

AN CRITICAL ANALYSIS OF SPEECH RECOGNITION OF TAMIL AND MALAY LANGUAGE THROUGH ARTIFICIAL NEURAL NETWORK

Kingston Pal Thamburaj ¹, Kartheiges ponniah ² and Ilankumaran Sivanathan³ MuniiswaranKumar ⁴

¹ Universiti Pendidikan sultan Idris 1; fkingston@gmail.com

² Universiti Pendidikan sultan Idris 2; kartheiges@fbk.upsi.edu.my

³ Universiti Pendidikan sultan Idris 3; ilankumaran@fbk.upsi.edu.my

⁴ Universiti Pendidikan sultan Idris 4; muniisvaran@fbk.upsi.edu.my

* Correspondence: kartheiges@fbk.upsi.edu.my; Tel.: (+6 0162691801)

Abstract

Human and Computer interaction has been a part of our day-to-day life. Speech is one of the essential and comfortable ways of interacting through devices as well as a human being. The device, particularly smartphones have multiple sensors in camera and microphone, etc. speech recognition is the process of converting the acoustic signal to a smartphone as a set of words. The efficient performance of the speech recognition system highly enhances the interaction between humans and machines by making the latter more receptive to user needs. The recognized words can be applied for many applications such as Commands & Control, Data entry, and Document preparation. This research paper highlights speech recognition through ANN (Artificial Neural Network). Also, a hybrid model is proposed for audio-visual speech recognition of the Tamil and Malay language through SOM (Self-organizing map) and MLP (Multilayer Perceptron). The Effectiveness of the different models of NN (Neural Network) utilized in speech recognition will be examined.

Keywords: ANN, NN, Speech Recognition, interaction, hybrid method

1. Introduction

Speech recognition (SR) is considered as the process used for converting the acoustic signal which was captured using a telephone or microphone to a set of words. The words recognized using this process can be further used for data entry, Document preparation, and as commands. The speech recognition can be categorized based on the parameters they are

1. Speech representation
2. Modeling classification
3. Lexical models
4. Language models
5. Training data
6. Acoustic models

The speech recognition process is grouped as

1. A single word recognition system
2. Continuous speech recognition

Spontaneous speech is sometimes very difficult to recognize. One of the common problems faced in speech recognition is that understanding the vocabulary size of the combined word. For this perplexity, the model is applied which recognizes the number of words that can flow in a language model. They are few other parameters that can affect the speech recognition model they are

1. External noise such as sound in the background of the person
2. Microphone placement of the speaker.

The phonetic variables can be demonstrated by acoustic differences using certain words they are true, butter, it, into. The proper SR model will 1st try to model the source of the variable in many ways. In signal representation, many researchers have developed speaker independent features for signaling, and to analyze the speaker-dependent characters. At an acoustic-phonetic level, the variability of the speaker will be modeled through a statistical method applied in the large data model. The word-level variable will allow alternate pronunciation of words through a pronunciation network.

The HMM (Hidden Markov Model) is the predominant model used in SR for the past 15 years. In HMM the generation of the frame-by-frame word, the surface acoustic realization is the 2 represented proves in probabilistically as Markov process.

NN (neural network) is the alternate process used for estimation frame-by-frame word score after obtaining the source they are combined with HMM model this process is also referred to us as a hybrid model.

(Nilsson & Ejnarsson, 2002) One of the important aspects of SR is to assist people with fractional disability. This would help them in their daily activity. By speech, they could control all electronic appliances such as fans, lights, machines, etc which they use for their domestic purpose. The architecture of the audiovisual speech recognition engine is shown in Figure 1.

