A Review on Acoustic Phonetic Approach for Marathi Speech Recognition

Rohini B. Shinde,
College of Computer Science and Information Technology, Latur, Maharashtra. India

V. P. Pawar, PhD.
Associate Professor in Computer Science Dept of SRTM University, Nanded, Maharashtra. India

ABSTRACT

This paper discusses the phoneme used in Marathi language as a possible basic unit of speech recognition, for which there is some empirical psychoacoustic support in the case of human and some engineering justification in the case of machines striving to imitate human abilities. For the purpose of the research described in this paper, a basic unit of speech recognition is the intermediate form of speech information around which much of the recognition processing is organized for human beings or for machines. The general opinion of phonetician and psycholinguists is that there is indeed such a unit with relatively few distinct types[1]. For this research a basic unit is ideally an output of acoustic-phonetic processing and an input to the lexical processing stages.

Keywords:

ANN (Artificial Neural Network), Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT), Linear Predictor Coefficients (LPC), Swara (Vowels in Marathi), Vyanjana (Consonants in Marathi), MLSR (Marathi Language Speech Recognition)

1. INTRODUCTION

Over the last 40 to 50 years, researchers have proposed many different types of intermediate units. Some of the possibilities include sub-phoneme units, phones with right or left context, biphones, diphones [2] and variations[3], dyads or transemes [4], avents[5], triphones [6],demisyllables [7], whole words and phrases.

Current research in psychoacoustics and psycholinguistics suggest that the syllable might be a basic unit of human speech perception.

The paper has five sections,

1) Introduction
2) Acoustic-Phonetic Approach
3) Acoustic-phonetic Approach to Marathi Language speech Recognition (MLSR)
4) Marathi speech sounds and features
5) C-V Structure in Marathi Language
6) Acoustic Phonemes classifier
7) Experiment for vowel classification
8) Conclusion
9) Future Work:
10) References.

2. ACOUSTIC-PHONETIC APPROACH

Acoustic is deals with the study of different sounds and phonetic is the study of phonemes in the language. The acoustic-phonetic is based on the theory of acoustic-phonetics that postulates that there exist finite, distinctive phonetic units in spoken language and that the phonetics units are broadly characterized by a set of properties that are manifest in the speech signal , or its spectrum over time. Even though the acoustic properties of phonetic units are highly variable, both with speakers and with neighboring phonetic units it is also called as co-articulation of sounds

Following are some steps taken in the acoustic-phoneic approach to speech recognition[11]

1. Segmentation and labeling phase : In first step segmentation is done along with labeling phase because it involves segmenting the speech signal into discrete region where the acoustic properties of the signal are representative of one phonetic unit.

2. Determination of valid words from segmentation: second steps attempts to determine a valid word from the sequence of phonetic labels produced in the first step

3. ACOUSTIC-PHONETIC APPROACH TO MARATHI LANGUAGE SPEECH RECOGNITION (MLSR)

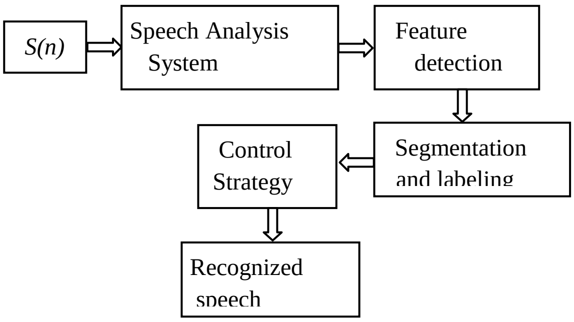


Fig:1 A block Diagram of acoustic-phonetic Marathi Speech Recognition

In the above diagram of the acoustic-phonetic Marathi Speech Recognition the first step is processing i.e. speech analysis system which provides an appropriate spectral representation of the characteristics of the time varying speech signal. The most common technique of spectral analysis are the class of filter bank methods and class of linear predictive coding (LPC) methods.

The next step in the processing is the feature-detection stage. In this step the spectral measurement are converted to a set of feature that describe the broad acoustic properties of the different phonetic units. The feature proposed for Marathi speech recognition are nasality, friction, formant location, voiced-unvoiced classification. The feature detection stage usually consist of a set of detectors that operate in parallel and use appropriate processing and logic to make the decision as to presence or absence, or value of a frame.

