



# Identifying Speaker with Pitch and Timbre

Students : Chi-Hsaun Fan, Yi-Pin Teng,  
Ka-Fun lao, Yi-Shan Chuang

Voice recognition is non-contact, non-intrusive, difficult to embezzle and easy to use. In this study, Auto-Correlation Function and Modulation Spectrogram which generated by Fourier Transform are used to find the features of a voice. After calculating the correlation coefficient between each pair of voices, Speaker can be recognized with these features of voiceprint. The threshold pitch period value 90.7 can recognize the gender correctly and the accuracy is 100%. Comparing correlation coefficients from different audios can recognize the speaker and the accuracy is 100% in vowels sound and 100% in numbers sound.

## Recognize gender with pitch:

The method of recognizing gender with pitch is shown in Figure 1. The male vocal folds are longer in length. The vibration is slower and performs lower pitch with longer period. The female vocal folds are shorter in length. The vibration is faster and performs higher pitch with shorter period.

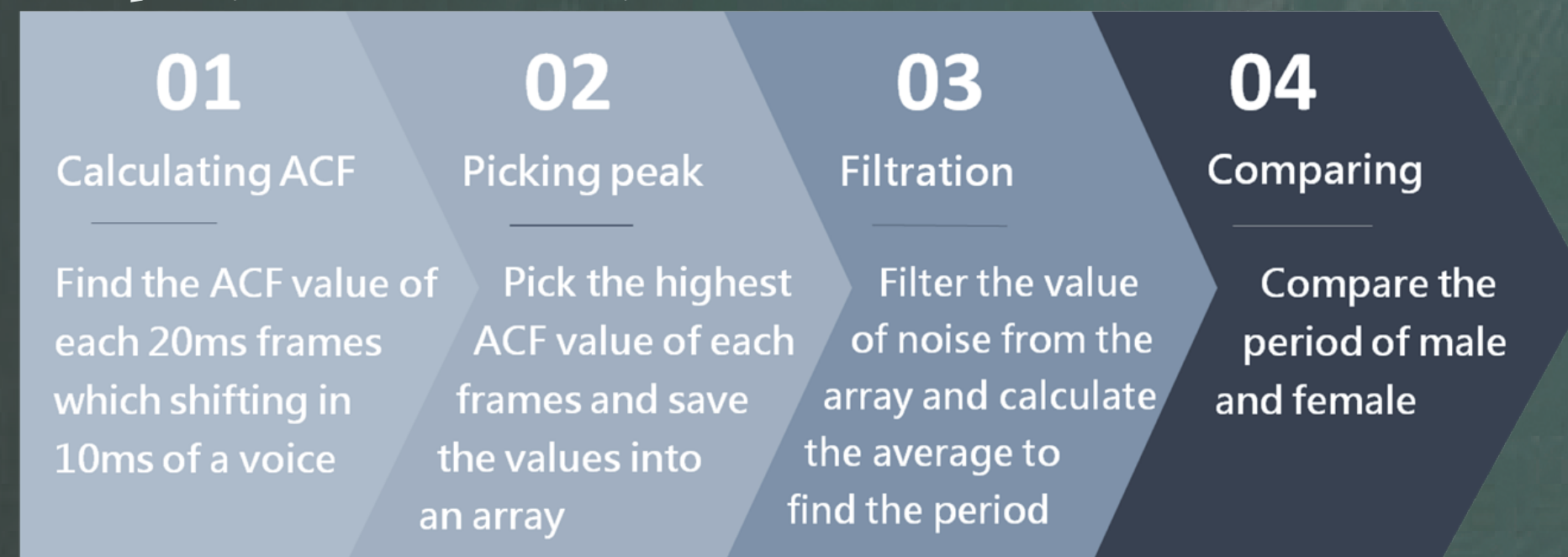


Figure 1. Method of gender recognition

## Recognize speaker with pitch and timbre:

The method of recognizing speaker using modulation spectrogram is shown in figure 2. Values in first column and second column of the matrix shown in Figure 3 reflect the feature of timbre. Values in 35\*48 matrix reflect the pitch. Calculating the correlation coefficients with the formula shown in Figure 4 of different audios in these three parts show the differences.



