Expert Systems with Applications 36 (2009) 6295–6300

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An expert system for speaker identification using adaptive wavelet sure entropy

Derya Avci

Bahcelievler Primary School, Elazig, Turkey

A R T I C L E I N F O

Keywords:
Discrete wavelet transform
Wavelet adaptive network based fuzzy inference system
Intelligent speaker identification
Speech signal
Feature extraction
Discrete wavelet entropy
Expert system

A B S T R A C T

In this study, an expert speaker identification system is presented for speaker identification using Turkish speech signals. Here, a discrete wavelet adaptive network based fuzzy inference system (DWANFIS) model is used for this aim. This model consists of two layers: discrete wavelet and adaptive network based fuzzy inference system. The discrete wavelet layer is used for adaptive feature extraction in the time–frequency domain and is composed of discrete wavelet decomposition and discrete wavelet entropy. The performance of the used system is evaluated by using repeated speech signals. These test results show the effectiveness of the developed intelligent system presented in this paper. The rate of correct classification is about 90.55% for the sample speakers.

1. Introduction

Speaker identification and speaker identification are complex problems. Both have fuzzy effect on them which adds to the hardness (Rabiner and Juang, 1993). During speaker identification, separation to the vowels makes it difficult for analysis. Speech/speaker identification is a problem studied for many years and of course numerous network models and digital signal processing techniques have been tested for this problem. Speech features which are usually obtained via Fourier transforms (FTs), short time Fourier transforms (STFTs) or linear predictive coding techniques are used for some kind of expert speaker identification (ESI). These methods accept signal stationarity within a given time frame and may therefore lack the ability to analyse localised events correctly. Moreover, the LPC method accepts a particular linear (all-pole) model of speech/voice production which severely speaking is not the case. Other methods based on Cohens general class of time-frequency distributions such as the Born-Jordan, Wigner-Ville, Cone-Kernel and Choi-Williams methods have also found use in speaker identification applications but have the drawback of introducing unwanted cross-terms into the representation. The discrete wavelet transform (DWT) copes with some of these problems. The DWT can ensure a constant-Q analysis of a given signal by projection onto a set of basis functions that are scale variant with frequency. Wavelets are shifted scaled version of original or mother wavelets. The wavelet families are commonly orthogonal to one another, important since this situation gives computational efficiency and ease of numerical implementation. Other factors influencing the selection of DWT over conventional methods include their ability to determine localised features. In addition, developments aimed at generalisation such as the Best-Basis Paradigm of Coifman and Wickerhauser (1992) provide more flexible and useful representations.

Expert speaker identification (ESI) has been under investigation by a large number of researches for about four decades (Reynolds, Quatieri, & Dunn, 2000). From a commercial viewpoint, expert speaker identification is a technology with potentially large market due to the applications of widely ranges from automation of operator-assisted service to speech-to-text aiding system for hearing-impaired individuals (Reynolds et al., 2000).

Most of the ESI works are generally supported by commercial corporations (Kadambe and Srinivasan, 1994). Opinions of the researchers, who study in speaker identification area, as summarized in Kadambe and Srinivasan (1994) and Mallat (1989) appear to ignore the benefits that can be gained by proper transformations of the input signal. The main task in Expert speaker identification (ESI) is to separate various speaker classes (Evangelista, 1994; Kadambe and Boudreaux-Bartels, 1992; Kadambe and Srinivasan, 1994; Mallat, 1989). In literature, some researchers have explored the use of wavelets to provide a richer feature space (Evangelista, 1993, 1994; Maes, 1994; Saito, 1994; Szu et al., 1992). Nevertheless, there is little evidence of widespread use of this technique (Kadambe and Srinivasan, 1994). In Saito (1994), preprocessing the data allows easier subsequent feature extraction and increased resolution. The signal was transformed from a time domain to a frequency domain using the Fourier transform by engineers (Buckheit and Donoho, 1995). Despite this is useful for some applications, this transform was not excess useful for expert speaker identification using real speech signals (Buckheit and Donoho, 1995). Because the Fourier transform tells us that a feature occurs somewhere in the signal, but not where. Wavelets which bring a new tool to the speech signal classification. It can be said that the benefits of using wavelets (Visser et al., 2003) which are the

E-mail address: derya2344@hotmail.com

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.07.012

new transforms are local; i.e., the event is connected to the time when it occurs. In studies wavelets used for speaker identification, it has been found that the original feature space can be augmented by the wavelet coefficients and will yield a smaller set of more robust features in the final classifier (Coifman and Wickerhauser, 1992; Visser et al., 2003; Wesfried and Wickerhauser, 1993).

Artificial neural network performance is depended on the size and quality of training samples (Kosko, 1992; Zhang, Walter, Miao, & Lee, 1995). When the number of training data is small, not representative of the possibility space, standard neural network results are poor (Nava & Taylor, 1996a). Incorporation of neural fuzzy techniques can improve performance in these cases (Nava & Taylor, 1996b).

Fuzzy theory has been successfully used for many speech/speaker identification applications (Zadeh, 1987). These applications show that fuzzy theory can be used to improve neural network performance. There are many advantages of fuzziness, one of which is the ability to handle blur data. Artificial neural network (ANN) is known to be excellent classifier, but its performance can be prevented by the size and quality of the training set. Efficiently, network results by using incorporation some fuzzy techniques and neural networks are obtained (Nava & Taylor, 1996a). So, standard ANN solution is poor, because the training sequence does not converge and the training results are vastly improved (Nava & Taylor, 1996a).

