



VOICE RECOGNITION FOR ENGLISH VOWELS USING ANN

¹Sagar Vasudev Fegade, ²Dr.Ashish Chaturvedi,

¹Research Scholar, ²Registrar ,

¹Department of CS and IT ,

¹Sabarmati University(Formerly Calorx Teacher's University) , Ahmedabad , India

Abstract : Artificial Neural Networks (ANNs) were developed in response to the neural networks found in the human brain and have found widespread application in speech processing. ANN applications include, among others, speech recognition, speech emotion recognition, language identification, speech enhancement, and speech separation. Similarly, given that human speech processing involves complex cognitive processes known as auditory attention, a growing number of papers have proposed ANNs supported by deep learning algorithms in conjunction with some mechanism to achieve symmetry with the human attention process.. This survey paper provides a comprehensive review of the various attention models used in the development of automatic speech recognition system for English vowels. This paper also describes a novel method for recognizing vowels. Lip features extracted via a combined method are fed into a neural network system for recognition.

IndexTerms - Voice recognition, artificial neural networks, Features extractions.

I. INTRODUCTION

Voice recognition of speakers by systems is that the problem of converting the knowledge content of the speech waveform of speakers into identifiable sets of features that carry all the possible discriminative information necessary for recognition of the speakers. the power of a recognition system to adequately recognize the voice of speakers essentially depends on the adequate capture of the time frequency and energy of the speech waveform and the way well the popularity model parameters are trained to supply the simplest sets of discriminations so on achieve accurate recognition[6]. With the arrival of technology, the thought of using the voice signals for the purpose of identification has found many useful applications in platforms like access control of data , access to banking services, secured database access system, remote access to telephone services, avionics, and automobile systems, etc[7],

Speech may be a natural mode of communication for people. We learn all the relevant skills during infancy , without instruction, and that we still believe speech throughout our lives. It comes so naturally to us that we don't realize how complex a phenomenon speech is. The human vocal tract and articulators are biological organs with nonlinear properties, whose operation isn't slightly below conscious control but also affected by factors starting from gender to upbringing to spirit . As a result, vocalizations can vary widely in terms of their accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed; moreover, during transmission[2,3], our irregular speech patterns can be further distorted by ground noise and echoes, also as electrical characteristics (if telephones or other equipment are used). of these sources of variability make speech recognition, even quite speech generation, a really complex problem. We review a number of the synthetic Neural Network (ANN) approaches utilized in speech recognition. Some basic principles of neural networks are briefly described also as their current applications and performances in speech recognition. Strengths and weaknesses of pure connectionist networks within the particular context of the speech signal are then evoked

The sub problems can be classified according to[11]

- isolated words - continuous speech
- vocabulary: small (< 50 words), medium (50 - 500 words), large (> 500 words)
- single speaker - multiple speakers
- noise: noise level, type of noise

A answer to the boundaries of 'hand-engineered' facets has emerged via mimicking features of organic neurons in synthetic neural networks (ANN). The workable of ANNs is currently being exploited with get admission to to massive trainable datasets, environment friendly gaining knowledge of algorithms, and effective computational resources. Few of these developments in ANN over the final decade have led to deep studying [16,17] that, in turn, has revolutionized various utility domains consisting of pc vision, speech analysis, biomedical picture processing, and on line market analyses. The fast success of deep studying over usual computing device getting to know may additionally be attributed to three factors. First, deep getting to know provides end-to-end trainable architectures that combine function extraction, dimensionality reduction, and last classification.

I. NEURAL NETWORKS (NN)

In 1980, Neural Networks (NN) were developed as an acoustic modelling method in the Automatic Speech Recognition System (ASR)[9]. Many features of Speech Recognition (SP), such as phoneme classification, have made use of Neural Networks (NN). The Neural Networks (NN) method is very useful in Speech Recognition Systems. The neural network method is used in the majority of research work. Isolated Word Speech Recognition, audiovisual speaker recognition, and speaker versioning are all used to classify phonemes.

Because speech is not stationary, Speech Recognition is a task-oriented domain. Various techniques are used depending on the task. Techniques such as MFCC, DTW[9], and others are commonly used for small datasets. Neural Network-based techniques are preferred for larger and more standardised corpora.

However, MFCC and DTW are popular, simple, and robust techniques. These techniques are used depending on the goal of the proposed system, the type of speech considered, and the size of the database.

