The Effect Of Combining Entropy Feature With MFCC On Isolated-Word Speech Recognition

طالب محمد جواد عباس
مدرس مساعد
كلية الرافدين الجامعة

Abstract

In this paper, the MFCC feature(technique) is examined for isolated-word speech recognition . The combination of several entropy feature with the above , techniques are studied . It is shown that this combination introduce a more accurate results.
The result of applying MFCC alone was 77.55%.The result of using feature entropy was 57.83%. But the result of combination of the above two features together was improved to 87.75%.

1.Introduction

A key issue for implementing an accurate speech recognition system is the set of acoustic features extracted from speech signal [1].

A feature is some measurable characteristic of the input which has been found to be useful for recognition .In many isolated-word recognition systems, the pattern takes the form of a set of time function- i.e. ,values for each feature are recorded over the length of the word rather than at particular points ;such patterns are often called templates [2].

The speech signal is a highly redundant signal. It carries linguistic message as well as other information about speaker, regarding their physiology , psychology ,etc.

Feature measurement, some times called feature extraction, is basically a data reduction technique. Ideally features should meet the following criteria :

Insensitive to extraneous variables (i.e. emotion ,state of talker etc);stable over long periods of time ;frequently occurring ;easy to measure and finally not correlated with other features [3].This is the result of the fact that better signal feature extraction leads to better recognition performance [4].

It is generally impossible to find features that meet all requirements at once, and compromises are inevitable. In fact, the selection of the best parametric representation of the acoustic data is an important task in design of any speech recognition system. Several parametric representation of speech signal such as LPC parameters, filter bank parameters and cepstral parameters has been the interest at many researchers,and other researchers was present a new algorithm based on the template matching technique and makes use of a new set of features ,which consists of combination of the fractal dimension with the Mel-frequency cepstral coefficients (MFCCs) [3].

MFCCs perform well when used for clean speech recognition. However, for noisy speech the recognition rates go down.

Augmenting the MFCC feature vector by dynamic features improves both discrimination and robustness of the MFCC-based recognizer. An alternative parameterization based on frequency filtering (FF) technique .By using FF , a significant improvement with even lower computational costs can be obtained for both clean and noisy speech recognition rates in comparison to MFCC [5].

In the paper, the isolated-word recognition introduced firstly, the principle of MFCC and entropies are presented after them. Then we deal with MFCC and entropies separately and after that we integrate them together. Finally we present the experiment results and conclusions.

2. Isolated-Word Recognition

A great amount of the work to date concerns the recognition of set of isolated-words, spoken by a limited number of speakers [6]. Recognition problems are conveniently up into the following categories. The sequence is in order of increasing difficulty.

A. Isolated-Word recognition. Recognition of words separated by pauses.

B. Word Spotting .Detection of occurrences of a specified word in continuous speech.

C. Connected speech recognition. Recognition of words without pauses in between.

D. Speech understanding. An elaboration of category A; create a pool of information about the utterance and draws on stored information about the language being spoken.

In all of these, the system may be talker dependent (i.e, trained for a particular user) or talker independent. At any level, talker-independent systems are much more difficult to implement and show poorer performance than talker-dependent systems. The simplified block diagram of such systems is shown in Figure 1.

