	3.0377	0.759
40206	0	-19.7823	-9.7559	6.4729	-7.2434	3.1521	-9.8955	14.0633	7.5550	17.998
40207	1	-12.1851	0.8128	-3.3694	0.8916	-1.8152	0.4398	-1.0136	1.0800	-1.455
40208	1	-21.1484	6.4136	-8.1151	12.7578	9.7780	4.2382	3.5982	5.0211	-5.125

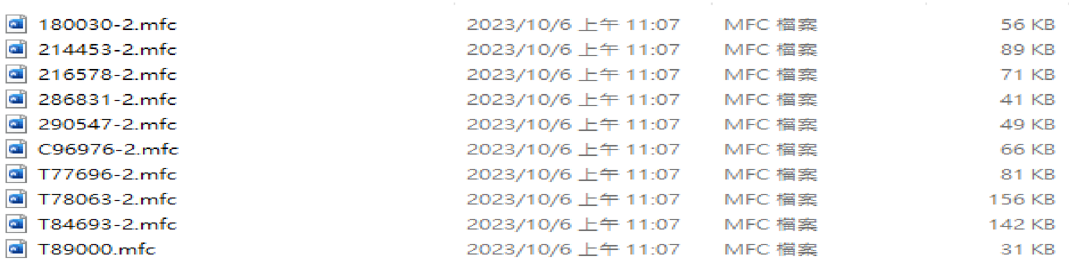
Reference results

Examples of Training and Testing Data

The training data is divided into two folders to store the mfcc files of normal and sick sounds.



22-2.mfc	2023/8/23 上午 06:11	MFC 檔案	465 KB
23-2.mfc	2023/8/23 上午 06:11	MFC 檔案	314 KB
27-2.mfc	2023/8/23 上午 06:11	MFC 檔案	287 KB
28-2.mfc	2023/8/23 上午 06:12	MFC 檔案	714 KB
30-2.mfc	2023/8/23 上午 06:12	MFC 檔案	261 KB
32-2.mfc	2023/8/23 上午 06:12	MFC 檔案	383 KB
33-2.mfc	2023/8/23 上午 06:12	MFC 檔案	182 KB
34-2.mfc	2023/8/23 上午 06:12	MFC 檔案	354 KB
36-2.mfc	2023/8/23 上午 06:12	MFC 檔案	270 KB
F_26-2.mfc	2023/8/23 上午 06:12	MFC 檔案	587 KB

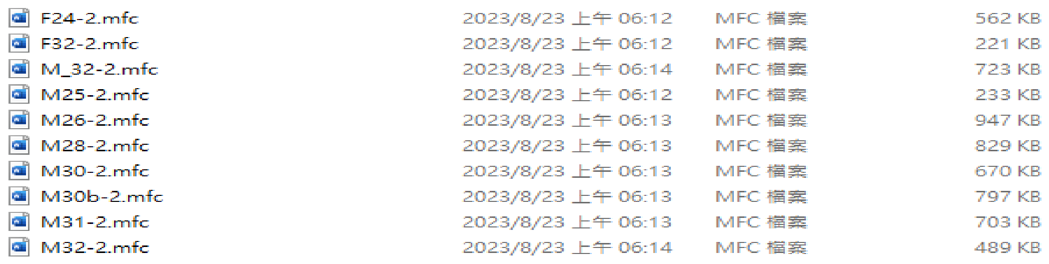


180030-2.mfc	2023/10/6 上午 11:07	MFC 檔案	56 KB
214453-2.mfc	2023/10/6 上午 11:07	MFC 檔案	89 KB
216578-2.mfc	2023/10/6 上午 11:07	MFC 檔案	71 KB
286831-2.mfc	2023/10/6 上午 11:07	MFC 檔案	41 KB
290547-2.mfc	2023/10/6 上午 11:07	MFC 檔案	49 KB
C96976-2.mfc	2023/10/6 上午 11:07	MFC 檔案	66 KB
T77696-2.mfc	2023/10/6 上午 11:07	MFC 檔案	81 KB
T78063-2.mfc	2023/10/6 上午 11:07	MFC 檔案	156 KB
T84693-2.mfc	2023/10/6 上午 11:07	MFC 檔案	142 KB
T89000.mfc	2023/10/6 上午 11:07	MFC 檔案	31 KB

