IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.
PP 12-16
www.iosrjournals.org

Speech recognition by Dynamic Time Warping

C.B.Kare[1], Mrs.V.S.Navale[2]
[1](Department of E&TC, AISSM's College of Engineering, Savitribai Phule Pune University, Pune, India)
[2](Department of E&TC, AISSM's College of Engineering, Savitribai Phule Pune University, Pune, India)

Abstract: *Automatic template matching has the interest of researchers for decades. Although there are various applications for which the technique is already used in daily life, a number of problems still have to be solved. At this moment, one of the biggest problems is the low user acceptance: the systems are not accurately enough; the mistakes that recognizers make are usually not very understandable to humans, which can frustrate the users of the systems. This is about the use of the Dynamic Time Warping (DTW) algorithm. We claim that the results of a recognizer based on the DTW-algorithm template matching are more "intuitive" to humans than the results of other recognizers. Because humans can understand the errors that system makes, this will probably improve the user acceptance of template matching systems. Furthermore, given that users are aware of what they have to do for the system to understand them, the recognizer is expected to yield better recognition performances. The way the system works also adds a number of new possibilities for applications of template matching A more human-like system, for example, can be of use in the field of controlling robotics .many type of embedded system through template matching.*

Keywords*- DTW; endpoint detection; match; short-time zero crossing, short-time energy*

I. INTRODUCTION

Recent advances in speech technology and computing power have created a surge of interest in the practical application of speech recognition.

Speech is the primary mode of communication among humans. Our ability to communicate with machines and computers, through keyboards, mice and other devices, is an order of magnitude slower and more cumbersome. In order to make this communication more user-friendly, speech input is an essential component.

There are broadly three classes of speech recognition applications, in isolated word recognition systems each word is spoken with pauses before and after it, so that end-pointing techniques can be used to identify word boundaries reliably. Second, highly constrained command-and-control applications use small vocabularies, limited to specie phrases, but use connected word or continuous speech.

Finally, large vocabulary continuous speech systems have vocabulary of several tens of thousands of words, and sentences can be arbitrarily long, spoken in a
Natural fashion. The last is the most user-friendly but also the most challenging to implement.

Although there are other advanced techniques in template matching recognition (hidden Markov modeling or neural network techniques), DTW is still used in the small-scale embedded systems (e.g. cell phones, mobile applications) because of the simplicity of its hardware implementation, the straightforwardness and speed of the training procedure. A simple template matching recognition scheme using DTW is represented in Fig. 1.2. The classic dynamic-time warping (DTW) algorithm uses one model (template) for each word to be recognized.

The main problem is to find the best reference template fore certain word. Choosing the appropriate reference template is a difficult task. In order to increase the recognition rate, a better solution is to increase the number of templates for the same word. But this technique also increases the memory size and the computing time. Another solution is combining the DTW technique with vector quantization (VQ) in order to create classes

II. Literature Review

Some reference papers are studied to understand the Dynamic time warping algorithm Review of some some paper is given below

Tiberius Zaharia, et.al. [1] proposed that DTW algorithm compares the parameters of an unknown spoken word with the parameters of one or more reference templates. The more reference templates are used for the same word, the higher is the recognition rate. But increasing the number of reference templates for the same word to recognize, leads to an increase in memory resources and computing time. The proposed algorithm is used in the learning phase and combines the advantages of DTW and Vector Quantization (VQ); instead of

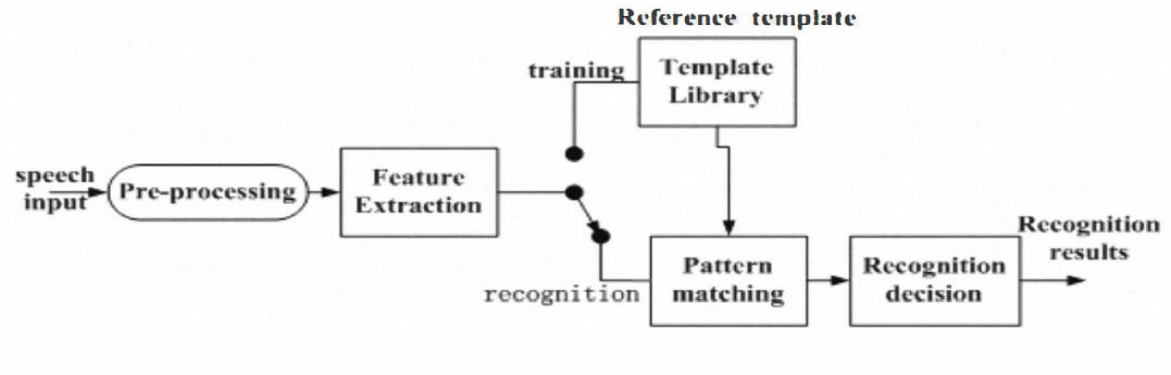
storing multiple reference templates, it stores only one reference model for each word and that reference is based on classes (like in the vector quantization method), each class is represented by a centroid (or codeword). In the recognition phase, the parameters of the unknown utterance are compared to the centroids of the reference model. This solution increases the speed of calculation in the recognition phase and reduces the quantity of used memory.Mfcc feature are used to speech and accuracy of this system is 80% [1].

