IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-3, Mar 2019 E-ISSN: 2347-2693

Vowel Recognition of Speech using Data mining

Susheel Kumar Tiwari[1*], Manmohan Singh[2], Rahul Sharma[3]

[1]Department of Computer Science Engineering, Millennium Institute of Technology & Science, Bhopal
[2]Department of Computer Science Engineering, Chameli Devi Group of Institution, Indore
[3]Department of Computer Science Engineering, Chameli Devi Group of Institution, Indore

Corresponding Author: sushiltiwari24@yahoo.co.in

DOI: https://doi.org/10.26438/ijcse/v7i3.11641167 | Available online at: www.ijcseonline.org

Accepted: 18/Mar/2019, Published: 31/Mar/2019

Abstract— Over the past few years, technology has become very dynamic. It is fuelling itself at an ever increasing rate. An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. Over the last decades many interesting techniques of Neural Network (NN) were introduced, and shown to be useful in many applications in different fields. Since neural network brings together techniques from different fields such as vowel recognition, pattern recognition, Character recognition, face recognition, pattern matching, image processing, signature verification, data compression, signal processing among many different sources. This paper presents a study survey of various method of vowel recognition. The methods included and analyzed in this survey are Knowledge Based Cascade Correlation (KBCC), Multilayer Perceptron, Formants, and Linear predictive features.

Keywords—Vowel,speech,Speech recognition (SR), Knowledge Based Cascade Correlation (KBCC), Multilayer perceptron (MLP), linear predictive (LP)

I. INTRODUCTION

Speech recognition (SR) is the translation of spoken words into text. Analysis and presentation of the speech signal in the frequency domain are of the great importance in studying the nature of speech signal and its acoustic properties. The prominent part of speech signal spectrum belongs to formants that correspond to the vocal tract resonant frequencies. The quality of some of the most important systems for speech recognition and speech identification as well as systems for formant based speech synthesis are Speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition", "ASR", "computer speech recognition", "speech to text", or just "STT". Some SR systems use "training" where an individual speaker reads sections of text into the SR system. These systems analyze the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription. Systems that do not use training are called "Speaker Independent" systems. Systems that use training are called "Speaker Dependent" systems.

II. SURVEY OF VARIOUS METHOD

We study several methods for vowel recognition included methods are Knowledge Based Cascade Correlation (KBCC), Multilayer Perceptron, Formants, Linear predictive features.

A. DESCRIPTION OF KBCC

Knowledge based cascade correlation (KBCC) is an abstraction of CC. As in cascade correlation (CC), candidates are installed on top of the network, just below the output; hence new units receive inputs from every non-output unit already in the network. KBCC is not bounded to a pool of candidate units that are single-valued functions. KBCC can recruit any multivariate vector-valued component. The connection scheme in KBCC **as** shown below is similar to the CC connection scheme, except that a hidden unit may have a matrix of weight connections **(as** opposed to a single vector) at their inputs and their outputs as shown in figure 1.

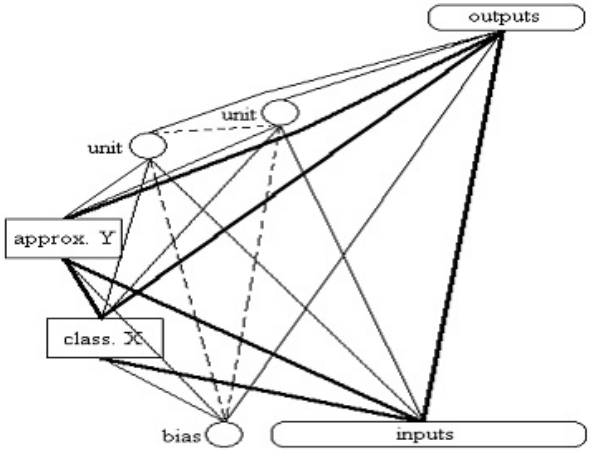

Figure 1: A KBCC network with four hidden units.

The first one is existing classifier X, the second one is existing approximator Y, and the two are single sigmoid units. A dashed line represents single weights, while a solid thin line represents weight vectors, and solid thick lines weight matrices [1].

To solve vowel recognition problem create six transfer scenarios from the CMU AI repository. Script was involved training networks based on the female data and then using them as sources to train target networks on male data. The other scripts is similar and completes all conversions of the three subsets [1]. The script is prepared using the following scheme. Starting with the three datasets (male, female, child), one dataset is used to train the source networks, and a different dataset for training the target networks. They produce six scripts. In order to compare KBCC with CC without knowledge, they add three more script where they trained CC nets on one of the datasets without any prior knowledge [1]. In this method KBCC is ready to receive and use its prior knowledge in the learning of a bulky and genuine new problem. In addition, the convenient of significant knowledge decrease KBCC learning time, without any loss of accuracy [1].

B. DESCRIPTION OF MULTI LAYER PERCEPTRON:

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. A multi-layer perceptron with one hidden layer was used to recognize the vowel sounds [3]. Multilayer perception with sigmoid nonlinearities and trained with the back propagation algorithm is appointed for the neural network. Multilayer perceptron contains several layers of neurons interconnected only between adjacent layers and no connection with in a layer [3]. The connection structure between layers for perception is shown in figure 2.