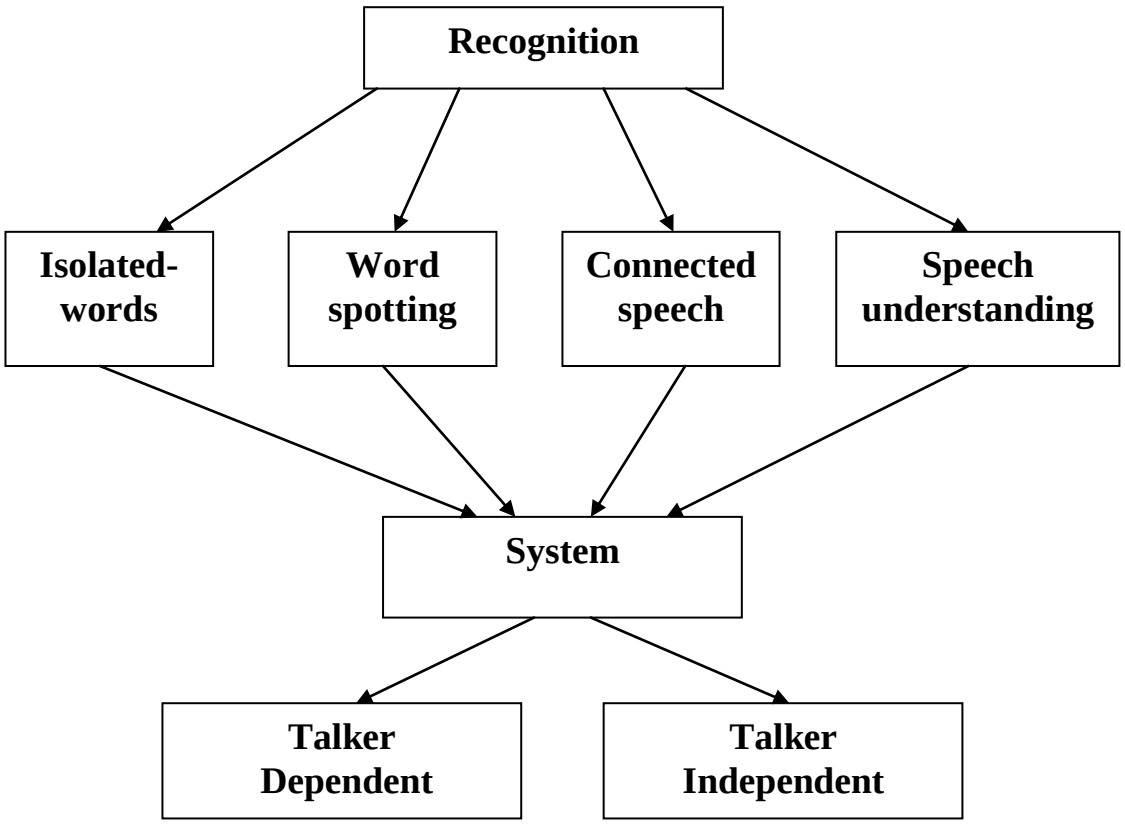

Fig 1: Block Diagram show the types of recognizer

We will concern our efforts on isolated-word recognition, because these present the fewest problems [2].

3. COMPUTATION of MFCC

Because of the known variation of the ears' critical band-widths with frequency, filters spaced linearly at low frequencies and logarithmically at frequencies have been used to capture the phonetically important characteristics of speech [7].

The MFCCs result from replacing the linear frequency scale by a scale, where the frequency range of input speech signal is divided into a bank of band pass filters [3].This result suggested that a compact representation would be provided by a set of Mel-frequency cepstrum coefficients. These cepstrum coefficients are result of a cosine transform of real logarithm of the short-time energy spectrum expressed on a Mel-frequency scale [7].

In MFCC, the main is that it uses Mel-frequency scaling which is very approximate to human auditory system [7].The basic steps are:

Step (1): The speech signal is framed into blocks of approximately (25) ms in length (sampling rate 11025 Hz), hamming window is applied to the framed signals (blocking into blocks of 256 samples).

Step (2):

a. Apply the pre-emphasis which done by filtering the speech signal through a simple FIR filter, such as the one described by eqn(1):

$$H(z)=1-0.9375.z^{-1} \qquad (1)$$

b. Using MATLAB function which below:

Y=FILTER(B,A,X) (this function made filters the data in vector X with the filter described by vectors A and B to create the filter data Y).

c. The parameters (filter) A and B depend on the transfer function that has only poles and can be written as [8]:

$$H(z)=\frac{1}{1-\sum_{k=1}^{p} a_k z^{-k}} \qquad (2)$$

Where a_K are the filter coefficients, P is the number of poles in the digital filter, and z is the z-transform representation.

d. After clarifying the parameters of MATLAB function in (b),we can applied and get (Y) when substituting the following values:

A=1 (Numerator of equation 2)

B=1-0.9375 (Denominator of equation 2)

X= vector or matrix contains data

Step (3): Apply Fourier transform to each frame and windowed signal, to obtain a short-time spectrum.

Step (4): Compute the power spectral density (PSD) using the following equation:

$$PSD(Y) = [abs(Y)]^2 \quad (3)$$

Step (5): Take the log of each frame result from step (4) above.

Step (6): Take the inverse DFT of log spectrum.

The result is a vector of cepstral coefficients, which is the feature vector of the speech frame.

The simplified block diagram of such steps is shown in Figure 2.