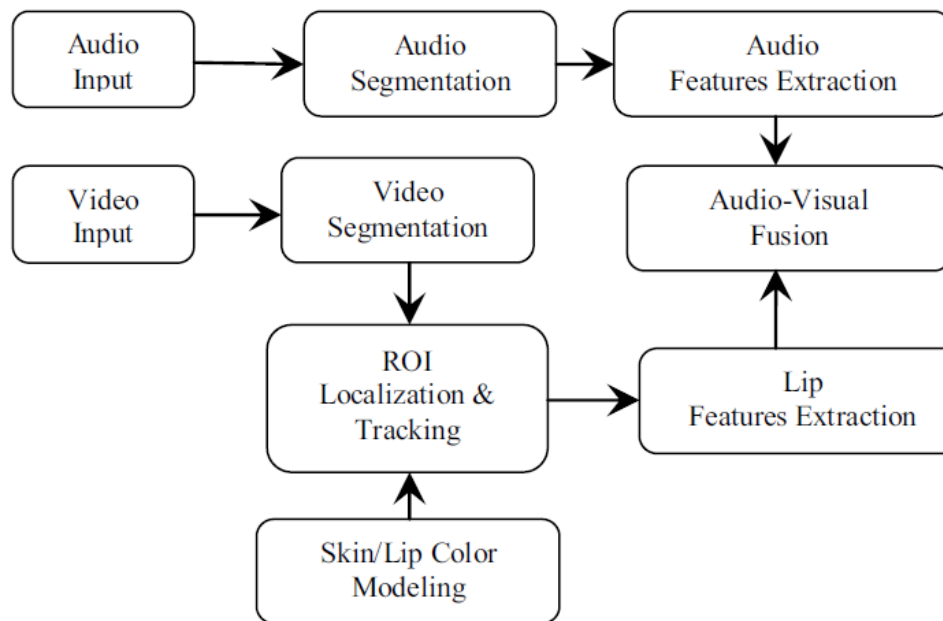


Figure 1: The basic block diagram of an audio-visual automatic speech recognizer

2. LITERATURE REVIEW:

(Srinivasan, 2011, Pal Thamburaj , K., & Sivanathan,2020) in the year 1950 1st attempt was made for automatic SR using a machine. (Kumar et al., 2020) in the year 1952 isolated digit recognition for a single speaker was introduced by Davis Biddulph and Balashek in Bell Laboratories. This model was based on measuring the spectral resonance for the vowel region of each word.

(Muhammad et al., 2018) in the year 1956 Olson and Belar tried to analyze SR by introducing 10 distinct syllables embodied along with 10 monosyllable words in RCA Laboratories. This model was based on spectral measurement in the vowel region.

(El Ouahabi et al., 2020) in the year 1959, Fry and denes introduced a phoneme recognizer for recognizing 4 vowels and 9 consonants. This model utilized a spectrum analyzer and pattern match for making recognition decisions. This model was based on statistical information on English Phonemes.

(Yavuz & Topuz, 2018) in the year 1959 Forgie and Forgie introduced a recognizer with 10 vowels embedded with a/b/-vowel-/t/ format.

(Rynjah et al., 2020) in the year 1960 Natkata and Suzuki introduced a hardware-based vowel recognizer. (Techini et al., 2017) in the year 1962 introduced a phoneme recognizer. (Shi et al., 2018) introduced a digital SR in 1963. (YUSOF, 2019) proposed a framework using time aligning with a pair of speech utterances.

(Kumar et al., 2020; Muhammad et al., 2018, K. Pal Thamburaj et al.,2015) many researchers have tried various approaches for speech recognition using ANN. This method was the most emerging method for SR and its classification. It is an information processing method. shown in fig 2 as follows.

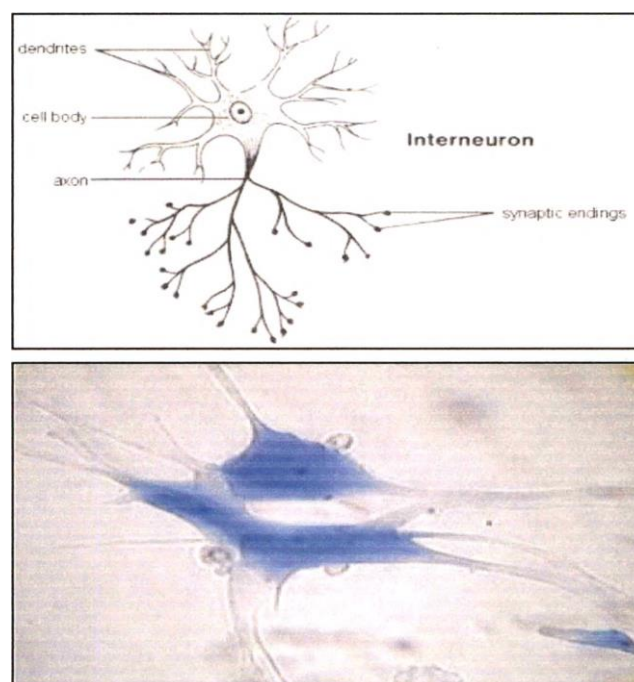


Fig: 2: Biological Neuron

(Dong & Li, 2020) the main disadvantage in SR is due to its ability to adaption for learning, ability to generalize, and non-linearity.

(Ali et al., 2020) the MLP (Multilayer Perceptron) is another popular method for SR in NN architecture. The basic model of MLP is illustrated in fig 3 as follows. MLP is the supervised learning method that adapts its value based on the training pattern.