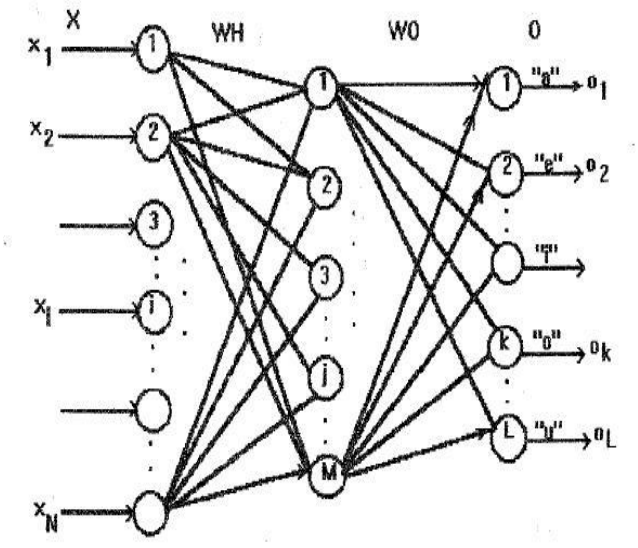

Figure 2: Connection structure for perceptron

In this method they formulate weight adjustment laws under the assumption that each entry "1" in the input layer will be computed as 0.98 and each "0" will be treated as 0.02.Neural network designed for classification required supervised training. The back propagation neural network is the most popular among many networks that satisfy the requirements for pattern classification. In this method performance test result created by putting input of the network a set of 100 vowels from each class, so 500 sounds and recognizes them [3].

A multilayer perceptron converges to an optimal classifier for vowel recognition. As parameters there are used the formant's frequencies of vocal sounds. Essential for precision recognition is using three formants as parameters of neural network. Binary coding is need for putting those numeric valued to the input of the network [3].

C. DESCRIPTION OF ISOLATED VOWEL RECOGNITION USING LINEAR PREDICTIVE FEATURES

Isolated vowel recognition is fundamentally a pattern recognition problem [4]. In this paper, isolated vowel recognition was studied with 3 different neural network classifiers, namely, the MLP, RBF and the PNN [4].

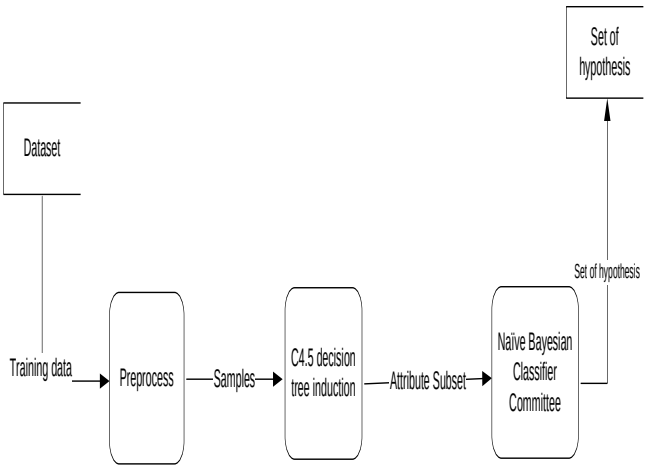

Figure 3

MLP is a three-layer network connected to each other by a matrix of weights and the layers are input layer, hidden layer, and output layer. The initial weights are assigned randomly. The output of each hidden layer node is the weighted sum of the input or feature vector passed through a nonlinear function. The nonlinear function that is commonly used is the sigmoid function [4].

The RBF network is a fully connected three-layered network and the layer which are used in RBF are input layer, a pattern layer, and output layer. The weights in the pattern layer are equal to the values of the training data. The weights in the output layer are initialized using small random values. The output of the pattern layer is computed by finding the distance between the input data and the pattern layer weights. It is then passed through a nonlinear Gaussian function. The outputs of the pattern layer are then

sent to the output layer and their weighted sum is computed [4].

In a PNN, the input layer is fully connected to the pattern layer, but the pattern layer is sparsely connected to the output layer. A PNN has a training algorithm that only requires one pass of the data through the network. The weights in the pattern layer are normalized versions of the input vectors. A pattern layer node is created for each data sample. It is connected to the output corresponding to its correct class. For classification, PNN computes the weighted sum of the pattern layer outputs. The outputs of the pattern layer are then passed through a nonlinear Gaussian function and stored according to the class that is associated with the particular node [4].

Various linear predictive features were examined (direct form predictor coefficients, reflection coefficients, log-area ratios and the cepstrum). Each linear predictive (LP) feature shows a different vowel recognition performance. Three different classifier fusion strategies (linear fusion, majority voting and weighted majority voting) were found to improve the performance [4].

D. DESCRIPTION OF RECOGNITION OF VOWELS IN CONTINUOUS SPEECH BY USING FORMANTS

Vowel Recognition through Formant Analysis in Serbian language, wherein they detect which of the five Serbian vowels is spoken by the Speaker [5].

