Laplacian Operator as Speaker Identification Parameter

***S. K. Jamil** ****S. M. Ali** ****M. S. Al-Zewary**

*Dept. of Physics, College of Science, Al-Mustanceriya University.
**Remote Sensing Unit, College of science, University of Baghdad.

Abstract:

New speaker identification test's feature, extracted from the differentiated form of the wave file, is presented. Differentiation operation is performed by an operator similar to the Laplacian operator. From the differentiated record's, two parametric measures have been extracted and used as identifiers for the speaker; i.e. mean-value and number of zero-crossing points.

استخدام معامل لابلاس كمعلمة للتحقق من أصوات المتكلمين

سلوان كمال جميل*، صالح مهدي علي و الزيواري**

**** قسم الفيزياء – كلية العلوم – الجامعة المستنصرية***

***** وحدة التحسس النائي – كلية العلوم – جامعة بغداد***

الخلاصة:

تم في هذا البحث استحداث خاصية جديدة للتحقق من أصوات المتكلمين مستخلصة من الحالة التفاضلية للجل الصوتي. تم تنفيذ عملية التفاضل بمعامل مشابه لمعامل لابلاس. تم حساب معاملين من التسجيل التفاضلي كمقاييس لتمييز الأصوات والتحقق من المتكلمين: تتمثل المعايير المشتقة هذه بالقيمة المعدل وعدد نقاط العبور الصفرية للتسجيل بعد أجراء عملية التفاضل.

Introduction

Speech is one of the most important tools ever used in communication. Several ways can be followed in categorizing speech through communication systems. For example, speech can be represented in terms of its message content or information. An alternative way of speech characterization is that based on the message information carried by the signal, e.g. the form of the acoustic wave. Although information theoretic ideas have played a major role in sophisticated communication systems, speech representation based on the waveform, or on some parametric models has also found to be useful in practical applications. In considering the process of speech communication, it is helpful to begin thinking of a message represented in some abstract form in the brain of the speaker and then converted to an acoustic signal. The information that is communicated through speech is, intrinsically, in a discrete nature; i.e., it can be represented by concatenation of elements from a finite set of symbols. The symbols from which every sound can be classified are called Phonemes. Each language has its own distinctive set of phonemes, typically ranged between (30 to 50). For example, English can be represented by a set of around 42-phoneme [1].

In the last four decades, considerable numbers of researches have been published aimed for developing methods of recognizing speaker, automatically, based on voice input alone. Such system could be considerable benefit in numerous situations; i.e. in both civilian and military environment. Generally, there are two primary functions of automatic speaker recognition; the first is speaker verification, where the system either accepts or rejects the identity claim made by the user based on his or her voice characteristics. While, the second is speaker identification, in which the system determines which speaker from a known set best matches the unknown input voice [2]. The use of the term Automatic Speaker Recognition (ASR) indicates both speaker verification and speaker identification. In fact a numerous ASR systems have been developed in the past 20 years, mostly they followed the same basic steps, illustrated figure (1).