The third step in the procedure is the segmentation and labeling phase where by the system tries to find stable regions and then to label the segmented region according to how well the features within that region match those of individual phonetic unit.

To illustrate the steps involved in the acoustic-phonetic approach to speech recognition, consider the phoneme lattice shown in given fig:2

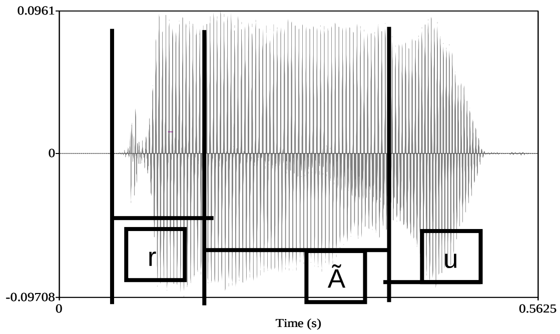


Fig:2 Phoneme Lattice for word string three(rhu)

In the Fig:2 example is taken of the numerical digit 0-10 in Marathi language. An important speech task is accurate digit recognition. This exploit the knowledge of acoustic phonetics to recognize first isolated digits, and next some connected digit strings. We first need a sound lexicon for the digits. The sound lexicon describes the pronunciation of digits in terms of the basic sound of Marathi.

4. MARATHI SPEECH SOUNDS AND FEATURES

Every language in a world have distinct speech sounds i.e. phonemes. Speech sounds of all languages are classified into vowels & consonants. Most popularly there are representing with some specific symbols generally called as Aksharas of that language. In phonetic terms & in linguistic terms vowels & consonants are defined as follows.

In phonetic terms, a speech sound is defined as a vowel. If in the production of it there is in the pharynx & the mouth no obstruction & no narrowing of a degree that would cause audible friction all these are vowels. All other sounds are taken as consonants.

The basic unit of the writing system in Indian languages are Aksharas, which are an orthographic representation of speech sounds. An Akshara in Indian language scripts is close to a syllable and can be typically of the following forms: C, V, CV, CCV, VC, VV, and CVC where C is Consonant & V is Vowel.[9]

The number of linguistically distinct speech sounds in a language is often a matter of judgment and is not invariant to different linguistics. Above tree structure displays list of phonetic symbols of Marathi language. As shown in table 1 & table 2 there are 12 vowels & 35 consonants.

Here we see the conventional set of 12 vowels, classified as front, mid, or back. Corresponding to the position of the tongue hump in producing vowels.

4.1 Swara (vowels)

In Marathi language there are total 12 swara (vowels) present which are denoted by following symbols

Sr. No.	Vowel	Pronunciation of vowel
1	अ	Pronounced as U in US
2	आ	Pronounced as A in ARM
3	ई	Pronounced as I in IT
4	ई	Pronounced as EA in EASY
5	उ	Pronounced as U in UTTAR PRADESH
6	ऊ	Pronounced as OO in OOZE
7	ए	Pronounced as E in EGG
8	ऐ	Pronounced as AI in AIR
9	ओ	Pronounced as O in OPEN
10	औ	Pronounced as OU in OUT
11	ऌ	Pronounced as RU in RULE
12	ऋ[1]	Pronounced as RU in RULE

Table 1 : Pronunciation of Vowels in Marathi

The vowel sounds are perhaps the most interesting class of sounds in Marathi. Most practical speech recognition systems rely heavily on vowel recognition to achieve high performance. To illustrate the this point consider the following section

Section 1

क.ण.त्य .ह भ.श.च .भ्य.स व ल.खन व.क्य ल.खन. ष.व.य प.र्ण ह.त न.ह.

Section2

.ओ.आ.इ आ.ए.आ अ.आ. . .ए.. .आ...ए.. आ .ई.आ. .ऊ. . ओ. .आ. ई

In section 1 we have omitted the conventional vowel letters, however, with a little effort the average reader can fill in the missing vowels and decode the section so that it reads.

कोणत्याही भाशेचा अभ्यास व लेखन वाक्यलेखनाषिवाय पूर्ण होत नाही.

In the section 2 we have omitted the conventional consents letters, the resulting text is essentially not decodable

4.2 Vyanjana (consonants)

In Marathi language, If in the production of sound there is in the pharynx & the mouth obstruction & narrowing of a degree that would cause audible friction, these sounds are consonants. In Marathi language generally consonants have the vowel V at the end, if this V vowel is removed from that Akashara then remaining is the consonant. There are 35 consonants in this language

e. g.