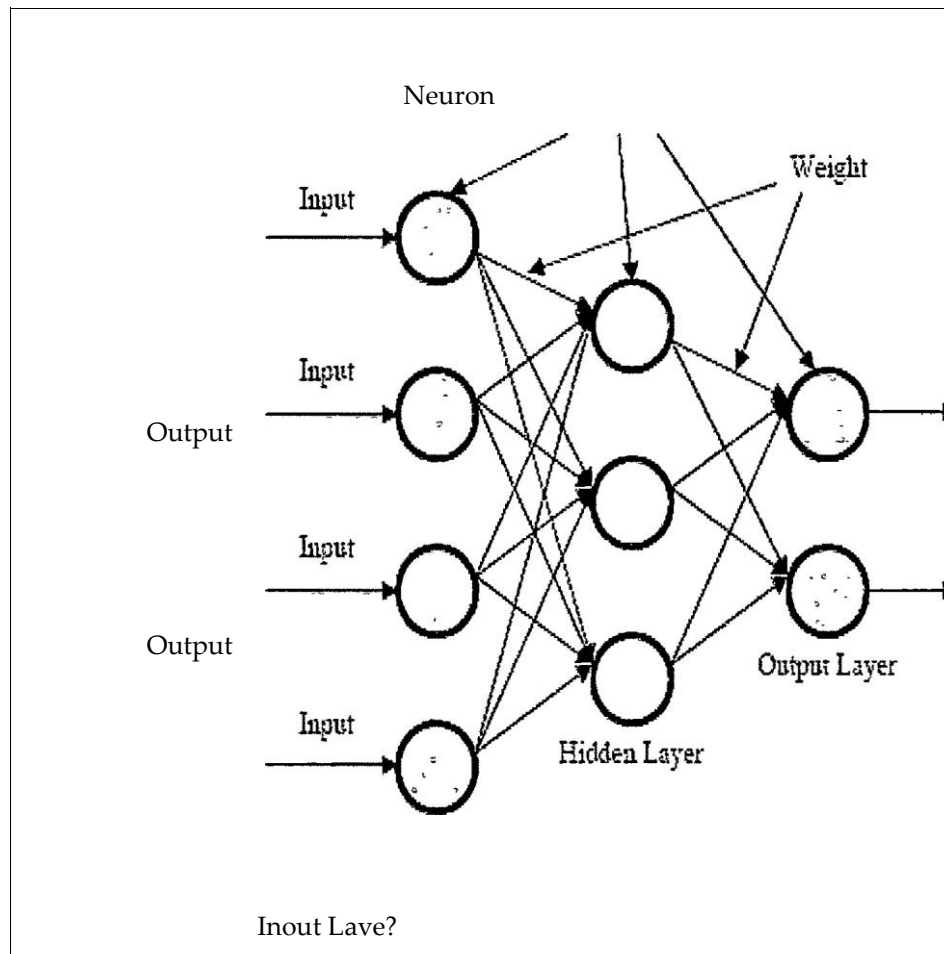


Fig 3: A basic architecture of Multilayer Perceptron (MLP)

3. Review of the approached method

ANN is a powerful computational device that can be used for massive parallelism which makes the system a very effective one. This model can learn and generalize for the given training data. It is very tolerant to a fault. Also adapts itself for noise. They can perform any kind of operation such as logic as well as symbolic. There are many types of ANN. Most of the ANN operators with a neural device with connected neuron are the Self Organizing Map (SOM), Multi-Layer Perceptron (MLP), and the Hopfield network. The main motive of this ANN network is to analyze the link between the input and output pattern. This process is achieved by modifying the link weight between the units.

4. Overview of selected language for study

4.1 Overview of Tamil Language

The Tamil language is one of the members of the Dravidian language which is predominantly spoken in the southern part of India. It is the official language of state Tamil Nadu and Puducherry located in the southern part. The Tamil Language is also considered the official language for countries like Singapore and Srilanka. Also, many Tamil language speaking people are found in South Africa, Malaysia, Fiji, and Mauritius. It is declared as the classical language of India in 2004. This is due to 3 important reasons they are it is one of the ancient languages, it has an independent tradition, and it comprises of ancient literature. It was found that almost 65 million people are speaking the Tamil language in the 21st century.

Tamil language writing was found in inscriptions and potsherds from the 5th century BC. The Tamil language has 3 periods based on the lexical and grammatical change they are

1. The Old Tamil from 450-700 BC.
2. The Middle Tamil from 700-1600 BC
3. The Modern Tamil from 1600 BC

The Tamil Language writing style evolved from Brahmi Scripts. The shape of Tamil letters changed over time. The main addition of letters for the Tamil language was incorporated from Grantha letters. Spoken Tamil also changed over time especially during the phonological structural change of words. Within Tamil Nadu, the phonological difference is found between different districts located in the north, south, west, and east of Tamil Nadu.

4.2 Overview of Malay Language

The Malay language is one of the members of the Austronesian language family. This language is largely spoken by 33M people of Sumatra, Borneo, and the Malay Peninsula. It is widely spoken by people of Malaysia and Indonesia. The Malay language shows major resemble Sumatra but it is related to other Austronesian languages such as Java, Borneo, and Cham Language of Vietnam.

The Malay language is the official language of the Republic of Indonesia, and Bahasa. Many version of Malay language is found they are the Bazaar Malay. Baba Malay, Kutai Malay, Banjarase, and Brunei Malay.

Affixes are demonstrated in constructions such as *di-bēli* “be bought” and *mēm-bēli* “buy” from the root form *beli* “buy!” and *kemauan* “desire” from *mau* “want.” Doubling may be used to mark the plural—for example, *rumah* “house” and *rumah-rumah* “houses”—or to form derivative meanings, as in *kekuningkuningan* “tinted yellow” from *kuning* “yellow” and *bērlari-lari* “run around, keep running” from *bērlari* “run.”

