SPEAKER IDENTIFICATION USING WAVELET ANALYSIS AND ARTIFICIAL NEURAL NETWORKS

Anupam Shukla Ritu Tiwari Hemant Kumar Meena Rahul Kala

Department of Information Technology,
Indian Institute of Information Technology and Management, Gwalior, MP, INDIA

Citation: A. Shukla, R. Tiwari, H. K. Meena, R. Kala (2009) Speaker Identification using Wavelet Analysis and Artificial Neural Networks, *Journal of Acoustic Society of India*, 36(1), 20-25.

Abstract

Speaker identification is one of the most developing areas. The area attracts the attention due to its variety of uses in security, password protection, etc. With the development of technology, more and more people look forward for the use of speech analysis for identity of speakers. Speech processing techniques play a key role in this regard. The speaker identification refers to knowing the identity of the person where the identity is recognized by the words he speaks. A speaker is made to speak a few words. Later the speaker can be identified any time when he speaks these words.

In this paper we use wavelet analysis to extract useful characteristics of any author. These extracted features are trained using neural networks. In testing, we make the speaker speak out the same set of words. The features are extracted from this data and are fed into the training of the neural network. The neural network tells us the identity of the speaker. In order to test the system, we recorded the voice data of the speakers. The features were extracted and used in neural networks. A high performance of 94.7% clearly shows the working of the algorithm and its results.

KEYWORDS: Speech recognition, Wavelet Analysis, Artificial Neural Networks, Wavelet Transforms, Speaker identity

1. Introduction

Identification of a person is a very traditional problem and various tools and techniques have been used for the identification of people like figure print recognition, face recognition, signature recognition etc. Speech recognition is another such type of recognition technique that identifies an author based on the words spoken by him. In this paper we have used Wavelet Analysis to extract the various features of the speaker. This is an excellent means of analysis of such type of signals and is advancement over the Fourier analysis or Short Time Fourier Analysis (STFT) [6, 8]. We use Artificial Neural Networks to identify the speaker by the characteristics extracted.

This paper is organized as follows. Section 2 deals with the motivation of the paper. Section 3 discusses the concept of Wavelet analysis and Artificial Neural Networks. Section 4 talks about the procedure followed. The results are given in Section 5. Section 6 gives us conclusion remarks.

2 Motivation

Speaker identification is one of the most developing fields of today [2]. A lot of work has already been done in this field using various tools and techniques. A lot of work in this field exists in Hidden Markov Models. These are completely statistical models. These try to predict the expected value of output when the historic data is known. A lot of work also exists in Wavelet and other transforms [1, 3, 4, 5]. Here people have tried to identify the speaker by using transformations and analysis. Neural Networks [7] have been extensively used for the machine learning. They provide a convenient way to train the network and test it with high accuracy.

3 Techniques Used

In this section we would discuss the various techniques used in this paper [9, 10. 11].

3.1 Analysis Techniques

In this paper we have used Wavelet transform to extract characteristics. This analysis is advancement over Fourier analysis and Short Time Fourier Analysis (STFT). We discuss the properties and the advances if this transform in the subsequent sections.

3.1.1 Fourier Analysis

Fourier analysis breaks down a signal into constituent sinusoids of different frequencies. Another way to think of Fourier analysis is as a mathematical technique for transforming our view of the signal from time-based to frequency-based. Fourier analysis has a serious drawback. In transforming to the frequency domain, time information is lost. When looking at a Fourier transform of a signal, it is impossible to tell when a particular event took place. This transform is given in Fig. 1.

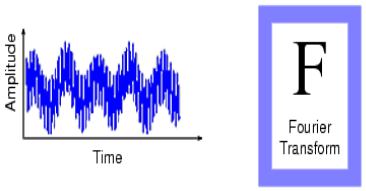

Fig.1: Fourier Transform

Fig.2: Short Time Fourier Transform

3.1.2 Short Time Fourier Analysis

Short Time Fourier Analysis is a technique called windowing the signal which maps a signal into a two-dimensional function of time and frequency. It uses Fourier transform to analyze only a small section of the signal at a time. While the STFT compromise between time and frequency information can be useful, the drawback is that once you choose a particular size for the time window, that window is the same for all frequencies. This transform is given in Fig. 2

3.1.3 Wavelet Analysis

The Wavelet Transform is a windowing technique with variable-sized regions. Wavelet analysis allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information. This is shown in figure 3.

