CHAPTER 5

FEATURE SELECTION

AND EXTRACTION
5.1 INTRODUCTION

In pattern recognition problems, one has to always consider the dimensionality of the pattern. Features are pattern descriptors, having lower dimension. These features are important in representing a pattern and characterizing the discriminating properties of pattern classes. The complexity of the algorithm increases with the number of patterns and their dimensions. Therefore, it is essential to restrict the number of patterns to the most important and having discriminatory information. This in turn will reduce the hardware requirement and also the cost of measurement extraction. The dimensionality reduction in most of the cases improves the performance [43].

The reduction in dimensionality is carried out in two ways. In the first one, the least contributing measurements to the class separability are identified and eliminated. Thus a subset of the superset is chosen by ignoring the redundant and the dispensable measurements. This process is called feature selection. In the second method, the measurements are mapped into a lower dimensional feature space to reduce the dimension. This is called feature extraction. The important points to be
considered for feature selection and extraction are the feature evaluation criteria, the dimensionality of the feature space, the optimization procedure and the form of mapping.

5.2 FEATURE SELECTION METHODS IN PATTERN RECOGNITION

5.2.1 Probabilistic Method

This method is considered in terms of error 'e' in the two class problem. When the two probability density functions overlap, the error will be maximum and when the PDF's are non-overlapping the error will be zero. There are several probability distance measures available of which Mahalanobis distance measure is very popular and commonly used in many speaker recognition methods.

$$d = (\mu_2 - \mu_1)^T \Sigma^{-1} (\mu_2 - \mu_1)$$  \hspace{1cm} (5.1)

Where \( \mu_i \)'s are mean vector and \( \Sigma \) is the covariance matrix.

5.2.2 Probabilistic Dependence Measure

In this approach two random variables are involved viz., a pattern vector and the class. An observation of the outcome of the former enables one to make a decision about the latter. Conditional density function is made use of in this method. In situations where the pattern vector and class are independent, the method fails in classification.
5.2.3 Entropy Method

Entropy measure is also employed for feature selection. As a posteriori probability is calculated to determine how much information has been gained from the experiment. If all the classes are equally probable, then the information gain is minimal and the entropy is maximum.

5.2.4 Interclass Distance Measures

In this method, it is assumed that the pattern vector of each class occupy distinct region in the observation space. The average pair-wise distance between the patterns in the set is a measure of class separability in this space. The Euclidean metric is a common interclass distance measure in speaker recognition problems.

5.3 FEATURE SELECTION CRITERIA IN SPEAKER RECOGNITION

Speaker recognition problems exclusively have to meet certain specific criteria for the feature selection, which depends on the measurements of speech characteristics. These criteria can be listed as shown below [10].

1. The features should occur naturally or frequently in normal speech.

2. These features should be easily measurable.
3. They must vary as much as possible among speakers, but must be as consistent as possible for each speaker.

4. They should neither change over time, nor be affected by the speaker's health.

5. They should neither be affected by reasonable background noise nor depend on specific transmission characteristics.

6. They should not be modifiable by conscious effort of the speaker or at least be unlikely to be affected by attempts to disguise the voice.

Practically it is impossible to incorporate all the above criteria to make the recognition system foolproof. Especially the last mentioned three are most difficult to adopt. So one has to make an engineering compromise.

5.4 FEATURE EVALUATION IN SPEAKER RECOGNITION

There are different popular methods used for feature evaluation in speaker recognition problems. The F-ratio analysis is one of these evaluation techniques for speaker discrimination ability of the features. In speaker recognition, a parameter is said to be good when the individual speaker probability distribution is as narrow and as widely separated as possible. The F-ratio can be defined in other words, as the ratio proportional
to the variances of speaker means to the mean of the speaker variances. When the individual speaker distributions are farther apart and narrower, the F-ratio value is higher and the parameters are selected as suitable parameters [5].

Discriminant analysis is another feature evaluation technique in speaker recognition. In this method new features are created with linear combination of the original features. The optimum linear transformation of the original feature space is determined by a combination of eigen vector analysis and F-ratio technique [23].

The draw back in these two feature evaluation methods is that the features with high F-ratio may not contribute much to the performance of the recognition system, than a feature with lower F-ratio [23].

Another evaluation technique uses a knock out method. The stress in this technique is that the features selected should contribute to the performance of recognition. If a set of N features are available for evaluation, the effectiveness of a subset of N-1 features is considered and the error performance is determined. By this method, the most effective subset (N-1) is chosen and the single feature is eliminated or knocked out. Then
a subset of \((N-2)\) is chosen and the same elimination technique is used. This process is repeated till all the features are knocked out and the effectiveness is arranged in the reverse order [23].

Dynamic programming methods are also used for determining feature effectiveness [51].

5.5 FEATURES USED IN THE PRESENT METHOD

Though there are many feature selection and evaluation methods, there are not many standardized features in the speaker recognition systems. The features used in different methods are dissimilar. Many features are claimed to be successfully used in different recognition systems. Using all these features in a single system is not practical due to various reasons. A possibility of using a few features in this method is studied. Attention is given to features which are easily measurable and implementable in a small computer system at a reasonably good performance. Avoiding very complex features, one can save a lot of computation time and make the system faster. An appropriate knock out method is adopted by critically examining the performance. A few measurements which are felt as simple are chosen for this purpose. Different combinations of these features are used in the method and performance is determined. The features whose contribution is small are left out.
The features selected in this method are

1. Short-time energy
2. Zerocrossings
3. Autocorrelation function (ACF)
4. Pitch by time domain symmetry and
5. Phoneme timings, transition times and their entropy from the utterances.