Figure 2. Method of speaker recognition

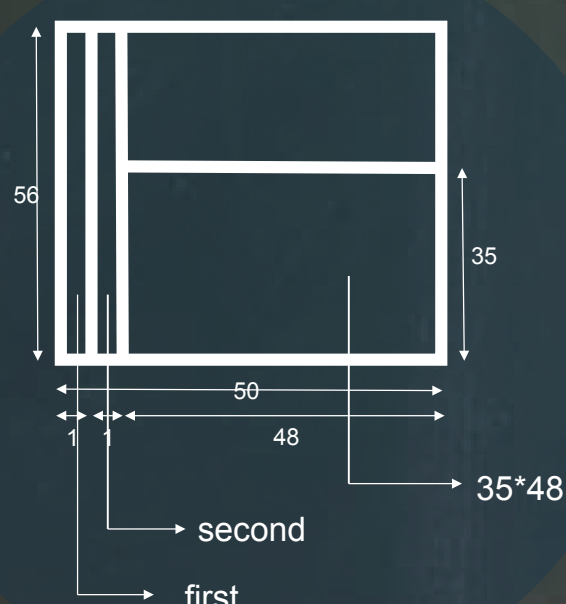


Figure 3. The DFT Coefficients Matrix (Modulation Spectrogram).

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Figure 4. The Correlation Formula.

## I. Abstract

## II. Introduction

## III. Approach

## IV. Experiment & Result

## V. Conclusion

II.

Physically, each of us has a unique voiceprint. Just as we each have fingerprints, eye prints, etc. Since fingerprint recognition becomes insecurity, eye print recognition and voiceprint recognition start receiving attention. Acquiring voice with voiceprint is more convenient and natural. The complexity of the algorithm for voice recognition and the cost of acquiring voice are low. There is vast development potential in voiceprint recognition although it is not common now. In this study, auto-correlation function is used to find the period and frequency of voice. Modulation spectrogram is used to find the pitch and timbre of voice.

IV.

## Experiment I: Gender recognition

The threshold period value shown in Figure 5 is used to recognize the gender of voices. The data of testing set shown in Figure 6 prove that the threshold is effective.



Figure 5. Threshold period value

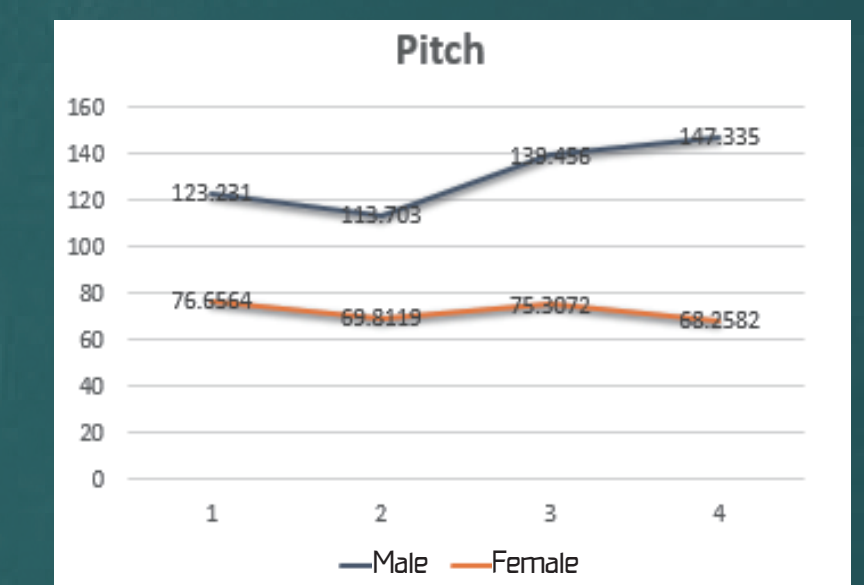


Figure 6. Period of different testers

## Experiment II: Speaker Recognition

In this experiment, each person has two records. Four persons record vowels sound and seven persons record the sound of numbers 0-9. One of the records is used to calculate the correlation coefficients of each word in record with each record of all persons. Vote the highest correlation coefficients of each word as shown in Figure 7 to recognize who the speaker of the record is. All of the testers are recognized successfully as the results shown in Figure 8 and Figure 9.