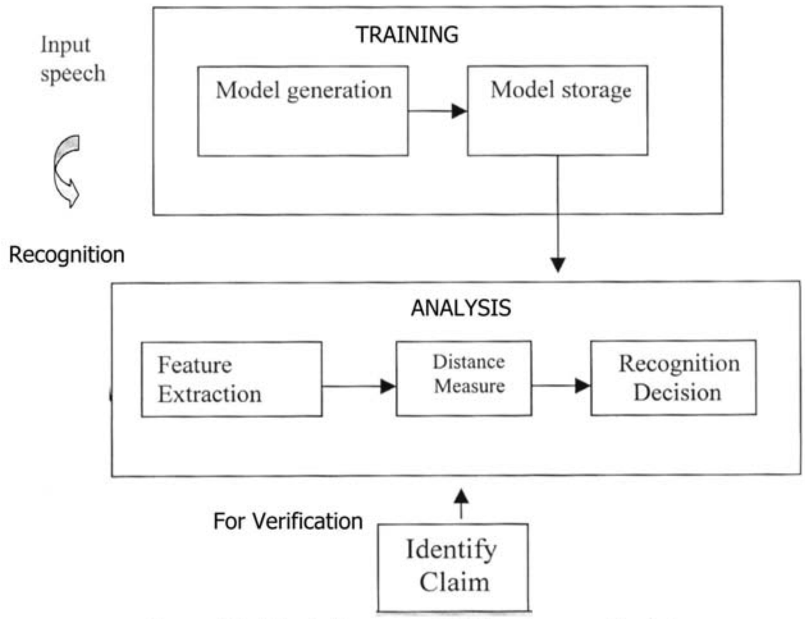

Figure (1): Block diagram of ASR system methodology

ASR System Design

Generally, to design an ASR system, the following points should be taken in consideration, these are: is the system designed to be used for identification or for verification, is the test text-dependent or text-Independent, what will be the tested speech's quality, where the test will carry on, what is the required accuracy, how many speakers have to be identified, what will be the voice variability of each speaker and, finally, who will use the system, for the details see [3].

Differential “lAPLACIAN” Test Operator

At the beginning, it should be noted that our introduce “Laplacian” feature was not the only parameter adopted in our system that has been designed [3] to perform identification and verification operations. In fact, too many parametric test features were adopted, hopefully they will be presented as a series of short papers. The final one will contain the general procedures involved at that system. In our research, each processed record wave file is first divided into frames, each of lengths “256 elements”. Each frame, then, considered as a sequence of signal elements “S(i)” and their differentiated form is presented by:

$$\delta S(x) = (\mathrm{L}-1).S(x) - \sum_{\mathrm{i}=-\mathrm{L}/2}^{\mathrm{L}/2} S(x+\mathrm{i}), \quad \text{for} \quad \mathrm{i} \neq 0 \tag{1}$$

Where “L” is chosen to be any odd integer. As can be seen, equation (1) takes the conventional of “Laplacian” form when L=5; i.e.

$$\begin{aligned} \nabla^2 S(x) &= 4S(x) - \sum_{\mathrm{i}=-2}^{2} S(x+\mathrm{i}), \quad \text{for} \quad \mathrm{i} \neq 0 \\ &= 4S(x) - [S(x-2) + S(x-1) + S(x+1) + S(x+2)] \end{aligned} \tag{2}$$

In fact, we have adopted the length “L=25” in performing eq.(1). Instead of using all differentiated values as identifiers, only two parameters were adopted, these are; the differentiated record’s mean “μ”, presented by;

$$\mu = \frac{1}{M} \sum_{j=1}^{M} \left[\frac{1}{N} \sum_{i=1}^{N} F_j(i) \right] \tag{3}$$

Where, M” represents number of frames within each record, and “N” is the number of differentiated elements per each frame.

The second feature that has been adopted was the Zero-Crossing Feature (ZCF). This feature is computed as the number of times the differentiated signals change their attribute; i.e. the signal sign changed from "–" to "+" or vice versa. The record’s “ZCF” value is determined from;

$$\mathrm{ZCF} = \sum_{\mathrm{i} \in \mathrm{R}} (\text{if} \quad \mathrm{S(i-1)} \times \mathrm{S(i+1)} \leq 0) \tag{4}$$

Where; R refers to the whole record’s elements.

During our research course, the test is performed on “text-dependent and independent” of more than 50 voluntaries. Different differential records of three of these peoples were selected, illustrated in figure (2). Table (1) demonstrates the mean and zero-crossing values of different speakers. The first couples of files belong to the same persons. As can be seen, pairs of “ZCF” values belong to the same person are very close [e.g. (1416,1699), (2658,2538) etc.). The next five files belong to other persons, their “ZCF” values, as seeing, ranged (from 2658 to 2152).

