



Proposed Speech Analyses Method Using the Multiwavelet Transform

Baydaa Jaffer Al-Khafaji

Dept. of Computer Science/ College of Education for pure Science (Ibn Al-Haitham) /
University of Baghdad

Received in : 7April 2014, Accepted in :14 April2014

Abstract

Speech is the first invented way of communication that human used age before the invention of writing. In this paper, proposed method for speech analyses to extract features by using multiwavelet Transform (Repeated Row Preprocessing).The proposed system depends on the Euclidian differences of the coefficients of the multiwavelet Transform to determine the best features of speech recognition. Each sample value in the reference file is computed by taking the average value of four samples for the same data (four speakers for the same phoneme). The result of the input data to every frame value in the reference file using the Euclidian distance to determine the frame with the minimum distance is said to be the "Best Match". Simulation program using visual basic has been in order to determine final results.

Keyword: Multiwavelet Transform, Euclidian distance, Hamming window, Repeated Row Preprocessing

Introduction

Speech communicate system consists of the speech mechanism and hearing system , Speech Mechanisms shows how Human speech is produced by complex interactions between the diaphragm, lungs, throat, mouth and nasal cavity [1] . The processes which control speech Phonation is the process of converting air pressure into sound via the vocal folds, or vocal cords as they are commonly called , Resonation is the process by which certain frequencies are emphasized by resonances in the vocal tract ,and Articulation is the process of changing the vocal tract resonances to produce distinguishable sounds. In general, in its most simplistic form speech can be viewed as a stochastic process involving two principal dimensions - time and frequency. The complexity of the speech recognition task lies in the fact that a given utterance can be represented by an effectively infinite number of time- frequency patterns. A human speech signal is produced by moving the vocal tract articulators towards target positions that characterize a particular sound [2]. There are several viable methods currently used for speech recognition including template matching, acoustic-phonetic recognition and stochastic processing. The human auditory system is based on a time-frequency analysis of sounds. The information received by the human ears can be described most conveniently as non-linear auditory responses to frequency selectivity and perceived loudness [3]. The Multiwavelet Transform is widely used in the extraction of features in all the types of signals, especially for extracting features in the speech sound signal. The wavelet and multiwavelets transformations are directly applicable only to one – dimensional signals. The two main categories of methods for doing this are separable and non separable algorithms which are mostly applied for 2-D signals,where Separable methods simply work on each dimension in series. The typical approach is to process each of the rows in order and then process each column of the result. Non-separable methods work in both signal dimensions at the same time. While non-separable methods can offer benefits over separable methods, such as savings in computation, they are generally more difficult to implement [4] .

Preprocessing for multiwavelets

As mentioned before , multiwavelets filter banks require a vector –valued input signal .There is a number of ways to produce such signal from 2-D data signal perhaps the most obvious method is to use adjacent rows and columns of the data [4] .

However, this approach does not work well for general multiwavelets and leads to reconstruction artifacts in the lowpass data after coefficient quantization.

This problem can be avoided by constructing "constrained" Multiwavelets , which possess certain key properties . Unfortunately, the extra constraints somewhat restrictive; data compression tests show that constrained Multiwavelets do not perform as well as some other Multifilters.

Another approach is to first split each row or column into two half-length signals and then use these two half signals as the channel input into the Multifilter.

A naive approach is to simply take the odds samples for one signal and the even samples for the second signal. This approach doesn't work well because it destroys the assumed characteristics of the input signal.

It is generally presumed that data will locally well-approximated by low-order polynomials, usually constant, linear ,or quadratic. The high-pass filter is designed to give a uniform zero output when the input has this form. Taking alternating data points as the filter alerts character of the input signal, hence the filter output will no longer be forced to zero ,reducing compression performance .But there is a way around this problem :one may first prefilter the two half length signals before passing them into the Multifilter .

Over sampled scheme: repeated row preprocessing

In multiwavelets setting , GHM multiscaling and multiwavelets function coefficients are 2*2 matrices ,and during transformation step they must multiply vector(instead of scalars) .This means that multifilter bank need 2input rows .

so the most obvious way to get two input rows from a given signal is to repeat the signal .Two rows go into the multifilter bank .This procedure is called "Repeated Row" which introduces over sampling of the data by factor of 2.

for a given scalar input signal {Xk} of length n (N is assumed to be power of 2 and so is of even length), repeated row preprocessing of this signal is by repeating the input stream with the same stream multiplied by a constant α .