III. SPEECH RECOGNITION TECHNIQUES

Speech-enabled applications over wireless networks and the World Wide Web (WWW) employing speech recognition are recently attracting more and more attention. At present google search is based on the words in text mode. ASR can be used to give the input words to google to be searched, through speech instead of text. Recently this application is also introduced in the google search. Managing the data of students, like entering the marks, attendance, and preparing progress reports is tedious process. If IWR is used for entering the data, the work load on the teacher will be reduced. Therefore ASR for IWR based HMM based word models is built for entering data of the students of a class and tested the performance of the system. However HHMs developed for word models can also used for connected word recognition and it is to spell the connected words as compared isolated words. Therefore, in this chapter the performance of the ASR system for Isolated words is given and next chapter will give the performance of connected word recognition.

Word Recognition is where a word uttered by the user has to be recognized by the speech recognition system. This is possible by the Reference Models stored in the database corresponding to each word intended to be recognized. Thus while performing the recognition an uttered word is compared to each of these models. Only the words having the Models in the system, can be recognized. If a new word is spelt for recognition, it will recognize as one of the word having model or simply give as a new word. The inputs to the Word Recognition system are stored MODELS and the MFCC features of the word uttered (TestFeatures).

The recognition process is simply matching the incoming speech with the stored Models In the recognition process, Forward Algorithm of Dynamic Time Warping, is used for calculating the Cost. All the MODELS (Reference Features) are given as Reference Features to the DTW, one after the other along with the features of the word uttered. The MFCC features of the word uttered for recognition are the test features applied to the DTW algorithm. Thus the DTW algorithm gives a cost for each model and the test features. The Model with the lowest distance measure (cost) is the recognized word. The word corresponding to the model with lowest cost is the recognized word. Hence the best match (lowest distance measure) is obtained from dynamic programming.



Block diagram of speech recogination

MFCC Technique
Mel Frequency Cepstral Coefficients (MFCC):

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior. For frequencies lower than 1 kHz, human ear hears tones with a linear scale instead of logarithmic scale for the frequencies higher that 1 kHz. In other words, MFCC is based on known variation of the human ear's critical bandwidth with

frequency. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech.

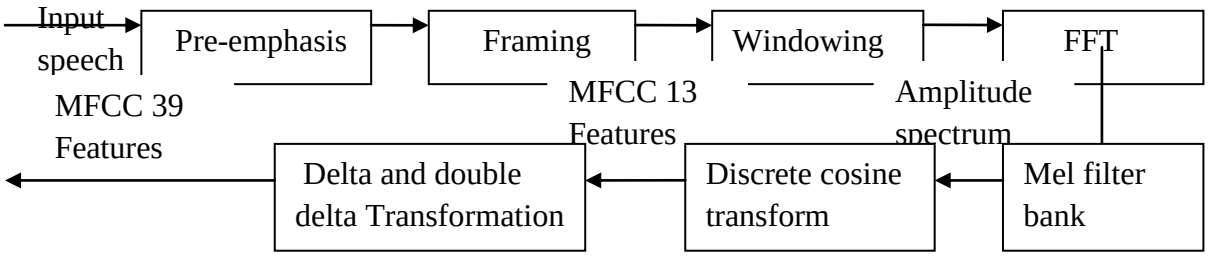


Figure Block diagram of MFCC extraction

Mel-spectrum:

MFCC is based on the human auditory system. The human perception of the intensity of speech or audio signal with respect to energy does not follow a linear scale. Similarly the human perception of distinguishing two speech or audio signals at different frequencies does not follow a linear scale. Thus for each tone with an actual frequency f measured in Hz, a subjective pitch is measured on a scale called the 'Mel Scale'. The Mel frequency scale is linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz. As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.

The speech signals have most of their energy in the low frequencies. It is also very natural to use a Mel-spaced filter bank to analyze a speech.

