Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 95 (2016) 245 – 252

Procedia Computer Science

Complex Adaptive Systems, Publication 6
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2016 - Los Angeles, CA

Voice Identity Finder Using the Back Propagation Algorithm of an Artificial Neural Network

Roger Achkar*, Mustafa El-Halabi*, Elie Bassil*, Rayan Fakhro*, Marny Khalil*

*Department of Computer and Communications Engineering,
American University of Science and Technology, AUST
Beirut, Lebanon

Abstract

Voice recognition systems are used to distinguish different sorts of voices. However, recognizing a voice is not always successful due to the presence of different parameters. Hence, there is a need to create a set of estimation criteria and a learning process using Artificial Neural Network (ANN). The learning process performed using ANN allows the system to mimic how the brain learns to understand and differentiate among voices. The key to undergo this learning is to specify the free parameters that will be adapted through this process of simulation. Accordingly, this system will store the knowledge processed after performing the back propagation learning and will be able to identify the corresponding voices. The proposed learning allows the user to enter a number of different voices to the system through a microphone.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: Frequency; Neural Network; Backpropagation; Voice Recognition; Multi-layer Perceptrons.

1. Introduction

Voice recognition has been around for decades. However, it is surprising that only few people actually worked on it taking into account that it is very difficult for systems to succeed in understanding and differentiating human voices. It is known that voice recognition is a complex phenomenon. Voice is a dynamic process without clear distinguished

* Roger Achkar. Tel.: +9613997118; fax: +9611339302.
E-mail address: rachkar@aust.edu.lb

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2016.09.322

parameters [1]. Hence, distinguishing tones and vibration is harder than it sounds. For this reason, there is a need to build a system which will be capable of recording the voice of a person, listen to it, and eventually recognize it. In order for this to be achieved, lots of efforts and implementations are required [2].

Invented by Apple, Siri which is a computer program that works as an intelligent personal assistant and knowledge navigator was a hit in the field of voice recognition. Although helpful in answering questions and making recommendations, Siri recognizes the user based on the name provided while creating the Apple ID. Speech-to-text systems such as automated dialing and smart guidance systems for blind are also based on voice recognition but still lack machine intelligence and the ability to identify the user without referring to pre-provided written, registered or saved data of an account [3], [4], [5].

In machine learning, Artificial Neural Networks or ANNs are a family of models inspired by biological neural networks (the central nervous system, in particular the human brain) which are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. The accuracy of identification in neural networks is controllable because it is based on the number of samples fed to the system in the learning process. The higher the number of samples is, the more the accuracy of identification will be. Performing better than any other approaches, neural network is elected to tackle the problem of voice identification.

Hence, most common approaches to voice recognition are the fundamental frequency and the formant frequencies. The fundamental frequency is the lowest frequency of the voice signal, whereas the formant frequencies are fivefold; they are the resonances of the vocal tract system [6]. These basic parameters are to be considered as the experiential knowledge stimulated by the learning process of the neural network.

The main reason for using these approaches is to have an extreme accuracy in recognizing the voice. For this reason, the concept of voice recognition is initially used in order to allow the MATLAB software to understand the voice of the person [7]. The software itself will play the voice captured and will output the corresponding graph. These graphs help in calculating the five frequencies assigned to a specific person. Consequently, the program will save five frequencies for each input sentence on a text file; these frequencies represent the raw data of the learning process, so that the software only recognizes familiar voices. As a consequence, as soon as the person speaks any sentence to the computer through the microphone, the program will acquire the five different frequencies which are assigned to this person. The same person must enter different sentences with different voice levels for the learning to take as much examples as possible. Then, the program that was implemented in JAVA will take these input frequencies and perform the learning process. Therefore, no matter what condition a person may speak in, the program will be capable of identifying to whom each set of frequencies belongs.

2. Neural Network

The neural network used in this project is a Multilayer Perceptron (MLP). The MLP that is used in this system is made up of 6 input neurons, 1 hidden layer consisting of 4 neurons, and 2 output neurons.

Multilayer Perceptrons can be used in various applications that require classifications. Moreover, the MLPs are parallel structured networks that are fast in evaluating many examples and are robust on noisy training data [8]. In general, MLPs need non-linear function in order to undergo non-linear computations in order to perform the input-output mapping. The most common non-linear function used is known as the sigmoid activation function. The latter is used as teaching examples in MLPs. It makes computations easier than arbitrary activation functions. Moreover, it is used in MLPs to give logistic neurons real-valued output that is bounded function of their total input [9].

Accordingly, using the sigmoid function and the input-output mapping, the MLP performs the popular algorithm known as the error-correction learning rule. Basically, the aforementioned rule consists of two passes: forward pass and backward pass [10]. The backward pass or the back propagation is a learning procedure used in neural networks that repeatedly adjust the weights of the connections. This is done in order to minimize the measure of the difference between the actual output and the desired output [11].