II. VOWEL RECOGNITION

Each language's basic elements are vowels and consonants. Furthermore, the difference in pronunciation of a word uttered by people speaking different mother tongues is primarily due to differences in vowels and how they are pronounced[12][13]. Farsi (Persian mother tongue) has six distinct vowels denoted as , , ا, آ, ای, او which are very similar to the English vowels a ,e ,o ,â ,i and u respectively. In this section a method is presented for vowel recognition using the lip features extracted in the former sections and applying them to an appropriate neural network. The developed method is then used to recognize the 6 major Farsi vowels. To create a suitable database, 8 people (2 males and 6 females) were asked to utter 42 monosyllabic Persian words, each of which contained a single vowel. The utterances' scenes were then recorded at the standard rate of 30 frames per second.

Non-native speakers of a language's speech are usually distinguished by acoustic phonetic and phonological deviations from the norm. This field of study has attempted to identify these deviations in order to characterise various accents. For many years, the English language[22] has been studied from the standpoint of both native and non-native accents. The speech of non-native speakers may have several acoustic-phonetic characteristics that are not found in the speech of native speakers. The purpose of this paper[22] is to investigate the impact of Bengali accent on English vowel recognition. The spectral characteristics of various English vowel sounds are seen to be heavily influenced by Bengali-accented speech. The nature of each vowel sound in Bengali accent differs from that of native accents. These findings are extremely useful in developing an English vowel recognizer that can successfully work with Bengali-accented people. The recognition method used here is very simple, and the recognition accuracy is also very good.

This field of study has endeavored to recognize these deviations to portray different accents. For a long time, the English language[22] has been concentrated from the outlook of both local and non-local accents. The discourse of non-local speakers might have a few acoustic-phonetic attributes that are not found in the discourse of local speakers. The motivation behind this paper[22] is to research the effect of Bengali inflection on English vowel acknowledgment. The phantom qualities of different English vowel sounds are believed to be intensely affected by Bengali-emphasized discourse. The idea of every vowel sound in Bengali pronunciation contrasts from that of local accents. These discoveries are very valuable in fostering an English vowel recognizer that can effectively work with Bengali-emphasized individuals. The acknowledgment strategy utilized here is extremely basic, and the acknowledgment precision is additionally generally excellent.

III. FEATURE EXTRACTION AND METHOD

Every 10 ms, the feature extraction produces an acoustic vector that represents the salient speech feature of a 30 ms window. The cepstrum coefficients, the delta-cepstrum coefficients, i.e. the estimation of their temporal derivative, the delta-energy, and the delta-delta energy are popular features. The vector quantizer detects clusters in the set of acoustic vectors and determines a consultant vector for every cluster. This vector is coded, and the recognizer is fed a string of codebook vectors. The facts drift is significantly reduced as a result of the incoming speech sign. The vector quantization step can be skipped if enough time and computing power are available. Vector quantization is normally carried out with the aid of the classical K-means algorithm, however the **Kohonen neural community (Kohonen, 1992)** has additionally proved to be at least as efficient, however at a lot decrease computational fee for the training.

The most extensively used recognizers are primarily based on hidden Markov models. It is assumed that the speech utterances are produced through a **Markov process**, whose states are no longer without delay observable. For a given Markov mannequin of, say, a word, the probability that the pronunciation of the phrase produces a sure speech sign can be determined. Conversely, for a given speech utterance that represents a word[11], the probability that the utterance has been produced when announcing a sure word, can be calculated in the equal way. Therefore, specific hypotheses of phrases can be tested and the most possibly can be chosen. This is the attention technique primarily based on hidden Markov models. Choosing a non zero likelihood that the Markov process remains in the identical country at the subsequent time instant, the time variability of speech can efficiently be modelled. The major disadvantage of hidden Markov fashions is their relatively modest discriminative electricity for classification.

IV. CONCLUSION

Speech may be a natural mode of communication for people. We learn all the relevant skills during infancy , without instruction, and that we still believe speech throughout our lives. It comes so naturally to us that we don't realize how complex a phenomenon speech is. Artificial Neural Networks (ANNs) were developed in response to the neural networks found in the human brain and

have found widespread application in speech processing. ANN applications include, among others, speech recognition, speech emotion recognition, language identification, speech enhancement, and speech separation.