So the input leangth-2 vector is formed from the original as

$$\begin{pmatrix} X_k \\ \alpha X \end{pmatrix} \text{ Where } k=0.1.2\dots n-1 \quad \text{---- (1)}$$

Here α is constant: it is typically chosen so that if $X_k=C=\text{constant}$, for all k, then the output from the high-pass multifilter is zero .This can always be done if the system has approximation order higher than zero. For the GHM case , $\alpha=1/\sqrt{2}$ is selected since

$$[G0+G1+G2+G3] \begin{pmatrix} C \\ \frac{C}{\sqrt{2}} \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 8 & -\frac{10}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C \\ \frac{C}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{---- (2)}$$

The output from the low-pass multifilter is simply a scaled version of input,

$$[H0+H1+H2+H3] \begin{pmatrix} C \\ \frac{C}{\sqrt{2}} \end{pmatrix} = \frac{1}{5} \begin{pmatrix} \frac{6}{\sqrt{2}} & 4 \\ 4 & \sqrt{2} \end{pmatrix} \begin{pmatrix} C \\ \frac{C}{\sqrt{2}} \end{pmatrix} = \sqrt{2} \begin{pmatrix} C \\ \frac{C}{\sqrt{2}} \end{pmatrix} \quad \text{----(3)}$$

Speech recognition via multiwavelet

There is an extensive amount of information in the speech signal only some of which is relevant to the selection of the correct machine response. A critical task is to extract from all that information all those parts that convey the message.

* Feature extraction is done by Multiwavelet Transform.

However, it is extremely difficult to do this. An objective of speech analysis for speech recognition is therefore to extract useful information from the signal by approaching directly in the time domain, or by some related techniques in the frequency domain .In some way or another most proposed schemes are based on, or can be related to, short-time frequency analysis of the speech waveform.

Proposed speech recognition system using multiwavelet transform

1- Preprocessing the data

The first thing to be done before getting the data to be processed in the system is data acquisition processing. The data to be processed is stored on the computer as a "WAVE PCM sound files 8-bit mono" as shown in figure (1).

2- Data framing

It may simply consist of placing the data into fixed-size blocks, thus the speech samples are taken in frames small enough not to cover too many changes in this speech signal but large enough to discover meaningful information about the signal. Each block is of 256 bit as shown in figure (2)

3- Choosing data windowing

Two main types of window functions that had been used in speech recognition namely the **Hamming window** which is used here

$$W_n = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) & \text{if } 0 \leq n \leq N-1 \\ 0 & \text{if otherwise} \end{cases} \quad \text{-----(4)}$$

and the **Rectangular window**

$$W_n = \begin{cases} 1 & \text{if } 0 \leq n \leq N-1 \\ 0 & \text{if otherwise} \end{cases} \quad \text{---- (5)}$$

4-Feature extraction is done by multiwavelet transform

Using the over-sampled scheme of processing, the DMWT matrix is double in dimension compared with the input which should be a square matrix (of 4*1 minimum).

5- Creating the reference

Each sample value in the reference file is computed by taking the average value of four samples for the same data (four speakers for the same phoneme).

6- Finding the best match

Finding the distance of the input data to every frame value in the reference file using the Euclidian distance measurement function $E=(x-ref)^2$. The frame with the minimum distance is said to be the "Best Match" the table of result shows in figure (3).

A flow diagram for the proposed speech analyze method via multiwavelet shown in figure (4)

Conclusions

This paper presents a new technique for speech analyze through the use of multiwavelet to extract features by using (Repeated Row Preprocessing). The proposed system depends on the Euclidian distance measurement for finding the distance of the input data to every frame value in the reference file to compute the average value of four samples for the same data (four speakers for the same phoneme) to determine the best result. We can apply this method to a wider set of Phonemes (more than 7 speakers).

Also it has been seen that the bandwidth of the hamming window is about twice the bandwidth of the rectangular window of the same length.

This method could be developed to use as a preprocessing for the development of a speech - to - text recognition system.

References

1. Steven, E. and Golowich(1998) A Support Vector/HMM approach to phoneme recognition,October,14Bell labs,Lucent Technology.
2. Mohamed Debyeche (2000) Phoneme recognition system basec on HMM with distributed VQ codebook,IEEE Trans on comm., 28.
3. Jenkins, W.K,(1986) ,Recent advances in residue number techniques for recursive digital filtering, IEEE Transactions on Acoustics, Speech, and Signal Processing,ASSP-27, 19-30, February.
4. Waled, A. Mahmoud ,Quantization Techniques for the classification and recognition of speech signals(1986),ph.d thesis,University of Wales.