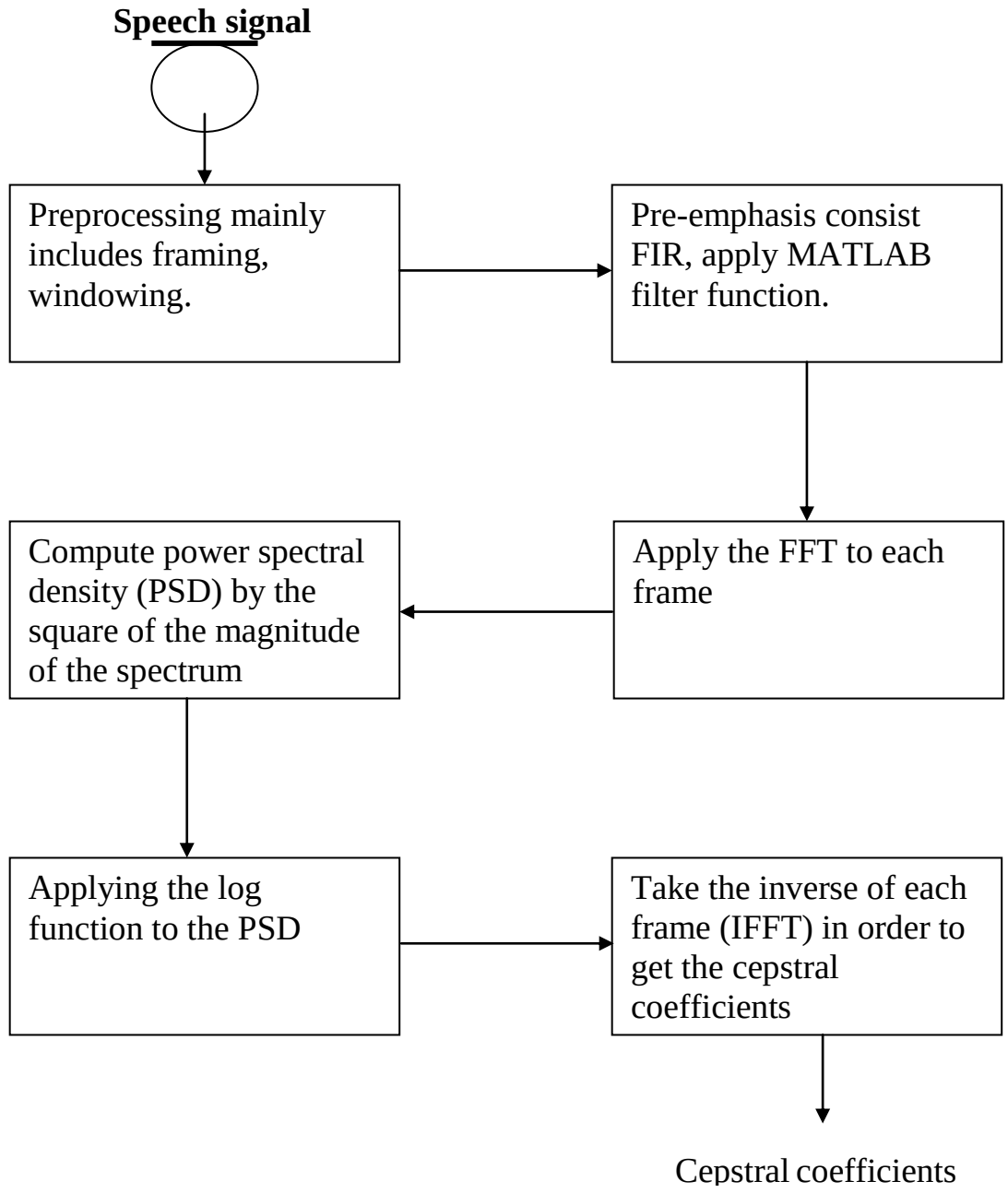

Fig 2 : Block diagram show steps of MFCC .

4. COMPUTATION of Entropies

Entropy is a quantitative measure of how uncertain the outcome of a random experiment is. Its definition and interpretation was introduced by C.E. Shannon (1948) and N.Wiener (1961) and is an integral part of textbooks on statistics, e.g. Papouls (1984).
This idea of measuring the uncertainty of random experiment was extended to a discrete time random signal by means of notion of prediction [9].

In this section we briefly review the entropies used is this paper.
Given a signal x , its Shannon entropy is defined as

$$H = -\sum_{i=1}^{M} P_i \log(P_i) \qquad (4)$$

Where p_i is the probability that the signal belongs to a considered Interval [10]. The relative entropy between two probability pi and ri corresponding to different frames of same signal, read as

$$D(p\backslash r) = \sum_{i=1}^{M} p_i \log\left(\frac{P_i}{r_i}\right) \qquad (5)$$

And

$$Dq(p\backslash r) = \frac{1}{1-q}\sum_{i=1}^{M} p_i\left[1-\left(\frac{P_i}{r_i}\right)^{q-1}\right] \qquad (6)$$

5. The Database

The data was originally recorded using:

- 11025Hz sampling frequency (sampling rate).
- 16 speakers (8 male and 8 female).
- Uttered 7 Arabic Words.
- Part of there data were designated as training material and the rest as test material.