5. Breif description of phonemes of Tamil and Malay Languages :

The basic theoretical unit for describing how to bring linguistic meaning to the formed speech, in the mind, is called phonemes

5.1 PHONEMES OF Tamil LANGUAGE:

The principal Tamil phonemes can be represented in the chart form as given below:

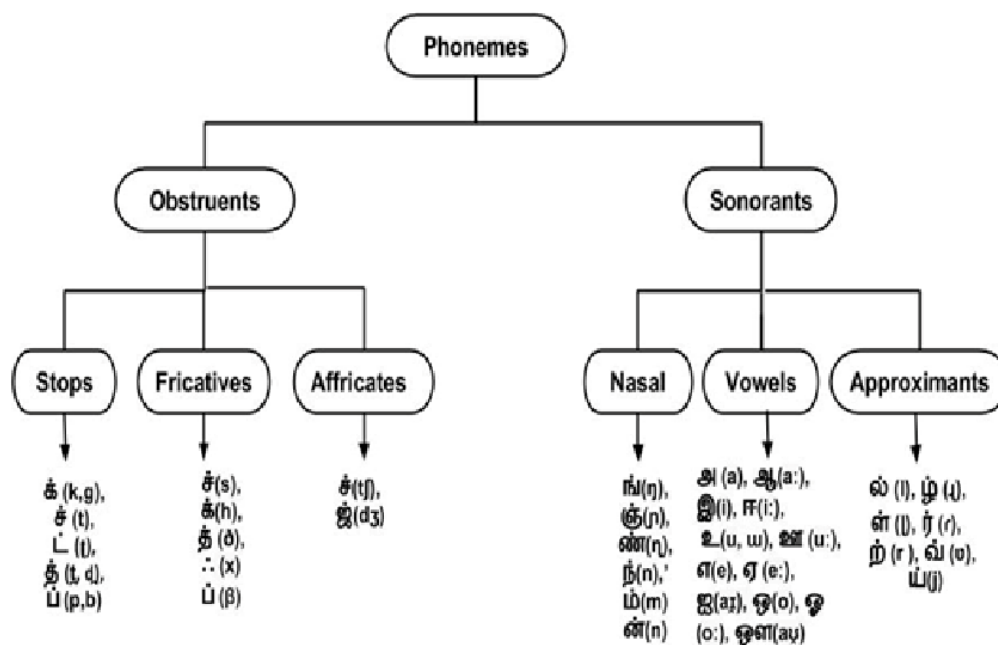


Fig 4 Tamil Phonology Chart

You can learn to write your name in Tamil

The Tamil Language originated with the Indus Valley Civilization over 5000 years ago! Canada is home to over 300,000 Tamils. It is the language of over 70 million people living mainly in South India, North and East of Sri Lanka, and around the world. The Tamil Alphabet has 247 letters (12 vowels + 18 consonants + 216 (=12 x 18) combinations + 1 special character)

Consonants	Vowels											
	1	2	3	4	5	6	7	8	9	10	11	12
	A	AA	I	EE	U	UU	E	AE	AI	O	OA	AU
	akk	ca	ca	ki	ke	pu	bu	ge	ga	pe	pa	cu
1 B, P: as in nIP	ப	பா	பி	பே	பு	பு	பெ	பை	பொ	பு	பொ	பு
2 C, S, J: as in catCH	ச	சா	சி	சே	சு	சு	செ	சை	சொ	சு	சொ	சு
3 GNA: as in puNCH	ஞ	ஞா	ஞி	ஞே	ஞு	ஞு	ஞெ	ஞை	ஞொ	ஞு	ஞொ	ஞு
4 D, T: as in daD	ட	டா	டி	டே	டு	டு	டெ	டை	டொ	டு	டொ	டு
5 G, H, K: as in cloCK	க	கா	கி	கே	கு	கு	கெ	கை	கொ	கு	கொ	கு
6 L: as in fILL	ல	லா	லி	லே	லு	லு	லெ	லை	லொ	லு	லொ	லு
7 L: as in buLL	ள	ளா	ளி	ளே	ளு	ளு	ளெ	ளை	ளொ	ளு	ளொ	ளு
8 L: as in caRL	ர	ரா	ரி	ரே	ரு	ரு	ரெ	ரை	ரொ	ரு	ரொ	ரு
9 M: as in jaM	ம	மா	மி	மே	மு	மு	மெ	மை	மொ	மு	மொ	மு
10 N: as in peN	ந	நா	நி	நே	நு	நு	நெ	நை	நொ	நு	நொ	நு
11 N: as in puN	ன	னா	னி	னே	னு	னு	னெ	னை	னொ	னு	னொ	னு
12 NTH: as in paNTHer	ந்	நா	நி	நே	நு	நு	நெ	நை	நொ	நு	நொ	நு
13 NG: as in baNG	ங	ஙா	ஙி	ஙே	ஙு	ஙு	ஙெ	ஙை	ஙொ	ஙு	ஙொ	ஙு
14 R: as in puRR	ர்	ரா	ரி	ரே	ரு	ரு	ரெ	ரை	ரொ	ரு	ரொ	ரு
15 R: as in KITE = tRee	ற்	ரா	ரி	ரே	ரு	ரு	ரெ	ரை	ரொ	ரு	ரொ	ரு
16 Th: as in paTH	த்	தா	தி	தே	து	து	தெ	தை	தொ	து	தொ	து
17 V, W: as in giVE	வ	வா	வி	வே	வு	வு	வே	வை	வொ	வு	வொ	வு
18 Y: as in baYl or YArD	ய	யா	யி	யே	யு	யு	யெ	யை	யொ	யு	யொ	யு