क्	ख्	ग्	घ्	ङ्	
च्	छ्	ज्	झ्	ञ्	
ट्	ठ्	ड्	ढ्	ण्	
	थ्	द्	ध्	न्	
प्	फ्	ब्	भ्	म्	
य्	र्	ल्	व्	ष्	
श्	स्	ह्	ळ्	क्ष्	ज्ञ्

Table 2: consonants after removing the vowel *V*

if the V vowel mixed in above consonant the Akasharas are form. Which are given below

[1] In Marathi language _ alphabet is used in some words only like fir§.k ekr§.k

क	ख	ग	घ	ङ	
च	छ	ज	झ	ञ	
ट	ठ	ड	ढ	ण	
त	थ	छ	ध	न	
प	फ	ब	भ	म	
य	र	ल	व	ष	
श	स	ह	ळ	क्ष	ज्ञ

Table 3: consonants after adding the vowel *V*

These consonants are grouped into 5 classes depending upon their pronunciation & articulatory system. [15]

5. C-V STRUCTURE OF SYLLABLE IN MARATHI LANGUAGE

The syllable is made up of one or more than one speech sounds. Speech sounds are either vowels or consonants. The vowel element is essential to the structure of a syllable. That is a syllable is not possible without the vowel element.

CV STRUCTURE	Word	Formation of word
C	व	व
V	आ	आ
CV	दोन	द. ओ. न
CCV	कमान	क. म. आ. न
VC	आज	आ. ज
VV	आई	आ. ई
CVC	लोभ	ल. ओ. भ

Table 4: C-V structure of syllable

6. ACOUSTIC PHONEMES CLASSIFIER

From the fig it is clears that classification of phoneme is little bit similar to English phonemes. Similar to English Marathi phonemes also classified in 3 categories.

1. Swar (Vowel)
2. Swaradi (Diphthongs)
3. Vryanjan (Consonant)

According to this classification the vowels are long Vowels आ, ई, ऊ Short Vowels ;;अ, इ, उ ऋ द्ध, Mixed vowels;ए, ऐ, ओ, औ Diphthongs include ;अं आ ॲ ऑ Where as consonants classified as Nasal consonant ङए ञ ए णए नए म , Semivowels;य र ल व , Fricatives: Voiced fricatives;कए खए च छ ट ठ त थ प फ , & Unvoiced fricatives;ग घ ज झ ड ढ द धए ब भ ,Whisper(ह) , Independent consonant(ळ)

क्ष ज्ञ these are not the consonant or vowels they forms as
क्ष त्र क. श. अ
ज्ञत्र द. न. य. अ

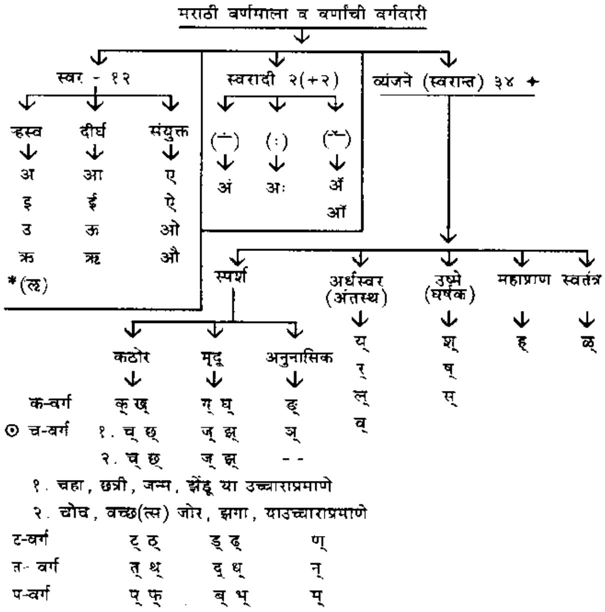
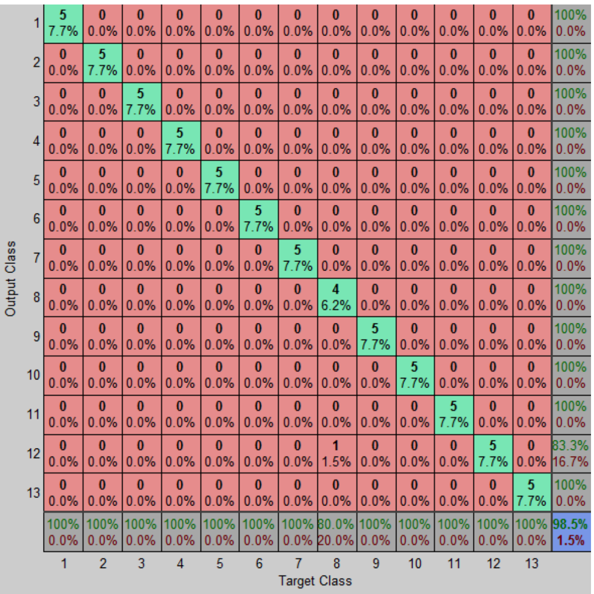