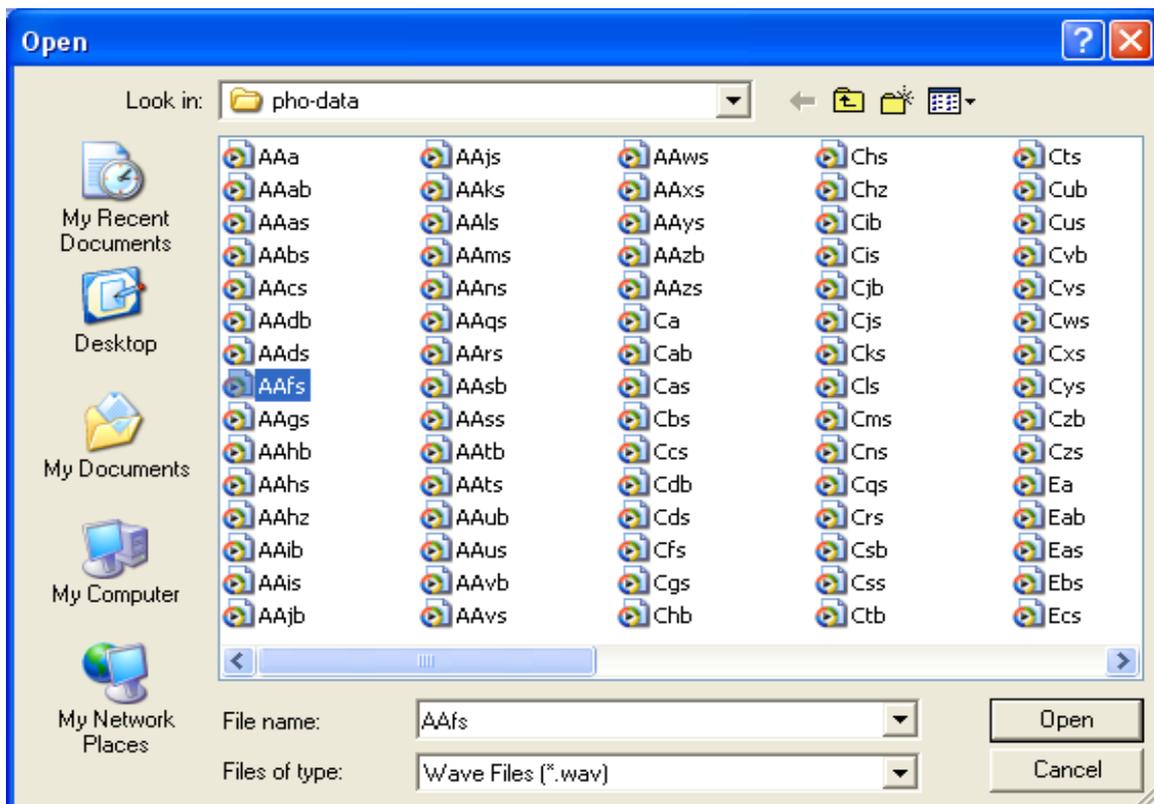


Figure No. (1): shows the opening of a wave file interface in the proposed system