Srirangam, 2006

Websites:
www.tamilnation.org
www.tamilcanadian.com
www.tamilnet.com

Fig 5 Tamil Phonemes

Test Words	Phonetic Transcription	No of Syllable	Words Classification
Amma	[əməmə]	2	Open Nasal
Appa	[əbəbə]	2	Open Plosives
Bhandhu	[bɦəndɦu]	2	Close Nasal Plosives
Kan	[kənə]	1	Open Nasal Plosives
Kathavu	[kəɖəɦu]	3	Open Plosives
Kathu	[kəɖɦu]	2	Close Plosives
Mani	[məni]	2	Open Nasal
Maram	[məməmə]	2	Open Nasal
Veedu	[ve:ɛ:du]	2	Open Plosives
Paal	[paələ]	1	Close Plosives

Fig 6 Word Classification of Tamil Language

Examples of Tamil Vowels

Vowels	Special Character
அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ	ஃ
Consonants	
க ங ச ஞ ட ண த ந ப ம ய ர ல வ ழ ள ற ன்	
Vowel-Consonants	
க ங ச ஞ ட ண த ந ப ம ய ர ல வ ழ ள ற ன்	

Fig 7 Tamil Language Vowels

5.2 Phonemes of Malay language

The principal Malay phonemes can be represented in the chart form as given below:

Malay consonant phonemes						
		Labial	Alveolar	Palatal	Velar	Glottal
Nasal		m	n	ɲ	ŋ	
Plosive/Affricate	voiceless	p	t	tʃ	k	(ʔ)
	voiced	b	d	dʒ	g	
Fricative	voiceless	(f)	s	(ʃ)	(x)	h
	voiced	(v)	(z)			
Approximant	central			j	w	
	lateral		l			
Trill			r			

Fig 8 Malay Phonemes

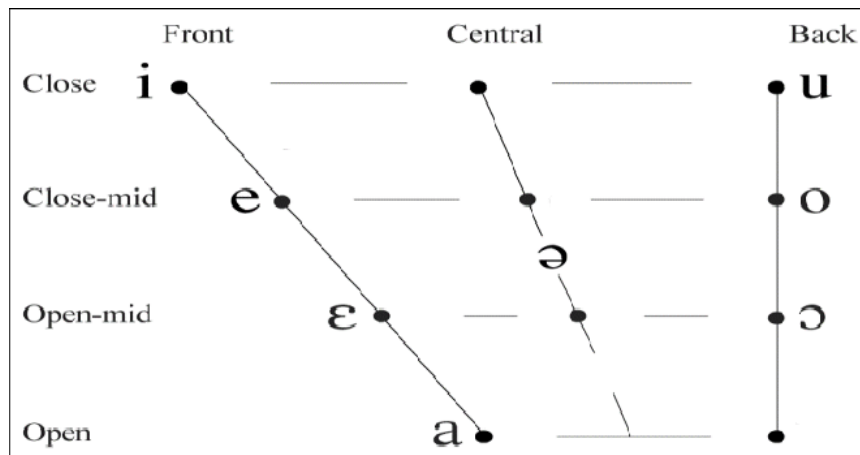


Fig 9 Malay Phonology chart

Examples for Malay Vowels

Vowels & Diphthongs

a	e	i	o	u	ai	au	oi
[a/ɑ/ə]	[e/ɛ]	[i]	[o]	[u]	[ai]	[au]	[oi]

5.3 Syllables Structure In Tamil and Malay Language

A syllable is the set of pronunciation which has 1 vowel sound with or without sounding the consonants. It is preceded by low sonority onset later followed by another set of low sonority coda.

Syllable	Example
V	அ, ஆ, இ, ஈ, உ, ஊ.....
CV	க, கா, த, தா, ல, லை, லை, தொ
V.CV	ஆகா, இலை, ஊசி,
CV.C	தட், டிப், நிப், வன், கன்,
V.C	ஆல், இன், உன், அப்,
CV.CC	வீழ்ந், வாழ்க், கிர்த்

Fig 11 Tamil syllable structure

Type of Syllable	Syllable structure	Example word
Monosyllabic	CVC	gam, cik, sup
	CV	di, ke, si
Disyllabic	CV+CV	beli, cuba, biru
	V+ CV	ini, itu, api
	V+ CVC	amat, ikan, ubat
Trisyllabic	CV+ CV+CV	lelaki, berasa, kerapu
	CV+V+CV	suara, buaya, kuasa
	CV+ CV+V	semua, ketua, deria
Tetrasyllabic	CV+CV+CV+CV	matahari, daripada, serigala
	CV+CVC+CV+VC	perempuan,
	V+ CV+ CV+ CV	apabila, adakala,

Fig 12 Malay Syllable structure

6. Phonemes Classification for speech recognition

ANN contribution in SR is considered as a most important process. This can be done using phoneme classifier, probability estimator of Speech recognizers, and isolated word recognizers. This chapter tries to highlight the segmentation of isolated words into phoneme utilizing MLP and clustering features using SOM.