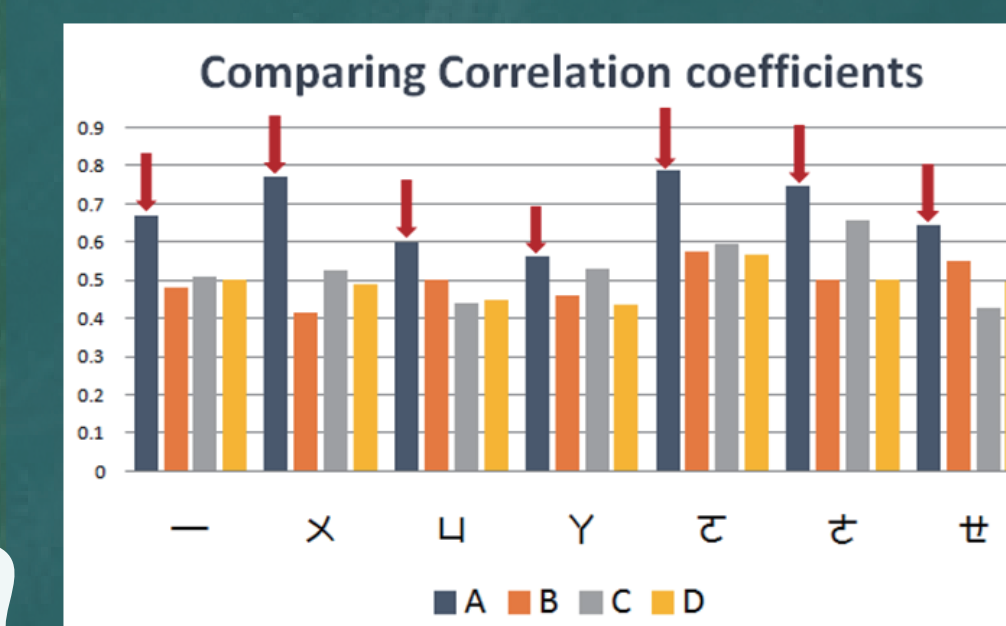


Figure 7. Voting method

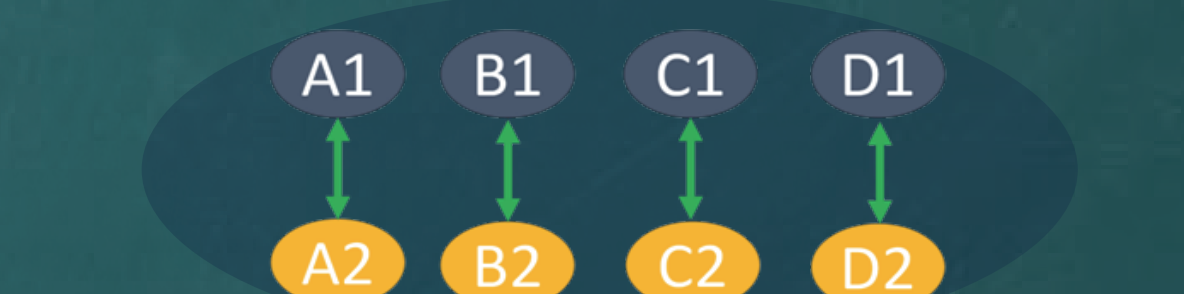


Figure 8. Result of recognition with vowels

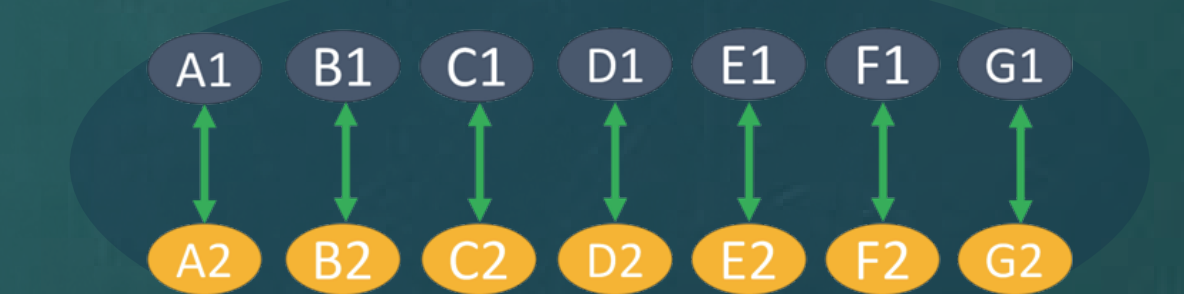


Figure 9. Result of recognition with numbers

V.

According to the result of the experiments, the accuracy of gender recognition and speaker recognition are 100%. The common voice recognitions nowadays are speech recognitions like Apple Siri, Voice search from Google, Samsung S Voice, etc. These technologies perform Speech To Text. For speaker recognition, since some of words contain consonants in the pronunciation, there is no pitch period in consonants and it's difficult to specify any features. Improving speaker recognition in this study can be applied to Voice Unlock, specific vocal remove, etc.