Table (1) ***Means and zero-crossing values for different differentiated records***		
File-name	***Mean value***	***Number of zero-crossing points***
Person1-a	6	1416
Person1-b	4	1699
Person2-a	4	2658
Person2-b	4	2538
Person3	5	2251
Person4	5	2122
Person5	4	2152
Person6	5	1557
Person7	4	2174
Person8	4	1940

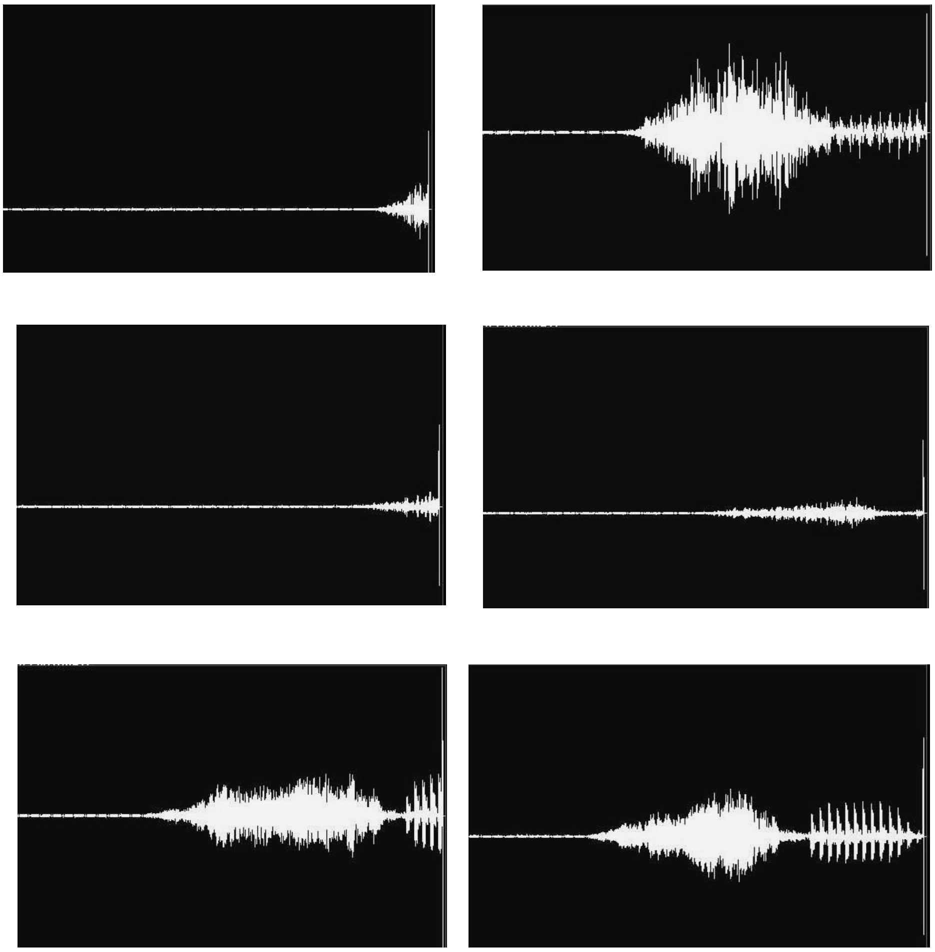

Figure (2): Differential forms of different records, using Laplacian operator, each two adjacent plots belong to the same person.

References

[1]Atal, S., (1972), "Automatic Speaker Recognition based on Pitch Contours," J. Acoustics, Soc. Am., Vol. 52.

[2]Mohankrishnan N., Shridhar M. and Sid-Ahmed M. A., (1982), "A Composite Scheme for Text-Independent Speaker.

[3]Recognition," Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Proc. Al-Zewary, M.S. (2002), "Computer System for Identification and Verification of Human Voices," Ph.D. thesis, University of Baghdad, College of Science, Dept. of Physics.