The features energy, zerocrossings and ACF are successfully used in one or other forms in the earlier methods [19,20,22,25,50,92] while time domain symmetry measures is a new technique used in this method. The duration of speech was also considered in some earlier methods [1,10]. In some features, the measurements taken are different from those adopted in the earlier methods.

5.5.1 Feature Extraction

While some of the simpler techniques are adopted directly from previous methods, others are extracted using simple techniques as explained below.

For finding the short-time energy, a segment of fixed length (600 samples) at maximum amplitude region of each word in the phrase is chosen and the energy is determined.
The number of zero-crossing within a length of four pitch periods is determined. The number of positive peaks within this same period is also taken as a feature measurement.

The Autocorrelation Function of 512 points is determined using an FFT method. From this ACF, the distance between the first two prominent peaks, slope between these peaks and number of smaller peaks within these peaks are taken as measurements.

5.5.1.1 Time domain symmetry

Since time domain symmetry is a new approach for pitch detecting, it is discussed in detail. This feature is mainly selected because, the fundamental frequency or pitch is a very important speaker discriminating feature in many speaker recognition systems [5,10,18,23,33,37].

It is an accepted fact that the voiced sound has a periodicity and determination of this periodicity leads one to pitch detection. There are different standard algorithms used for this purpose viz., Autocorrelation, Center clipping method and cepstrum method [144,145]. In most of these methods, some type of transformation is involved and also the computation involved is lengthy. So a new approach was felt necessary and
has been tried for detection of pitch which in turn is also a speaker dependent measurement. The symmetry check is purely a time domain method and mathematically less complex.

The basic principle is that of extraction of the periodicity of the voiced signal. It is well accepted that speech signal is a complex signal and it is the convolution product of the fundamental frequency generated by the vocal chords and the harmonics generated from the vocal tract. Thus there are two significant components viz., the slowly varying signal corresponding to the vocal chords and rapidly varying signal component corresponding to the vocal tract harmonics. In other words one component corresponds to the pitch and the other component corresponds to formants. Fig.5.1 shows a typical voiced sound segment. In this method, the main task is to filter out the effects of vocal tract and extract only the slowly varying fundamental frequency component.

An algorithm has been developed for this pitch detection. This algorithm works on a window based operation. The symmetry in the window is checked every time and if it shows a symmetry, this window is considered for pitch. Initially a small window of 32 samples is considered. Correlation coefficient of the samples in that window is determined. If the correlation coefficient exceeds a threshold value, the periodicity is checked
Fig. 5.1 A Typical Voiced Speech Segment /a/
with other symmetry measures, these symmetry measures are average, rms and rate of change of the samples in the window. For this purpose, the samples in the window is divided into equal halves and these measurements are determined. The difference in each measure gives a symmetry measure. The three symmetries are determined by

\[ A = \frac{1}{N/2} \sum_{i=1}^{N/2} x_i - \frac{1}{N/2} \sum_{j=N/2+1}^{N} x_j \left( \frac{1}{N} \sum_{i=1}^{N} x_i \right) \]  

\[ R = \left( \frac{1}{N/2} \sum_{i=1}^{N/2} (x_i^2)^{1/2} - \frac{1}{N/2} \sum_{j=N/2+1}^{N} (x_j^2)^{1/2} \right) \left( \frac{1}{N} \sum_{i=1}^{N} (x_i^2)^{1/2} \right) \]  

\[ C = \left( \sum_{i=1}^{N/2} (x_i - x_{i+1}) - \sum_{j=N/2+1}^{N} (x_j - x_{j+1}) \right) \left( \sum_{i=1}^{N} (x_i - x_{i+1}) \right) \]  

where A, R and C are normalized symmetry coefficients corresponding to Average, Root Mean Square, and Rate of change. N is the total number of samples in the window considered and \( x_i \) is the ith sample.

The total symmetry of the window can be computed to be

\[ T = \frac{1}{3} (A + R + C) \]
From eqn. (5.5) it can be observed that when $T = 0$, the signal corresponding to the samples in the window shows an absolute symmetry and when $T$ approaches 1 it shows maximum asymmetry or lack of periodicity. A threshold value of $T$ is fixed by trial and error and the window of samples falling within this threshold is taken as periodic. The period corresponding to half the number of samples in the window ($N/2*ts$, where $ts$ is the sampling time of the signal) is taken as the pitch period. If there is no symmetry in the window considered, the window size is changed and symmetry is again checked in the same manner. The complete algorithm, experiment and results are given in Appendix A. Upon confirmation of consistency the technique can be accepted as a good and easy technique for determining pitch period.

Once the pitch period is detected, the intermediate measures viz., average, RMS and rate of change of symmetry coefficients and pitch number are taken as valid measurements for speaker recognition.

An attempt is also made to define a set of relationships between the features selected in the current work and elsewhere and a new model for speech production. This is presented in a subsequent chapter after presenting the model.
5.6 CONCLUSION

A method is explained in this chapter which selects features similar to those established method but with different techniques of measurements and evaluation. A suitable strategy for knock out of features is also developed. A reasonably good performance is obtained using a set of features chosen in this method.