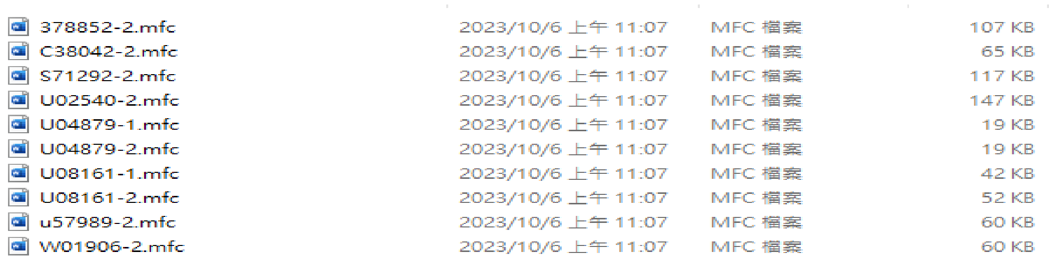
mfcc file of normal sound

mfcc file of sick sound

The testing data is divided into two folders to store the mfcc files of normal and sick sounds.



F24-2.mfc	2023/8/23 上午 06:12	MFC 檔案	562 KB
F32-2.mfc	2023/8/23 上午 06:12	MFC 檔案	221 KB
M_32-2.mfc	2023/8/23 上午 06:14	MFC 檔案	723 KB
M25-2.mfc	2023/8/23 上午 06:12	MFC 檔案	233 KB
M26-2.mfc	2023/8/23 上午 06:13	MFC 檔案	947 KB
M28-2.mfc	2023/8/23 上午 06:13	MFC 檔案	829 KB
M30-2.mfc	2023/8/23 上午 06:13	MFC 檔案	670 KB
M30b-2.mfc	2023/8/23 上午 06:13	MFC 檔案	797 KB
M31-2.mfc	2023/8/23 上午 06:13	MFC 檔案	703 KB
M32-2.mfc	2023/8/23 上午 06:14	MFC 檔案	489 KB



378852-2.mfc	2023/10/6 上午 11:07	MFC 檔案	107 KB
C38042-2.mfc	2023/10/6 上午 11:07	MFC 檔案	65 KB
S71292-2.mfc	2023/10/6 上午 11:07	MFC 檔案	117 KB
U02540-2.mfc	2023/10/6 上午 11:07	MFC 檔案	147 KB
U04879-1.mfc	2023/10/6 上午 11:07	MFC 檔案	19 KB
U04879-2.mfc	2023/10/6 上午 11:07	MFC 檔案	19 KB
U08161-1.mfc	2023/10/6 上午 11:07	MFC 檔案	42 KB
U08161-2.mfc	2023/10/6 上午 11:07	MFC 檔案	52 KB
u57989-2.mfc	2023/10/6 上午 11:07	MFC 檔案	60 KB
W01906-2.mfc	2023/10/6 上午 11:07	MFC 檔案	60 KB

mfcc file of normal sound

mfcc file of sick sound

Experiment 3

Classification using K-Nearest Neighbor (KNN) algorithm

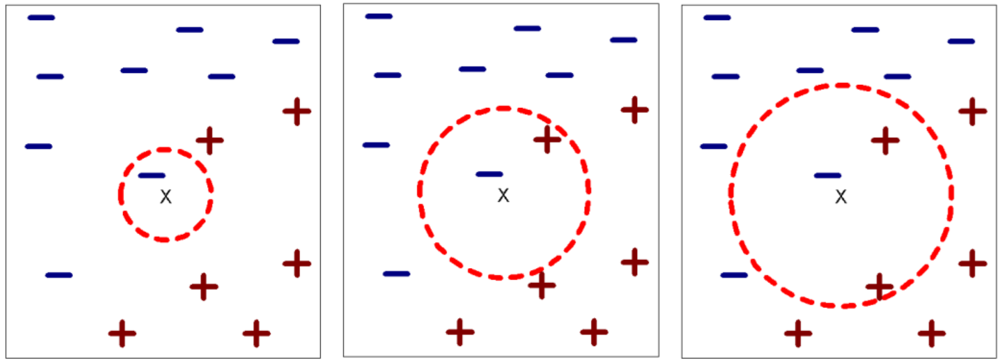
1 Set the k value, which is defined as k close neighbors, where k is set to 1

2 $D_k = \sqrt{(X_0^T - Y_0)^2 + (X_1^T - Y_1)^2 + \ldots\ldots + (X_{25}^T - Y_{25})^2}$, where $X_0^T \sim X_{25}^T$ is the 26-dimensional T-th column test data, $Y_0 \sim Y_{25}$ is the 26-dimensional training data, D_k is the k nearest distance between the testing and training data(Please refer to matlab distance function dist)

3 When k = 3, 5, 7..., a voting system is adopted, and the label that appears more often is the predicted category.

- Ex. When there are 3 neighbors around a test sound frame,
  Among them, 2 are sick and 1 is normal.
  Then the predicted sound frame is regarded as sick.