The following approximate formula computes the Mels for a given frequency *f in* Hz:

$$mel(f) = 2595 * \log_{10}(1 + \frac{f}{700})$$

IV. FIGURES AND TABLES

The fig.1 shows aerial car image captured by aerial vehicle. These types of aerial car images are used for car detection techniques.

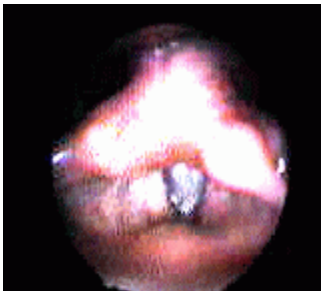


Figure 3.1 Closed glottis

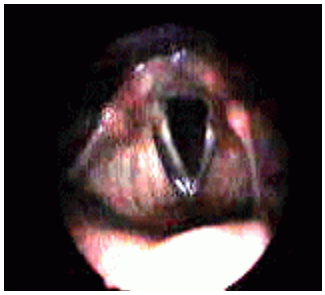


Figure 3.2 Open glottis

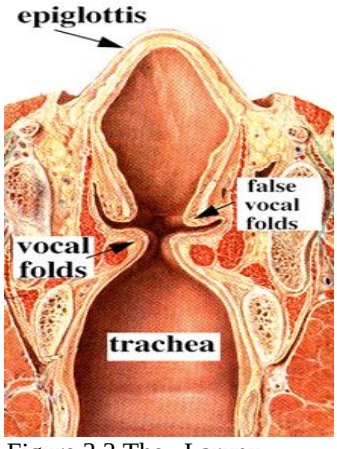


Figure 3.3 The Larynx

V. COMPARISON

Word spoken(10 times)	word correctly recognized	Accuracy in %

Zero	**7**	70%
One	**6**	60%
Two	**7**	70%
Three	**6**	60%
Four	**6**	60%
Five	**6**	60%
Six	**7**	70%
Seven	**7**	70%
Eight	**6**	60%
Nine	**7**	70%
Ten	**7**	70%
Eleven	**7**	70%
Twelve	**6**	60%
Thirteen	**6**	60%
Fourteen	**6**	60%
Fifteen	**5**	50%
Sixteen	**6**	60%
seventeen	**7**	70%
Eighteen	**7**	70%
Nineteen	**7**	70%
Twenty	**8**	80%
Thirty	**8**	80%
Forty	**7**	70%
Fifty	**8**	80%
Sixty	**7**	70%
Seventy	**7**	70%
Eighty	**7**	70%
Ninty	**7**	70%
	Total=190	**Overall accuracy=68%**

VI. CONCLUSION

This algorithm presents a new and simple training technique to prepare the reference templates for DTW-based speech recognition systems. Significant improvements have been obtained with this training technique. This DTW is used for small vocabulary application. Using DTW it is possible to recognize the word of different length.MFCC is the best feature for speech recognition using this feature we can achieve the high accuracy. Using the 13 MFCC and DTW accuracy is 68%,Using the 26 MFCC and DTW accuracy is 78%, and Using the 39 MFCC and DTW accuracy is 90%.

Acknowledgements

The author would like to thank Mrs. V.S.Navale for valuable support and guidance for completing this survey paper work.

REFERENCES

[1] Tiberius Zaharia, Svetlana Segarceanu, Marius Cotescu, Alexandru Spataru "Quantized Dynamic Time Warping (DTW) algorithm"IEEE.2010.

[2] Guangyu Kang "Variable sliding window DTW speech identification algorithm" IEEE.2009.
[3] Oscar, Luis Mengibar-Pozo,Andrzej,Pacut "A new algorithm for signature verification system based on DTW" IEEE .2008
[4] Ivan Kraljevski, Slavcho Chungurski, Zoran Gacovski, Sime Arsenovski Perceived templete matchingquality estimation using DTW algorithm 16th Telecommunications forum TELFOR 2008.
[5] Heng-Da Cheng et al. , A VLSI architecture for dynamic time-wrap recognition of handwritten symbols, IEEE ASSP, Vol. 34, N.3, Jun 1986.
[6] G. R. Quenot et al. , A Dynamic Programming Processor for Templete matchingRecognition, IEEE JSSC, Vol. 24, N(F9)q.20, April 1989.
[7] Stan Salvador, Chan, FastDTW: Toward Accurate Dynamic Time Warping in Linear Timeand Space, IEEE Transactions on Biomedical. Engineering, vol. 43, no. 4
[8] Tiberius Zaharia, Svetlana Segarceanu, Marius Cotescu, Alexandru Spataru