Fig.3: Wavelet Transform

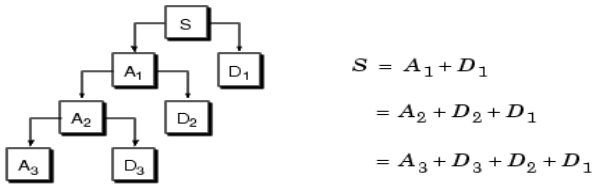

Fig.4: Wavelet Packet Analysis

Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques miss, aspects like trends, breakdown points, discontinuities in higher derivatives, and self-similarity. The wavelet packet method is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail, and the process is repeated. For n-level decomposition, there are n+1 possible ways to decompose or encode the signal. This is given in fig. 4.

3.2 Artificial Neural Networks

The artificial neural networks are an excellent means of machine learning. Learning in the context of artificial neural network involves the reputed training of the system to learn the given data. Testing in the neural networks is the step of giving unknown inputs to the neural network. These inputs test the performance of the neural network. A neural network must give desired answers to the inputs provided at this step.

4. Procedure

In this section we discuss the procedure followed for the experiment. The basic concept used is that we use Wavelet Analysis to extract features. These speakers were made to say a set of works one after the other. We applied Wavelet transform to each of the words spoken by the speaker. This gave us a series of features extracted. These series of values were extracted from the speech.

This formed a good database for the training of the neural network. The neural network was trained for each speaker. Then we tested the system. The input consisted of all the words spoken in order by the speaker. The output was collected and the speaker was identified. In the next sections we discuss the details of the Wavelet Analysis and the Artificial Neural Networks.

4.1 Wavelet Analysis

The wavelet analysis is used for analysis of the given input speech signal. Here the signal refers to a particular word spoken by a particular speaker. The Wavelet analysis consists of a detail and approximation. An approximation can be further broken down into detail and approximation. Hence we take a series of levels. A level here represents the degree of detail in the analysis. For this problem we fix the number of details to be 5. These are numbered as D_1, D_2, D_3, D_4, D_5 and A_5. Hence for every word spoken, we extract a total of six features.

4.2 Artificial Neural Networks

The purpose of the neural network is to learn the data for any speaker first. The learning is followed by the testing. We discuss the inputs and outputs in the next section.

4.2.1 Input

We know that every speaker speaks a total of 5 words. Every word is a collection of 6 features extracted. Hence there are a total of 30 features for every speaker. These form the 30 inputs for the neural network. These are numbered as I_1, I_2, I_3, I_4…. $I_{30.}$ In this paper we calculate the values of each of the input from the data extracted by the Wavelet Analysis. We have a huge amount of data for every input which is the training data. The general formula is given below. In this formula we have basically tried to normalize the inputs and get their ranges between 0 and 1. This can be well worked upon by the neural network. We subtract the mean and divide by the largest number in set for that particular input.

$I_i=(V_i - \text{Mean}(V_{ij})) \ / \ (\text{Max}(V_{ij}) - \text{Mean}(V_{ij})),$ for all j

Here I_i is th i^{th} input of the neural network

V_i is the ith feature extracted from Wavelet Analysis
Mean(Vi) is the mean of all Vij found in the training data set
Max(Vi) is the maximum of all Vij found in training data set for all j in data set

4.2.2 Output

In this problem we have created a classificatory output pattern. In this system there are as many numbers of outputs as are the number of classes. The outputs are numbered O_1, O_2, O_3…O_n where n is the number of classes. Each output O_i represents the chances of the output to be the i_{th} class. This varies from -1 to 1. The higher the output, the more will be the chances of the class to which the output belongs to be the class of the output. Hence in this way we need to find the maximum output and the class corresponding to it is the final class. For the training phase, the output of the class to which the input belongs is taken as 1, the output of the other classes are taken as -1. Hence if there are a total of 10 speakers, the output for the first speaker considered will always be <1,-1,-1,-1,-1,-1,-1,-1,-1,-1>.

5 Results

In order to prove the algorithm, we recorded data of 20 speakers. They were made to say 5 words each 'ab', 'is','baar','aap'and 'apne'. The speakers repeated these words and all of them were recorded. The Wavelet Analysis kit of Matlab was used for the Wavelet Analysis. The number of levels was specified to be 6. This gave us 6 different values found from wavelet analysis. The results of some the inputs of the wavelet analysis are given in Table 1. The neural network had a total of 1 hidden layer. There were a total of 48 neurons in the hidden layer.