6.1 PHONEME SEGMENTATION:

Segmenting continuous speech into particular phonemes is one of the basic processes in speech processing mainly in

1. Speech recognition
2. Speech synthesis
3. Speech database
4. Speech analysis

Reliable and accurate phoneme segmentation is the common factor in the requirement of the application to satisfy the needs. Many methods have been used for phoneme segmentation but some of them showed better performance because of having some phonetic knowledge. But certain methods based on rules and regulation is very difficult to optimize the performance. The performance of the segmentation degrades in real-time application. To face the disadvantage NN based approach must be implemented with the Conventional Rule-based method. This method will give a significant performance under the presence of disturbance and noise.

The MLP in phoneme segmentation has 1 hidden layer and 1 output layer. They also have 72 feature parameters for 4 consecutive frames and this is served as an output dataset. The output layer has 1 one and it has the right to decide on the 2nd frame whether it is a phoneme boundary or not. The hidden and output layer is used as the activation function. The number of nodes in the hidden layer will be changed according to the performance of the experiment.

6.2 SOM model for speech recognition model

The main aim of this model is the decrease the dimension of feature vectors by utilizing the Self-Organizing Map (SOM) in SR. fig (13) illustrates the SOM model.

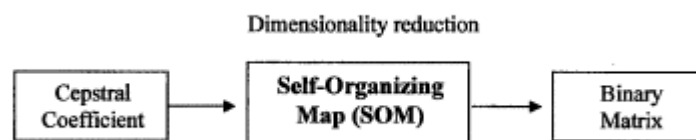


Fig 13 SOM model

The dimensionality of acoustic features has been reduced by introducing them with a recognizer block. Through which the classification of the model seems to be very simple. [kohonen] proposed a framework using NN which is used for generating the self-organized property of the unsupervised learning process known as the SOM. All the input vectors are included in the network without expecting the preferred output. Enough number of input vectors will be added

to the network from input to output nodes of the weight vector. The weight vector will be arranged based on the topology order of nodes present in the network.

[17] SOM model stores the topology order in the original space. The main motive of this model is to utilize the output of the self-organized map with speech processing output block which will reduce the feature vector through which the original behavior of the feature vector will be obtained and preserved.

Now the accurate number of neurons for the SOM model will be obtained. The obtained optimal size of SOM will ensure the Self-organized map with a sufficient number of neurons.

6.2.1 SOM ARCHITECTURE:

The SOM architecture is shown in fig (). It consists of 1 layer of neurons. The SOM model is arranged in 2x1 lattice and 2x1 lattice. This model helps in identifying the similarity measure through Euclidean distance measurement.

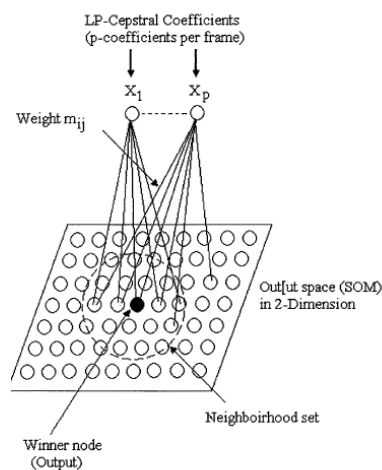


Figure 5.2: The 2-D SOM architecture

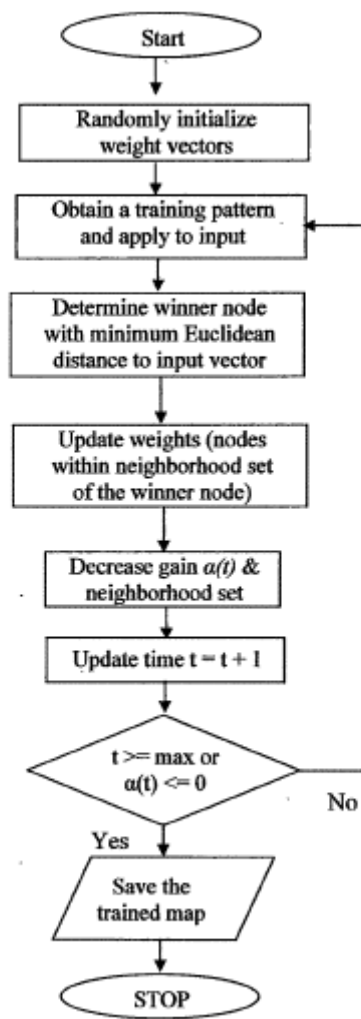


Fig 14. Flow Chart for SOM learning algorithm

7. Experimentation for speech recognition

In this model, the feature vector of phoneme segmentation is obtained from 5 input through the consecutive frame. The difference in the inter-frame is found between the 2 consecutive frames, and 4 inter-frame is placed in a range between 40 to 5. For every 5 consecutive frames, the interface difference is obtained and out of 4 interframes, one interframe contains 18 elements. These 72 element acts as the input for the MLP phoneme segmenter. In this present research for about 12 hidden nodes in MLP has been used. the MLP model obtained an accuracy of 65 percent for 10 m/sec and 83 percent for 20 m/sec duration. The feature vector for phoneme will be isolated.

The features obtained from the MLP will now undergo a time wrapping mechanism which is utilized to make them equal in length and then later passed for the Kohonen SOM model with 6 clusters.