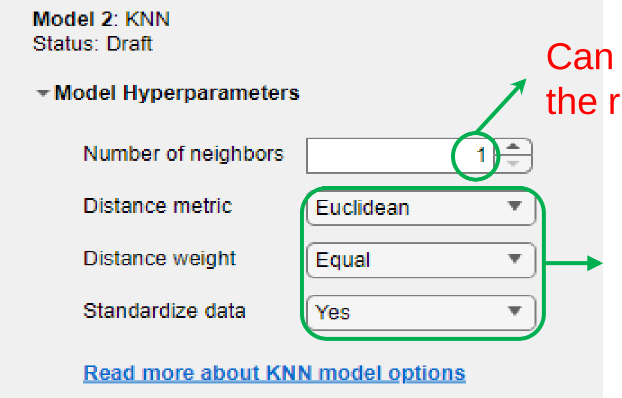
4 Use predicted labels to compare with test labels



(a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

KNN Operating Procedures

- Use label_person.m to generate Voicetrain data.
- In Matlab, click APPS->Classification Learner->New Session
- Select Voicetrain for Dataset and Resubstitution Validation (No Validation)->Start Session for Validation Scheme.
- Select Fine KNN as the model and set the parameters as shown under.
- Once completed, you can start training.
- Save the model.



Model 2: KNN
Status: Draft

Model Hyperparameters

Number of neighbors: 1
Distance metric: Euclidean
Distance weight: Equal
Standardize data: Yes