After showing steps of finding MFCC and equations (4,5,6) for entropies in (4) and database specification which we deal with in (5), we get the result mentioned in Table(1) for each technique separately ,and combination during using MATLAB and writing programs within available capability in it.Taken into consideration that the distance type used was Mahalanobis type.

Table1 : The result of applying individual feature

MFCC	77.55%
Shannon	53.10%
D-Entropy	59.18%
DQ_ Entropy	61.22%

Table2: The result of combining two features

MFCC+ Shannon	81.63%
MFCC+D-Entropy	75.51%
MFCC+DQ	81.63%

Table3 : The result of combining three features

MFCC+SH+D	83.67%
MFCC+SH+DQ	83.67%
MFCC+D+DQ	81.65%

Table4 : The result of combining four features

MFCC+SH+D+DQ	87.75%

6. Conclusions

a. A new set of feature was used for isolated-word speech recognition under the template matching technique

b .Experimental results in table (1,2,3,4) showed significatly improved recognition accuracy .

c. In fact , from table (2) , table (3) , and table (4) we can notice that the recognition performance is greatly improved when MFCC and (Shannon , D-entropy , DQ-entropy) are combined together and used as a single set of features .

d. In particular , the recognition rates are significantly increased , for isolated - words, during the first try from 62.76% to 79.59% (from table (1), table (2)), and when addition of MFCC is done to pair of entropies ,it increased to 82.99%. Finally the last table (table 4) show the result was becomes 87.75%.

7.References

[1] Marcos Fa ndez-Zanuy , Anna Exposito "Nonlinear Speech Processing Applied to Speaker Recognition "
Cost277 wepsite:http://www.ee.ed.ac.uk/~cost277.

[2] Thomas W.Parsons "Voice And Speech processing"
McGraw –Hill Book Company, 1987.

[3] S.Fekkai.M.Al-Akaidi "New Features For Speaker Independent Speech Recognition".
Faculty of Computing Sc. & Engineering, De Montfort University, Leicester, LEI 9BH, UK. 2001.
Email:mma@dmu.ac.uk.

[4] S.M.Ahadi,H.Sheikhzadeh,R.L. Brennan & G.H.freeman."An Efficient Front-End For Automatic Speech Recognition".
EE Dept., Amirkabir University of Technology, Tehran, Iran.
Dspfactory Ltd., Waterloo, Ont., Canada. 2003
Email:smahadi@dspfactory.com,
freeman@pce.uwaterloo.ca

[5] Dusan Macho, Climent Nadeu, Javier Hermando, Jaume Padrell."Time And Frequency Filtering For Speech Recognition In Real Noise Conditions".
Universitat Politecnica de Catalunya Barcelona,Spain. 1998.
Email: dusan@gps.tsc.upc.es

[6] Mahmoud , w .A "Quantisation Techniques For The Classification And Recognition of Speech Signals ". A thesis submitted to the University of Wales , 1986.

[7] Li Tan and Montri Karnjanadecha . "Modified Mel-Frequency Cepstrum Coefficient ".
Department of Computer Engineering , Faculty of Engineering
Prince of Songkhla University Hat Yai,Songkhla Thaiiland,90112. 2001.
E-mail:Litan212@hotmail.com,montri@coe.psu.ac.th

[8] Rabiner .L.R and Schafer.R.W,
" Digital Processing Of Speech Signals ",
Englewood Cliffs, NJ: Prentice Hall, 1978.

[9] Wolfgang Wokurek "Entropy Rate-Based Stationary, Non-Stationary Segmentation of speech ".
Institute of Natural Language Processing ,University of Stuttgart, Germany. 2000.
http://www.ins.stuttgart.de./~wokurek.

[10] Hugo L. Rufiner . Maria E.torres.

"Introducing complexity Measures In Nonlinear Physological Signals :Application To Robust Speech Recgnition"
http://www.elsevier.com.locate physic. Argentina
Received 2 May 2003, received in revised from 1 july 2003.

المستخلص:

في هذا البحث فأن خواص MFCCدرست لغرض تمييز الكلام ذو المفردة المعزولة .وكذلك تم دراسة الخواص العديدة من entropyودمجها مع الأسلوب (الخواص) الواردة أعلاه.لقد ظهر أن عملية الدمج قدمت نتائج اكثر دقة، في حين كانت نتيجة MFCCوحدة هي ٧٧.٥٥%، ونتيجة استخدام خواص entropyهي ٥٧.٨٣%. ولكن نتيجة الدمج بين الأسلوبين معا (دمج جميع لكلا الأسلوبين) ظهور تحسن كبير في التمييز والتي هي ٨٧.٧٥%.