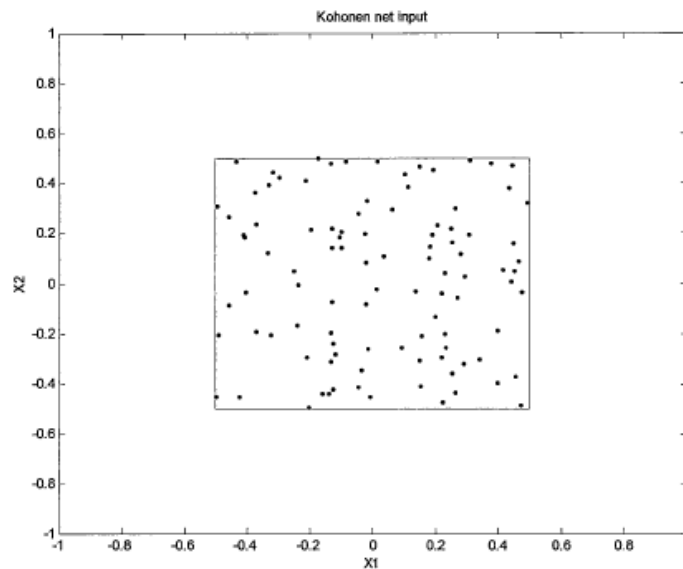


Fig 15 Kohonen SOM model

Fig (15) shows the Kohonen SOM model with input for the Tamil phoneme format /aa/ of the word 'Aalu'. The input for this model has been taken from the speech input pattern and it will be saved like an array. The input array is normalized for making the network work efficiently. The weights will range between +1 and saved into the array weight based on the Dimension of the SOM model. The training process will be carried out until it reaches the maximum epoch. The input vector from the input array will be selected on a random basis for learning and functioning to determine the winner node which is placed in the closest distance when compared to all other input nodes. Again the weight vector of the node placed in the closest distance will be made as to the new winner. This process will continue until it reaches the maximum epoch.

8. Conclusion

The neural network is a promising technique for speech recognition. The research directions in this area are fairly diverse and almost none of the existing approaches outstandingly dominate over the others. Indeed, speech recognition is well known for being a complex pattern recognition problem that can usually be divided into several sub-problems. It is extremely important to have a broad and thorough understanding of the nature of each sub-problem and accordingly choose the most appropriate neural network model and training algorithm to deal with it. The artificial

Neural Network Approach is the most highly relied on one. Besides, some fundamentals of Neural Network is reviewed, based on the topology and type of learning. It is expected that the present study has contributed towards the development of the recognition of Tamil and Malay words by using neural networks. The proposed model combines two neural networks namely SOM and MLP. The evaluation of the performance of the proposed model is made through its recognition accuracy.

Reference

1. Nilsson, Mikael, and Marcus Ejnarsson. "Speech recognition using hidden markov model." (2002).
2. Srinivasan, A. "Speech recognition using Hidden Markov model." *Applied Mathematical Sciences* 5.79 (2011): 3943-3948.
3. Pal Thamburaj , K., & Sivanathan, S. (2020). Marapu vali elututal tiranum tarkala elututal tiranum oru parvai [A study on modern and traditional writing skill]. *Muallim Journal of Social Sciences and Humanities*, 4(4), 141-146. <https://doi.org/10.33306/mjssh/103>
4. Kumar, Rajesh, et al. "Murmured Speech Recognition Using Hidden Markov Model." *2020 7th International Conference on Smart Structures and Systems (ICSSS)*. IEEE, 2020.
5. Muhammad, Hariz Zakka, et al. "Speech recognition for English to Indonesian translator using hidden Markov model." *2018 International Conference on Signals and Systems (ICSigSys)*. IEEE, 2018.
6. K. Pal Thamburaj, L. Arumugum and S. J. Samuel, "An analysis on keyboard writing skills in online learning," *2015 International Symposium on Technology Management and Emerging Technologies (ISTMET)*, Langkawai Island, 2015, pp. 373-377, doi: 10.1109/ISTMET.2015.7359062.
7. El Ouahabi, Safâa, Mohamed Atounti, and Mohamed Bellouki. "Optimal parameters selected for automatic recognition of spoken Amazigh digits and letters using Hidden Markov Model Toolkit." *International Journal of Speech Technology* 23.4 (2020): 861-871.
8. Yavuz, Erdem, and Vedat Topuz. "A phoneme-based approach for eliminating out-of-vocabulary problem of Turkish speech recognition using Hidden Markov Model." (2018).
9. Rynjah, Fairriky, Bronson Syiem, and L. Joyprakash Singh. "Khasi Speech Recognition using Hidden Markov Model with Different Spectral Features: A Comparison." *Available at SSRN 3515823* (2020).
10. Techini, Elhem, Zied Sakka, and Medsalim Bouhlel. "Robust Front-End Based on MVA and HEQ Post-processing for Arabic Speech Recognition Using Hidden Markov Model Toolkit (HTK)." *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2017.
11. Shi, Lin, et al. "Hidden Markov model based drone sound recognition using MFCC technique in practical noisy environments." *Journal of Communications and Networks* 20.5 (2018): 509-518.
12. Rashmi, S., M. Hanumanthappa, and Mallamma V. Reddy. "Hidden Markov Model for speech recognition system—a pilot study and a naive approach for speech-to-text model." *Speech and Language Processing for Human-Machine Communications*. Springer, Singapore, 2018. 77-90.
13. YUSOF, NORMIZA BINTI MOHD. "ISOLATED MALAY SPEECH RECOGNITION USING FUZZY LOGIC." (2019).
14. Winursito, Anggun, Risanuri Hidayat, and Agus Bejo. "Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition." *2018 International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2018.