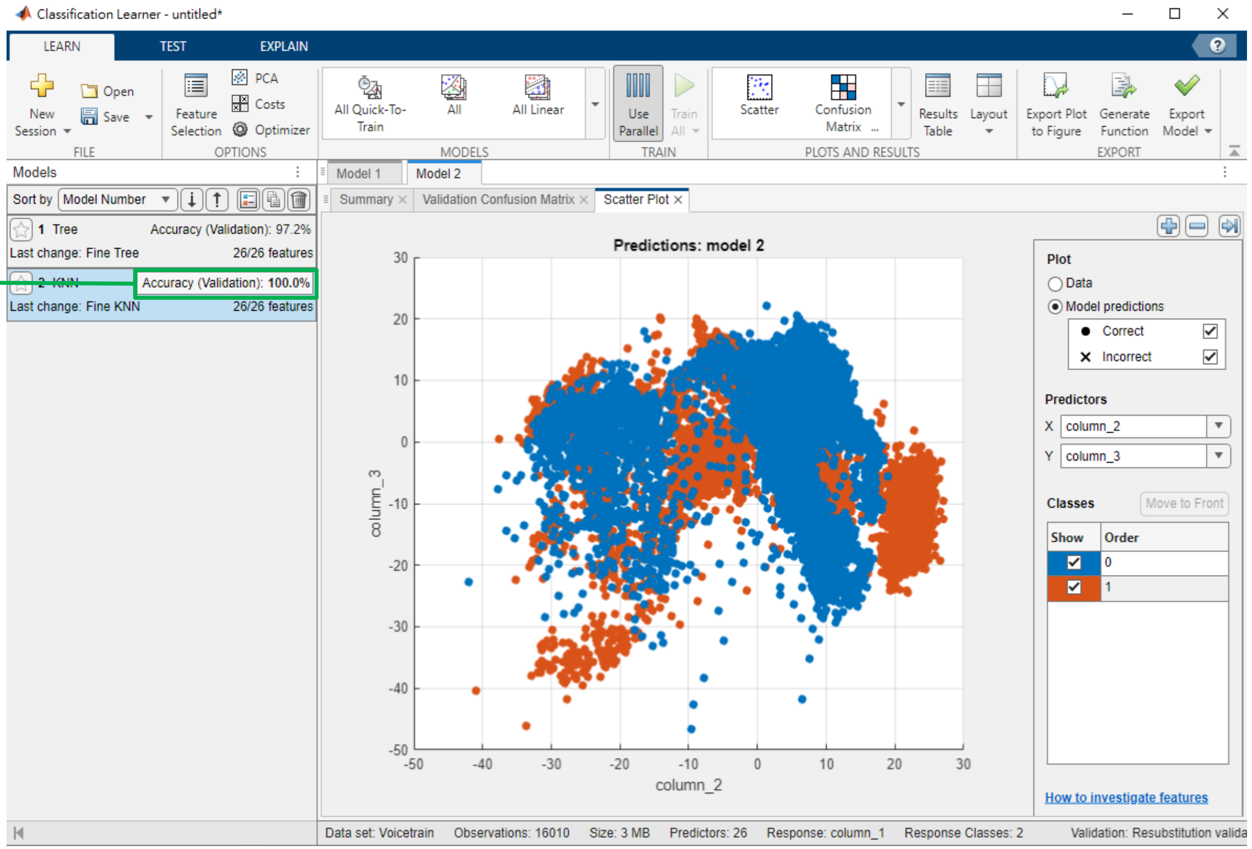
Read more about KNN model options

Can be adjusted according to the requirements of the handouts

Keep original settings

KNN Execution Result Description

Not the required accuracy



1 Tree Accuracy (Validation): 97.2%
Last change: Fine Tree 26/26 features

~~2 KNN~~ Accuracy (Validation): 100.0%
Last change: Fine KNN 26/26 features

Plot: ○ Data ● Model predictions; ● Correct ☑; × Incorrect ☑

Predictors: X column_2; Y column_3

Show	Order
☑	0
☑	1

Data set: Voicetrain Observations: 16010 Size: 3 MB Predictors: 26 Response: column_1 Response Classes: 2 Validation: Resubstitution valida

Please create a predict.m to predict.

$$\text{Accuracy} = \frac{\text{the number of correctly guessed sound frames}}{\text{the total number of sound frames}}$$

Experiment 4 (1/2)

Replace the sound frame characteristics with people and compare with test data(For details, please refer to word)

1 Use the sound frame feature to replace the person

2 Compare people who predicted the state to people who actually did it

3 Record the accuracy of the prediction person and sound frame (question 2)

Experiment 4 (2/2)

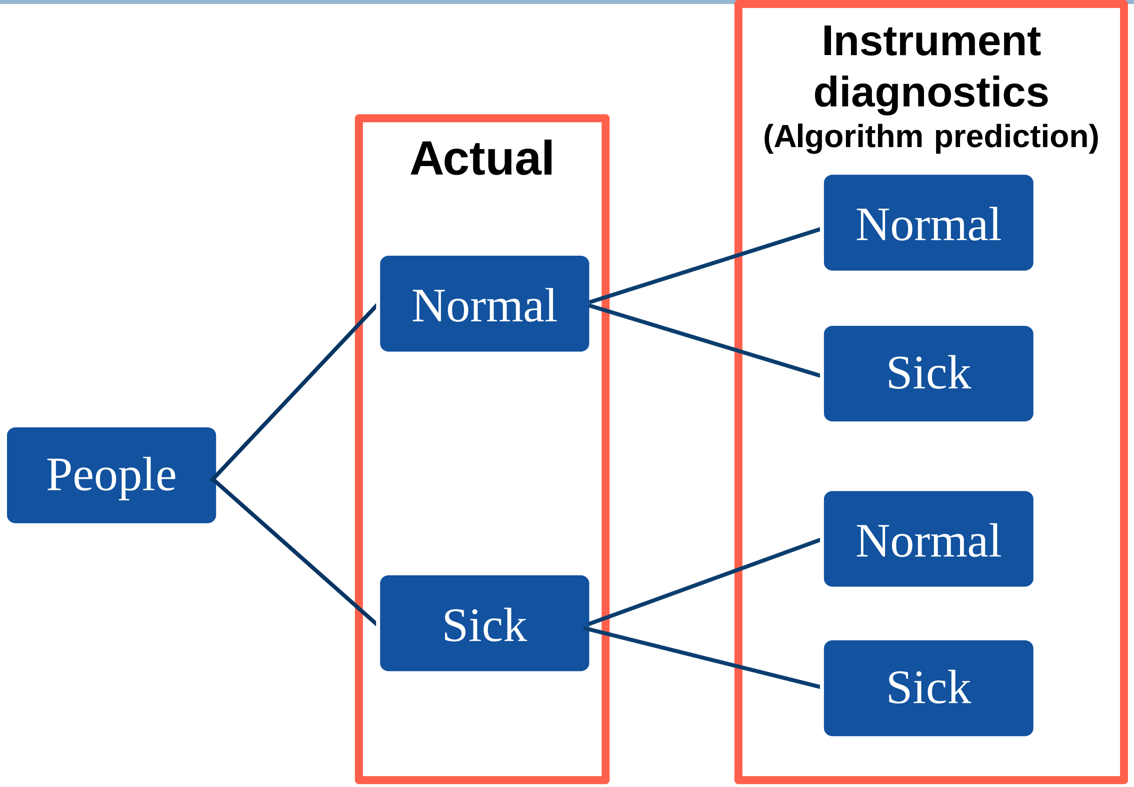
Set the sound frame ratio threshold (R) to 0.5 to indicate the predicted state (normal or sick)