Formants are the distinguishing or meaningful frequency components of human speech and of singing. By definition, the information that humans require to distinguish between vowels can be represented purely quantitatively by the frequency content of the vowel sounds. In speech, these are the characteristic partials that identify vowels to the listener. Most of these formants are produced by tube and chamber resonance, but a few whistle tones derive from periodic collapse of Venturi effect low-pressure zones. The formant with the lowest frequency is called f1, the second f2, and the third f3. Most often the two first formants, f1 and f2, are enough to disambiguate the vowel. These two formants determine the quality of vowels in terms of the open/close and front/back dimensions (which have traditionally, though not entirely accurately, been associated with the position of the tongue). Thus the first formant f1 has a higher frequency for an open vowel (such as [a]) and a lower frequency for a close vowel (such as [i] or [u]); and the second formant f2 has a higher frequency for a front vowel (such as [i]) and a lower frequency for a back vowel (such as [u]). Vowels will almost always have four or more distinguishable formants; sometimes there are more than six. However, the first two formants are most important in determining vowel quality, and this is often displayed in terms of a plot of the first formant against the second formant, though this is not sufficient to capture some aspects of vowel quality, such as rounding.

In this method we analyze that they use Linear Predictive Coding (LPC) method Linear Predictive Coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters [5].

In this paper [5] they presented a simple method for recognizing the five vowels of the Serbian language in continuous speech. The method they used is based on recognition of frequencies of first three formants that are present in vowels. By Using of LPC method for determining the frequencies and amplitudes of formants In speech, we have set the frequency ranges of formants F1, F2 and F3 for all vowels and defined the areas that vowels occupy in F1-F2-F3 space [5].

III. CONCLUSION

This paper presents survey on various methods of vowel recognition and the methods which we have used are Knowledge Based Cascade Correlation (KBCC), Multilayer Perceptron, Formants, and Linear predictive features. Accuracy and time of vowel recognition varies for different methods. We analyze that the Knowledge Based Cascade Correlation (KBCC) method are faster as compare to other methods.

REFERENCES

[1] Franqois Rivest and Thomas R. Shultz "Application of Knowledge-based Cascade-correlation to Vowel Recognition".

[2] Mihaela Grigore "Vowel recognition with non linear perceptron".

[3] Hult, G. "Some vowel recognition experiments using multilayer perceptrons"

[4] Jeff Byorick, Ravi P. Ramachandran and Robi Polikar "Isolated Vowel Recognition Using Linear Predictive Features and Neural Network Classifier Fusion"

[5] Biljana Prica and Sini˘sa Ili´c "Recognition of Vowels in Continuous Speech by UsingFormants"

[6] François Rivest and Thomas R. Shultz "Knowledge-based Cascade-correlation: A Review"

[7] Thomas R. Shultz and Francois Rivest "Knowledge-based Cascade-correlation: Varying the Size and Shape of Relevant Prior Knowledge"

[8] *Hua Nong TING and Jasmy YUNUS "*speaker independent malay vowel recognition of children using multi layer perceptron"

[9] Buckingham D, Shultz TR (2000) The developmental course of distance, time, and velocity concepts: A generative connectionist model. J Cog and Dev 1: 305–345

[10] Rivest F, Shultz TR (2002) Application of knowledge-based cascade-correlation to vowel recognition. IEEE Internat World Congr on Comp Intell, pp. 53–58

[11] Allison B (2007) The I of BCIs: next generation interfaces for brain–computer interface systems that adapt to individual users. In: International conference on human-computer interaction. Springer, Berlin, Heidelberg, pp 558–568

[12] Birbaumer N, Cohen LG (2007) Brain–computer interfaces: communication and restoration of movement in paralysis. J Physiol 579(3):621–636
[13] Faradji F, Ward RK, Birch GE (2009) Plausibility assessment of a 2-state self-paced mental task-based BCI using the no-control performance analysis. J Neurosci Methods 180(2):330–339
[14] Nijboer F, Sellers EW, Mellinger J, Jordan MA, Matuz T, Furdea A, Halder S, Mochty U, Krusienski DJ, Vaughan TM, Wolpaw JR (2008) A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol 119(8):1909–1916
[15] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. J Neural Eng. 2018;15: 056013.
[16] DaSalla CS, Kambara H, Sato M, Koike Y. Spatial filtering and single-trial classification of EEG during vowel speech imagery. i-CREATe 2009—International Convention on Rehabilitation Engineering and Assistive Technology. Association for Computing Machinery; 2009. pp. 1–4. doi: 10.1145/1592700.1592731
[17] Tzovara A, Murray MM, Plomp G, Herzog MH, Michel CM, De Lucia M. Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recognit. 2012;45: 2109–2122.
[18] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components—A tutorial. Neuroimage. 2011;56: 814–825.
[19] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. pp. 1725–1732.
[20] An X, Kuang D, Guo X, Zhao Y, He L. A deep learning method for classification of EEG data based on motor imagery. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag; 2014. pp. 203–210.