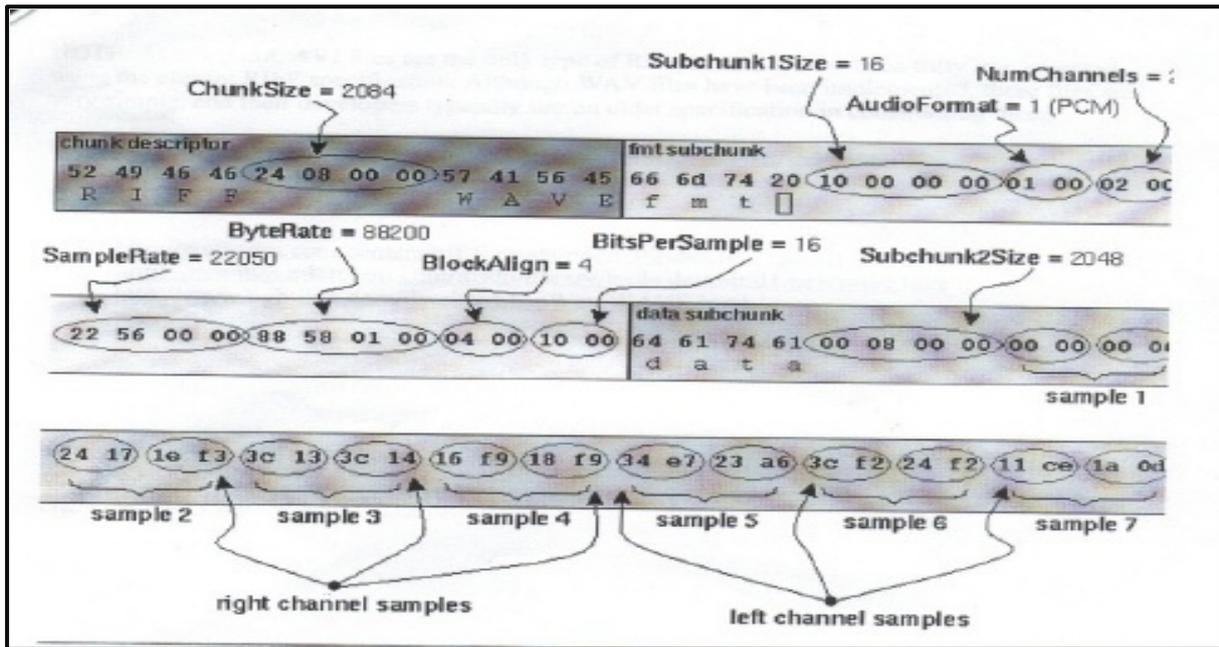


Figure No.(2)The data to be processed is stored on the computer as a " WAVE PCM sound files 8-bit mono

الحرف	الحرف	الحرف	الحرف	الحرف	اللفظة								
ش	س	ز	ر	ذ	د	خ	ح	ج	ث	ت	ب	ا	
203	156	111	222	125	99	117	85	216	190	188	204	22	ا
69	98	136	205	55	122	43	56	212	79	122	17	77	ب
54	123	76	89	46	87	123	34	260	33	7	189	67	ت
80	200	45	87	66	29	84	142	34	20	76	132	45	ث
32	50	122	99	65	112	55	66	13	98	34	88	132	ج

Figure No. (3)Table of result the frame with the minimum distance is said to be the best

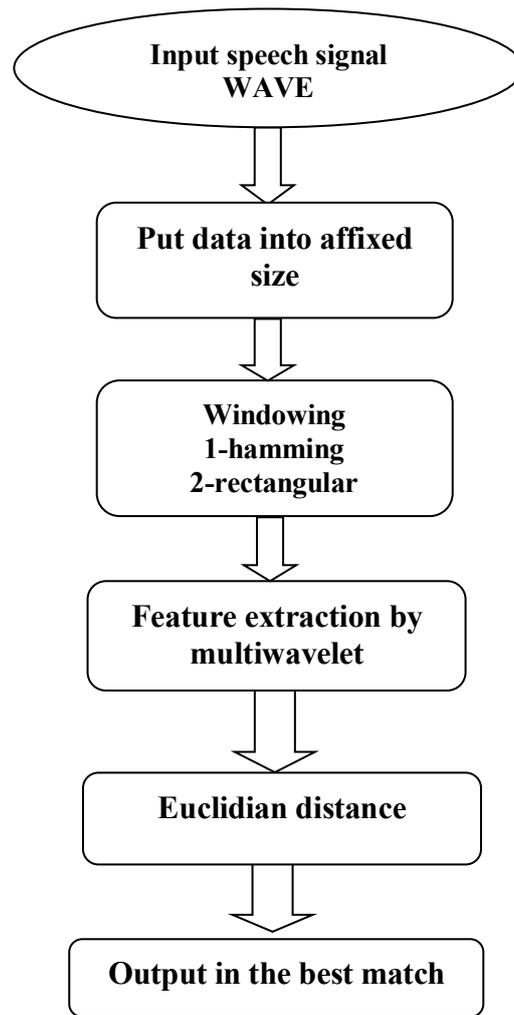


Figure No. (4):A flow diagram for the proposed speech analyze method by using multiwavelet

طريقة مقترحة لتحليل الكلام باستخدام التحويل متعدد المويجة

بيداء جعفر صادق الخفاجي

قسم علوم الحاسبات / كلية التربية للعلوم الصرفة (ابن الهيثم) / جامعة بغداد

استلم البحث في 7 نيسان 2014 ، قبل البحث 14 نيسان 2014

الخلاصة

الكلام هو اول طرائق الاتصال بين البشر وحتى قبل اختراع الكتابة. في هذه الورقة، نصف طريقة جديدة لتحليل خصائص الكلام المهمة وباستخدام التحويل متعدد المويجة وباحد انواعه التي هي طريقة تكرار تجهيز الصفوف النظام المقترح يعتمد على طريقة اقليدس في حساب المسافة ومعاملات التحويل متعدد المويجة في حساب فرق المسافة كل قيمة في فايل المصدر تحسب باخذ متوسط قيمة اربع عينات لترتيب البيانات نفسها (اربعة متحدثين للصوت نفسه) ولكي نحقق افضل النتائج نعتمد بتحقيق اقل مسافة بطريقة اقليدس، استخدمت لغة البرمجة فيجوال بيسك(6)

الكلمات المفتاحية:تحويل متعدد المويجة،طريقة اقليدس في قياس المسافة ، نافذة المبالغة،المعالجة بالصفوف المتكررة