

Speaker Identification using Artificial Neural Network

Rahul Badgajar, Sayali Bagave, Shalvi Baswant, Vishal Mhasagar, Prof. Mrs. A. J. Nirmal

Department of Electronics
DattaMeghe College of Engineering
Airoli, Navi Mumbai-400708

Abstract-Nowadays various tools and techniques are available for identification of a person, which includes finger print recognition, signature recognition, face recognition etc. One of the types of such techniques is speaker identification in which person is identified by the words he speaks. With more inventions in the technology people look forward towards speech analysis to know the identity of speaker.

Here, we are using recorded voice signals as inputs. Wavelet analysis is used for extraction of features of a speaker from the input signal. These extracted features are applied to artificial neural network for training. Once the network is trained, testing is carried out. In testing, we ask random speaker to speak out same set of words. The features extracted using wavelet analysis are given for testing of the network. Then neural network gives the identity of the person. We expect that algorithm we used for speaker identification would give acceptable performance.

I. INTRODUCTION

Speaker identification is useful in various sectors such as security, password protection, etc. Because of such various uses, it attracts the attention. Here, we have used Wavelet Analysis to extract features from the words spoken by speaker. We make use of Artificial Neural Networks for identification of the speaker by the characteristics extracted.

In this project we make use of Artificial Neural Networks along with Wavelet Transform for the speaker identification. Wavelet transform is used for the feature extraction. For every word spoken by the speaker, we extract six features. The extracted features are fed as input to the neural network for training. The neural networks learn the features and reach the performance goal. When a new data is applied as input to the neural network, it can identify the class to which it belongs.

II. SPEECH PROCESSING

Speech is an immensely information-rich signal exploiting frequency-modulated, amplitude-modulated and time-modulated carriers (e.g. resonance movements, harmonics and noise, pitch intonation, power, duration) to convey information about words, speaker identity, accent, expression, style of speech, emotion and the state of health of the speaker. All this information is conveyed primarily within the traditional telephone bandwidth of 4 kHz. The speech energy above 4 kHz mostly conveys audio quality and sensation. Speech sounds are produced by air pressure

vibrations generated by pushing inhaled air from the lungs through the vibrating vocal cords and vocal tract and out from the lips and nose airways. The air is modulated and shaped by the vibrations of the glottal cords, the resonance of the vocal tract and nasal cavities, the position of the tongue and the openings and closings of the mouth. The vocal tract is the cavity between the vocal cords and the lips, and acts as a resonator that spectrally shapes the periodic input, much like the cavity of a musical wind instrument [1].

III. ANALYSIS TECHNIQUE

Wavelet Analysis:-

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.

Wavelet analysis is capable of revealing aspects of data 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 [2].

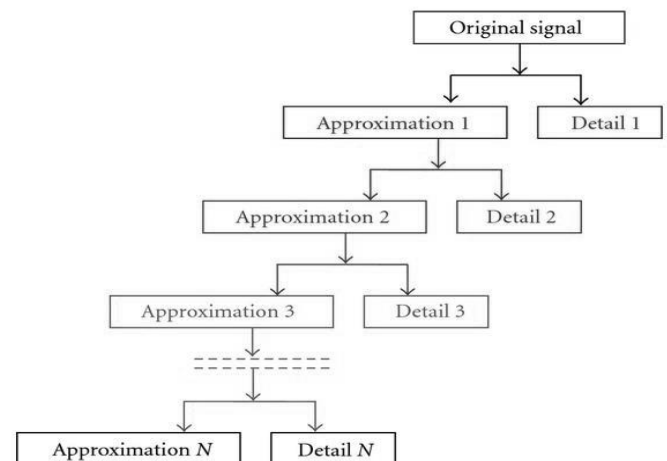


Fig. 1. Wavelet Decomposition Tree

A wavelet is a waveform of effectively limited duration that has an average value of zero. Compare wavelets with sine waves, which are the basis of Fourier analysis. Sinusoids do not have limited duration — they extend from minus to plus infinity. And where sinusoids are smooth and predictable, wavelets tend to be irregular and asymmetric.