S.No.	Spkr	Word	D1	D2	D3	D4	D5	A5	S.No.	Spkr	Word	D1	D2	D3	D4	D5	A5
1	A	1	1825	3642	7275	14542	29075	1825	**14**	C	4	1835	3661	7313	14617	29225	1835
2	A	2	1811	3614	7220	14432	28855	1811	**15**	C	5	1829	3649	7290	14572	29135	1829
3	A	3	1904	3799	7590	15172	30335	1904	**16**	D	1	1836	3664	7319	14629	29249	1836
4	A	4	1873	3737	7466	14924	29840	1873	**17**	D	2	1802	3596	7184	14359	28710	1802
5	A	5	1845	3682	7355	14702	29395	1845	**18**	D	3	1880	3752	7496	14984	29959	1880
6	B	1	1842	3675	7341	14673	29338	1842	**19**	D	4	1817	3626	7243	14477	28946	1817
7	B	2	1799	3590	7171	14334	28659	1799	**20**	D	5	1815	3622	7235	14461	28913	1815
8	B	3	1852	3696	7384	14760	29512	1852	**21**	E	1	1860	3712	7415	14822	29635	1860
9	B	4	1851	3694	7380	14752	29495	1851	**22**	E	2	1829	3650	7292	14576	29143	1829
10	B	5	1843	3678	7347	14685	29362	1843	**23**	E	3	1872	3736	7464	14920	29832	1872
11	C	1	1835	3662	7316	14623	29238	1835	**24**	E	4	1825	3642	7275	14542	29075	1825
12	C	2	1848	3687	7365	14722	29435	1848	**25**	E	5	1811	3614	7220	14432	28855	1811
13	C	3	1883	3757	7505	15001	29994	1883	**26**	F	1	1904	3799	7590	15172	30335	1904

Table1: the results of the Wavelet Analysis

When the algorithm was executed on the collected data, we got a performance of 97.5% using 20 speakers and 40 test cases (39 correctly identified). This clearly shows that the algorithm works well

and gives correct results on almost all inputs.

6. Conclusion

In this paper we proposed the use of Wavelet Transform and Artificial Neural Networks for the speaker identification. Wavelet transforms were used for the feature extraction. We extracted six features per word spoken by the author. These characteristics were given as a training data to the neural networks. The neural networks learnt the characteristics and reached the performance goal. When a new data was given as input to the neural network, it could make out the class to which it belongs. An acceptable performance level of 97.5% clearly show the fact that the algorithm can be used for the identification of the speaker.

In this paper we have proposed the use of Wavelet Analysis for the speaker identification. The same theory may also be used for the identification of words. Identification of author using multi-lingual words and the identification of speakers by giving any order of words needs to be done in future.

References

[1] Amar, C. Ben; Jemai, O, 'Wavelet Networks Approach for Image Compression', ICGST International Journal on Graphics, Vision and Image Processing, Vol. SI1, pp 37-45, May 2005

[2] Campbell, Jr. Joseph P; 'Speaker Recognition: A Tutorial.' Proceedings of the IEEE, Vol. 85, No. 9, pp. 1437-1462, 1997

[3] Long, C.J.; Datta S.; 'Wavelet Based Feature Extraction for Phoneme Recognition', Proc. of 4th Int. Conf. of Spoken Language Processing, pp 264-267, 1996

[4]Mallat, Stephane G., 'A Theory for Multiresolution Signal Decomposition: The Wavelet Representation', 674 IEEE Transactions On Pattern Analysis and Machine Intelligence.V OL. II, NO. 7. July 1989

[5] Mohanty, S.; Bhattacharya, S; 'Recognition of Voice signals for Oriya Language using wavelet Neural Network', ACM International Journal of Expert Systems with Applications, Vol 34, Issue 3, pp 2130-2147, April 2008

[6] Papakostas, G. A., Karras D. A., Mertzios B. G., and Boutalis, Y. S, 'An Efficient Feature Extraction Methodology for Computer Vision Applications using Wavelet Compressed Zernike Moments', ACM International Journal of Information Sciences, Vol 177, Issue 13, 2007

[7] Shukla, Anupam; Tiwari, Ritu, 'Fusion of Face and Speech Features with Artificial Neural Network for Speaker Authentication', IETE Technical Review, Vol 24, No 5, September-October 2007, pp 359-368

[8] Tang, Yuan-Liang; Hung, Chih-Jung, 'Recoverable Authentication of Wavelet-Transformed Images', ICGST International Journal on Graphics, Vision and Image Processing, Vol S11, pp 61-66, May 2005

[9] Torrence, Christopher;Compo, Gilbert P;.' A Practical Guide to Wavelet Analysis', Bulletin of the American Meteorological Society, Vol 79, pp 61-78, 1998

[10] Tzanetakis, George; Essl, Georg; Cook Perry, 'Audio Analysis using the Discrete Wavelet

Transform', In. Proc. WSES Int. Conf. Acoustics and Music: Theory and Applications (AMTA 2001) Skiathos, Greece, 2001
[11] Wassner, Hubert; Chollet, Geard; 'New Sepstral Representation using Wavelet Analysis and Spectral Transformation for robust speech recognition'