REFERENCE

- [1] Siddharth S. More, Prashantkumar L. Borde, Sunil S. Nimbhore, "A Review on Automatic Speech Recognition System in Indian Regional Languages", IJOCA (International journal of computer applications) vol. 181, No.4, July 2018
- [2] B.P Yuhas.et.al., Integration of acoustic and visual speech signals using neural networks. ,IEEE Communication Magazine November 1989
- [3] M. Barnard , E. J. Holden, R. Owens, " Lip tracking using pattern matching snakes," ,5th Conf. on Computer vision, pp.1-6, 2002
- [4] A.Waibel. Modular construction of time-delay neural networks for speech recognition, Neural Computation, 1,39-46 1989
- [5] S.Nakamura, Statistical Multimodal Integration for Audio-Visual Speech Processing. IEEE transaction on neural networks, 13(4), JULY 2002.
- [6] Rabiner, L. R., "Applications of speech recognition in the area of telecommunication", IEEE Proc., 1997, pp. 501-510.
- [7] Picone, J. W., "Signal modeling techniques in speech recognition," IEEE Proc., Vol. 81, No. 9, 1993, pp. 1215-1247.
- [8] Campbell, J. P. Jr., "Speech recognition: A tutorial", IEEE, Vol. 85, No. 9, 1997, pp. 1437-1462
- [9] Rajashri G. Kanke, Maheshwari A. Ambewadikar, Manasi R. Baheti, REVIEW ON SMALL VOCABULARY AUTOMATIC SPEECH RECOGNITION SYSTEM (ASR) FOR MARATHI
- [10] Vahideh Sadat Sadeghi, Vowel Recognition using Neural Networks, JAN 2006, research gate.
- [11] Jean Hennebert, NEURAL NETWORKS IN SPEECH RECOGNITION, January 1994
- [12] Dzmityr Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, Yoshua Bengio, End to end attention based large vocabulary speech recognition , www.arXiv:1508.04395 ,14 mar 2016
- [13] Mr. Himanshu N Patel, Dr. P.V. Virparia, "A Small Vocabulary Speech Recognition for Gujarati", International Journal of Advanced Research in Computer Science(IJOARICS), volume 2, No.1, www.ijarcs.info ,Jan-Feb 2011
- [14] Suhas R. Mache, Manasi R. Baheti ,C. Namrata Mahender, "Review on Text To Speech Synthesizer", JARCCCE(2015), Vol. 4, Issue 8, August 2015
- [15] Manasi Ram Baheti, Bharti W. Gawali, S.C. Mehrotra, "Automatic Speech Recognition For Task Oriented IVRS In Marathi", IOSR Journal Of Computer Engineering (IOSR JCE), www.Iosrjournals.Org, Mar-Apr. 2016
- [16] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, vol. 18, no. 7, pp. 1527-1554, 2006.
- [17] G. E. Hinton, "Learning multiple layers of representation," Trends in cognitive sciences, vol. 11, no. 10, pp. 428-434, 2007.
- [18] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, and A. Oliva, "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence," Scientific reports, vol. 6, pp. 1-13, 2016, Art no. 27755.
- [19] Waibel and K-F. Lee, editors. "Readings in Speech Recognition", Vol. 1, Morgan Kaufman Publisher, Inc., San Mateo, California, 1990
- [20] Weisfreid, M, V. Wickerhauser, "Signal Processing via Fast Malvar Wavelet Transform Algorithm", GRETSI, 1993, pp. 379-382.
- [21] Z.Zhao and C. G. Rowden, "Use of Kohonen Self-Organising Feature Maps for HMM Parameter Smoothing in Speech Recognition", IEE Proceedings-F, Vol. 139, No. 6, December 1992
- [22] Tanjin Taher Toma; Abu Hasnat Md. Rubaiyat; A. H. M Asadul Huq, Recognition of English vowels in isolated speech using characteristics of Bengali accent, 2013 2nd International Conference on Advances in Electrical Engineering (ICAEE), IEEE.
- [23] T.B.Martin, A.L.Nelson, and H.J.Zadell, Speech Recognition b Feature Abstraction Techniques , Tech.Report AL-TDR-64-176,Air Force Avionics Lab,1964.
- [24] M. Kirby and L. Sirovich, "Application of the Karhunen- Loeve procedure for the characterisation of human faces," IEEE Trans.
- [25] Pattern Analysis and Machine Intelligence, vol. 12, pp. 831-835, Dec.1990.
- [26] M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71-86, 1991.