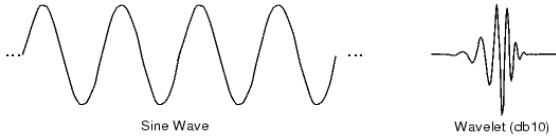


Fig. 2. Difference between sine wave and wavelet

Fourier analysis consists of breaking up a signal into sine waves of various frequencies. Similarly, wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet.

It also makes sense that local features can be described better with wavelets, which have local extent.

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower-resolution components as shown in fig. 1. This is called the wavelet decomposition tree [3].

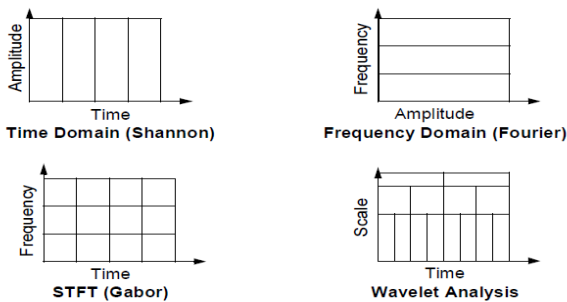


Fig. 3. Comparison of various transforms

IV. ARTIFICIAL NEURAL NETWORK

Since the invention of the digital computer, the human being has attempted to create machines which directly interact with the real world without his intervention. A neural network’s ability to perform computations is based on the hope that we can reproduce some of the flexibility and power of the human brain by artificial means.

Artificial neural network is a information processing system to model human brain. Artificial neural network contains the multiple layers of simple processing elements called neurons. It has large number of highly interconnected neurons which are the processing elements. Learning is accomplished by adjusting these strengths to cause the overall network to output appropriate result. Diagnostic systems, biochemical analysis, image analysis and drug development are the various areas where artificial neural network is used successfully [4].

The ANN’s are used in many important engineering and scientific applications, some of these are, signal enhancement, noise cancellation, pattern classification, system identification, prediction and control. Besides, they are used in many commercial products, such as modems,

image processing and recognition systems, speech recognition and bio-medical instrumentation, among others.

The artificial neural networks are an excellent means of machine learning. Learning with reference to artificial neural network involves the training of the system to learn the given data. In testing of neural networks, unknown inputs are given to the neural network. These unknown inputs test the performance of the neural network. A neural network should give desired output to the inputs provided for testing.

Types of learning:-

A. Supervised learning:-

In the supervised learning, the network is provided with input and output pairs. The training process continues until the network provides the expected response. The supervisor compares the output of the network with the expected one and determines the amount of the modification to be made in the weight. The objective is to decrease the difference between the answer of the network and the desired value.

B. Unsupervised learning:-

The artificial neural network with unsupervised learning does not require any external element to adjust the weight of the communication links to their neurons. Here the target output is not known for the given input. It is also called as Self learning network, since the process extracts the statistical properties of nodes and group in the similar network classes.