15. Dua, Mohit, Rajesh Kumar Aggarwal, and Mantosh Biswas. "Discriminative training using heterogeneous feature vector for Hindi automatic speech recognition system." *2017 International Conference on Computer and Applications (ICCA)*. IEEE, 2017.
16. Haridas, Arul Valiyavalappil, Ramalatha Marimuthu, and Vaazi Gangadharan Sivakumar. "A critical review and analysis on techniques of speech recognition: The road ahead." *International Journal of Knowledge-based and Intelligent Engineering Systems* 22.1 (2018): 39-57.
17. Vadwala, Ayushi Y., et al. "Survey paper on different speech recognition algorithm: Challenges and techniques." *Int. J. Comput. Appl.* 175.1 (2017): 31-36.
18. Londhe, Narendra D., Ghanahshyam B. Kshirsagar, and Hitesh Tekchandani. "Deep convolution neural network based speech recognition for Chhattisgarhi." *2018 5th international conference on signal processing and integrated networks (SPIN)*. IEEE, 2018.
19. Dong, Jinwei, and Shaohui Li. "English Speech Recognition and Multi-dimensional Pronunciation Evaluation." *Education Research Frontier* 10.3 (2020).
20. Chen, Z. R. "SPEECH RECOGNITION OPTIMIZATION OF INTERACTIVE SPOKEN ENGLISH INSTRUCTIONS." *Telecommunications and Radio Engineering* 79.14 (2020).
21. El Ouahabi, Safâa, Mohamed Atounti, and Mohamed Bellouki. "Optimal parameters selected for automatic recognition of spoken Amazigh digits and letters using Hidden Markov Model Toolkit." *International Journal of Speech Technology* 23.4 (2020): 861-871.
22. Gupta, Avishek, and Kamal Sarkar. "Recognition of spoken bengali numerals using MLP, SVM, RF based models with PCA based feature summarization." *Int. Arab J. Inf. Technol.* 15.2 (2018): 263-269.
23. Butt, Muheet Ahmed, et al. "MULTIPLE SPEAKERS SPEECH RECOGNITION FOR SPOKEN DIGITS USING MFCC AND LPC BASED ON EUCLIDEAN DISTANCE." *International Journal of Advanced Research in Computer Science* 8.9 (2017).
24. Dalsaniya, Nikunj, et al. "Development of a Novel Database in Gujarati Language for Spoken Digits Classification." *International Symposium on Signal Processing and Intelligent Recognition Systems*. Springer, Singapore, 2019.
25. Dalsaniya, Nikunj, et al. "Development of a Novel Database in Gujarati Language for Spoken Digits Classification." *International Symposium on Signal Processing and Intelligent Recognition Systems*. Springer, Singapore, 2019.
26. Nisar, Shibli, et al. "Pashto spoken digits recognition using spectral and prosodic based feature extraction." *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE, 2017.
27. Lounnas, Khaled, et al. "CLIASR: A Combined Automatic Speech Recognition and Language Identification System." *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*. IEEE, 2020.
28. Ouisaadane, A., Said Safi, and M. Frikel. "English Spoken Digits Database under noise conditions for research: SDDN." *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*. IEEE, 2019.
29. Netshiombo, Dakalo, et al. "Spoken Digit Recognition System for an Extremely Under-resourced Language."
30. Ali, Hazrat, et al. "Pioneer dataset and automatic recognition of Urdu handwritten characters using a deep autoencoder and convolutional neural network." *SN Applied Sciences* 2.2 (2020): 152.
31. Mukherjee, Himadri, Santanu Phadikar, and Kaushik Roy. "An ensemble learning-based Bangla phoneme recognition system using LPCC-2 features." *Intelligent Engineering Informatics*. Springer, Singapore, 2018. 61-69.
32. Mukherjee, Himadri, et al. "READ—a Bangla phoneme recognition system." *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*. Springer, Singapore, 2017.
33. Rai, Aishwarya, et al. "An efficient online examination system using speech recognition." *International Research Journal of Engineering and Technology* 4.4 (2017): 2938-2941.
34. Wang, Dong, Xiaodong Wang, and Shaohu Lv. "An Overview of End-to-End Automatic Speech Recognition." *Symmetry* 11.8 (2019): 1018.
35. Rathore, Hritika, and Jyotsna Sagar. "An Alternative Voice Communication Aid based on ASR." (2017).

36. Alyousefi, Sarah. *Digital Automatic Speech Recognition using Kaldi*. Diss. 2018.
37. Stipinović, Karlo. *Prepoznavanje glasa algoritmima za obradu signala*. Diss. University of Split. Faculty of Maritime Studies. Department of maritime electr. and. inform. technologies., 2019.