In this paper a novel method, which is an expert system for Turkish speaker identification, is introduced. It will aid to the expert speaker identification and enable further research of speaker identification to be developed. A combination of wavelet signal processing and adaptive network based fuzzy inference system (ANFIS) to efficiently extracting the features from pre-processed real speech signals for the purpose of ESI among variety speakers. An algorithm called the intelligent system is developed by using pattern identification approximations in this study. For expert speaker identification area, the novelties presented in this paper can be summarized as follow:

An effective adaptive feature extraction method developed in this study increases percentage of the correct word identification. The second novelty is the development of the discrete wavelet adaptive network based fuzzy inference system (DWANFIS) model as an efficiently classification method in ESI area.

Here, Turkish language speech signals are used for ESR. Experimental setup is used for obtaining the real Turkish speech signal data sets. At the ESR experiment set used in this study, speech signals are transmitted to the computer by using a microphone and an audio card which has 44 kHz maximum sampling frequency.

The paper is organized as follows: In Section 2, discrete wavelet transform, in Section 3, discrete wavelet adaptive network based fuzzy inference system, in Section 4, DWANFIS for speaker identification, in Section 5, obtained results are presented respectively and in Section 6, Discussions and Conclusion are given.

2. Discrete wavelet transform

Over the last few decades, wavelet analysis has been proven an effective signal processing technique for a variety of problems (Davis and Mermelstein, 1980). Especially, in feature extraction schemes designed for the purpose of speaker identification, wavelets have been used too much. DWT instead of discrete cosine transform in the feature extraction stage was used in the first approach (Tufekci and Gowdy, 2000). DWT was applied directly on the speech signal in second approach. According to this situation, either wavelet coefficients with high energy are taken as features (Long and Datta, 1996).

Because of its suitability for analyzing non-stationary signals, DWT has become a powerful alternative to the Fourier methods in many speech/speaker identification applications (Buckheit and Donoho, 1995; Coifman and Wickerhauser, 1992; Evangelista, 1993, 1994; Ha et al., 2005; Kadambe and Boudreaux-Bartels, 1992; Kadambe and Srinivasan, 1994; Maes, 1994; Mallat, 1989; Reynolds et al., 2000; Saito, 1994; Szu et al., 1992; Visser et al., 2003; Wesfried and Wickerhauser, 1993).

The main advantage of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time–frequency resolution in all frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets lack of the requirement of stationary (Coifman and Wickerhauser, 1992; Evangelista, 1993, 1994; Ha et al., 2005; Kadambe and Boudreaux-Bartels, 1992; Kadambe and Srinivasan, 1994; Maes, 1994; Mallat, 1989; Pati and Krishnaprasad, 1993; Reynolds et al., 2000; Saito, 1994; Szu et al., 1992; Visser et al., 2003).

DWT uses the fact that it is possible to resolve high frequency components within a small time window, and only low frequency components need large time windows. This is because a low frequency component completes a cycle in a large time interval whereas a high frequency component completes a cycle in a much shorter interval. Therefore, slow varying components can only be identified over long time intervals but fast varying components can be identified over short time intervals. DWT can be regarded as a continuous time wavelet decomposition sampled at different frequencies at every level or stage. The DWT functions at level m and time location t_{m} can be expressed as

$$d_k(t_k) = x(t)\psi_k\left(\frac{t - t_k}{2^k}\right) \tag{1}$$

where psi_k is the decomposition filter at frequency level m. The effect of the decomposition filter is scaled by the factor 2^k at stage m, but otherwise the shape is the same at all stages.

3. Discrete wavelet adaptive network based fuzzy inference system

Both artificial neural network and fuzzy logic are used in ANFIS's architecture. ANFIS is consisted of if-then rules and couples of input–output, for ANFIS training is used learning algorithms of neural network (Avci et al., 2005a,b).

Adaptive network based fuzzy inference systems (ANFIS) are systems that are constructed to make use of some organizational principles resembling those of the human brain (Avci et al., 2005a,b). They represent the promising new generation of information processing systems. Adaptive network based fuzzy inference systems are good at tasks such as pattern matching and classification, function approximation, optimization and data clustering, while traditional computers, because of their architecture, are inefficient at these tasks, especially pattern-matching tasks (Avci and Turkoglu, 2003; Jang, 1993). As for DWANFIS try to combine aspects of the discrete wavelet transformation for purpose of feature extraction and selection with the characteristic decision capabilities of adaptive network based fuzzy inference system approaches (Zhang and Benveniste, 1992). The discrete wavelet adaptive network based fuzzy inference system (DWANFIS) is constructed based on the discrete wavelet transform theory (Thuillard, 2000; Turkoglu et al., 2003) and is an alternative to adaptive network based fuzzy inference systems (Guler and Ubeyli, 2004). Discrete wavelet transform (Burrus et al., 1998) is a powerful tool for non-stationary signal analysis. Let $x(t)$ be a piecewise continuous function. Discrete wavelet transform allows one to decompose