A When actual/test sound is normal (0)

- When the predicted label is greater than 0.5, it is considered normal (0)
- When the predicted label is less than 0.5, it is considered sick (1)

B When actual/test sound is sick (1)

- When the predicted label is greater than 0.5, it is considered sick (1)
- When the predicted label is less than 0.5, it is considered normal (0)



Example

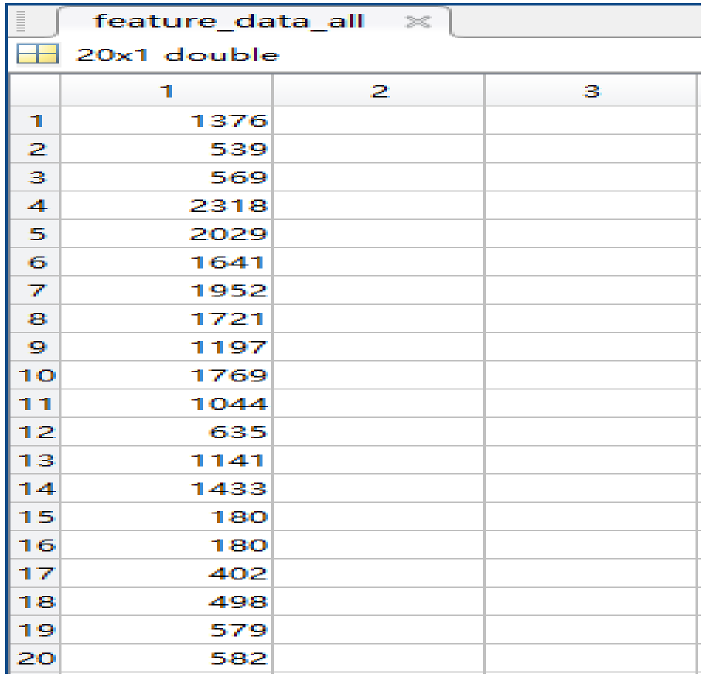
- The first patient has a normal voice (0), whose sound frame length is 3420, "0" has 1500 sound frames, and its threshold value

$$\frac{\text{Number of sound frames guessed}}{\textit{Number of all sound frames}} = \frac{1500}{3420} ≒ 0.439 < 0.5$$

This situation is A ②, the predicted outcome is sick(1)

Human Sound Frame Length

- Each of the 20 pieces of information in feature_data_all in label_person.m is the length of the person's sound frame. The length of the sound file is different, so the size will be different.



feature_data_all

20x1 double

	1	2	3
1	1376		
2	539		
3	569		
4	2318		
5	2029		
6	1641		
7	1952		
8	1721		
9	1197		
10	1769		
11	1044		
12	635		
13	1141		
14	1433		
15	180		
16	180		
17	402		
18	498		
19	579		
20	582		

Questions and Discussion

1 Try to draw the spectrogram and waveform diagram (refer to LAB6) of sick and normal sounds (pick one each) and explain the differences?

2 Try to adjust the parameters sample_rate, frame_time, and frame_move_time in mfcc_v2.m to explain the significance of parameter changes for capturing sound characteristics from mfc files.

Parameter Description ：

- sample_rate：The sampling frequency of a piece of audio (Unit：HZ)
- frame_time： The sampling time length of a sound frame (Unit：ms)
- frame_move_time： The time when one sound frame overlaps with the next sound frame (Unit：ms)

Questions and Discussion

3 Complete the table below and try to adjust the K value, setting it to 3, 5, or 7. What is the most accurate value?

K	1	3	5	7
Accuracy(frame)	77.8%	?	?	?
Accuracy(person)	75.0%	?	?	?

Adjust the K value and record the accuracy (the table is a reference value)

4 Use Experiments 3 and 4 to predict the 10 voice files in the folder other

Report Description and Allocation

- Please answer the Experiment 1 to Questions 4 in detail and upload the word file and m-file to the portal before the deadline, as well as submit the written report.

item	Score allocation
Experiment 1-4	40%
Question 1-4	40%
M file	20%