V. METHODOLOGY

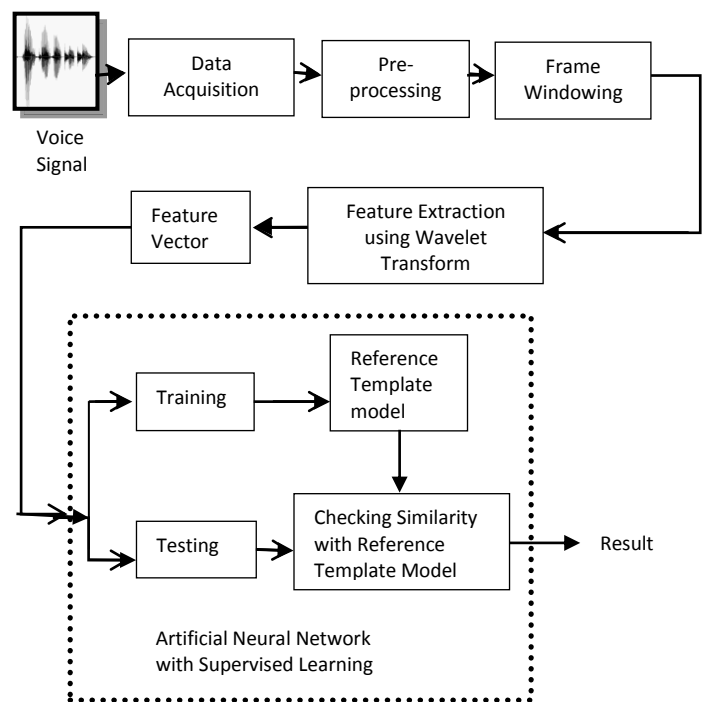


Fig. 4. Block Diagram

A. Data acquisition:-

The first stage consists of recording a speech file, from an esophageal speaker. The audio data is kept as a WAV file.

B. Preprocessing:-

The digital signal is low-pass filtered to reduce the background noise.

C. Wavelet Transform:-

The basic concept we use is Wavelet Analysis for extraction of features. These speakers are made to say a set of words one by one. Then Wavelet transform is applied to each word the speaker speaks. It provides a series of extracted features.

D. Artificial Neural Network:-

The features extracted using wavelet analysis forms a database for the training of the neural network. For each speaker, the neural network is trained and it undergoes testing. The input consists of all the words spoken by the speaker. The output is collected and the speaker is identified [5].

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 D1, D2, D3, D4, D5 and A5. Hence for every word spoken, we extract a total of six features.

The purpose of the neural network is to learn the data for any speaker first. The learning is followed by the testing [6].

VI. CONCLUSION

In this project, we make use of Artificial Neural Networks and Wavelet Transform for the speaker identification. The speaker identification refers to knowing the identity of the person where the identity is recognized by the words he speaks. We are going to use Wavelet transform for the feature extraction. These features will be applied as a training data to the neural networks. The neural networks learn the features and reference template model will be created. When a new data is given as input to the neural network, it can make out the class to which it belongs. Thus the speaker will be identified.

VII. ACKNOWLEDGEMENT

It is indeed a matter of great pleasure and proud privilege to be able to present this project. The completion of this project work is a milestone in student life and its execution is inevitable in the hands of guide. We are highly indebted the project-guide Prof. Mrs. A. J. Nirmal for her valuable guidance and appreciation for giving form and substance to this project.

We would like to tender our sincere thanks to the staff members for their co-operation. Really it is highly impossible to repay the debt of all the people who have directly or indirectly helped us for performing the project.

VIII. REFERENCES

- [1] PreetiRao, *Audio Signal Processing*, Chapter in *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, (Eds.) Bhanu Prasad and S. R. MahadevaPrasanna, Springer-Verlag 2007.
- [2] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, *Wavelet Toolbox for use with MATLAB*, 1996.
- [3] M. Sifuzzaman, M. R. Islam, M. Z. Ali, "Application of Wavelet Transform and its Advantages Compared to Fourier Transform", *Journal of Physical Sciences*, Vol. 13, 2009, 121-134.
- [4] Jure Zupan, "Introduction to Artificial Neural Network (ANN) Methods: What they are and how to use them", *Spain, Acta Chimica Slovenica* 41/3/1994, pp. 327-352.
- [5] Alfredo Victor Mantilla Caeiros and Hector Manuel Perez Meana, Esophageal Speech Enhancement Using a Feature Extraction Method Based on Wavelet Transform, *Mexico*.
- [6] AnupamShukla, RituTiwari, Hemant Kumar Meena, Rahul Kala, "Speaker Identification using Wavelet Analysis and Artificial Neural Networks", *Journal of Acoustic Society of India*, 36(1), 2009.