Fig 3 Classification of Phonemes (Marathi Language)[15]

7. EXPERIMENT FOR VOWEL CLASSIFICATION

The speaker recognition results were obtained using the generated database for this work. 10 features from each speaker's speech have been extracted using LPC. Total 65 samples of vowels (swara) are used for pattern recognition.65 samples are classified into 12 classes. In pattern recognition problems a neural network is used to classify inputs into a set of target categories. The proposed features have been tested on a Artificial Neural Network Using a MATLAB tool. The Neural Network Pattern Recognition Tool will help to select data, create and train a network, and evaluate its performance using mean square error and confusion matrices. The result of the Speaker recognition is shown below Fig 4(a), Fig 4(b) respectively in the form of confusion matrix and mean square error.



1	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 6.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
11	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	0 0.0%	100% 0.0%
12	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 1.5%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	0 0.0%	83.3% 16.7%
13	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 7.7%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	80.0% 20.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	98.5% 1.5%
	1	2	3	4	5	6	7	8	9	10	11	12	13	

Fig. 4 (a): Confusion Matrix displaying 98.5% recognition rate

8. CONCLUSION

In this paper experiment is done on only vowels in Marathi language.

Result of the experiment:- The recognition rate of person through selected SRS system is 98.5%.

Application of the result:- The selected SRS system can be applied for all Marathi language people who can speak and read Marathi language in appropriate form all the world. We hope the SRS created will serve as baseline system for further research on improving.

9. ISSUES OF ACOUSTIC SPEECH RECOGNITION SYSTEM FOR MARATHI LANGUAGE.

So many problems are associated with the acoustic phonetic approach to speech recognition.

1. Recognition Rate for a word or the continuous speech is less.
2. Vary in expressions while recording the utterance of speaker in different mode. (happy expressions, sad expressions, crying expressions, etc….)
3. Limitations regarding with health problem (basically in case of COLD)
4. Age wise variations may occur in speech so features may vary.
5. The method used for recognition system needs extensive knowledge of the acoustic properties.
6. Features are based on intuition and is not optimal in a well-defined and meaningful sense.
7. In case of Marathi Language the tone is vary from district to district.

Because of all these problems the Acoustic Phonetic Approach in Marathi Language is an interesting area for research work.

10. FUTURE WORK:

Future work of this paper is dedicated to Marathi Continuous Speech. Which may Useful for the Marathi language Recognition System (MLRS). Marathi is an Indo-Aryan language spoken by 90 million people all over the world & mainly used in Maharashtra state in India. There is a lot of scope to develop system using Indian-languages of different aspects and variations.

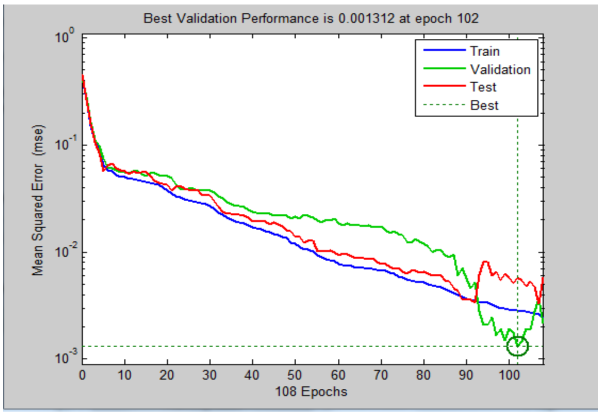


Fig. 4(b): Mean Square Error (MSE) displaying best validation performance at epoch 102

From the confusion matrix it is cleared that the 65 samples from the 13 speaker each is correctly classified into 13 classes only one sample is get misclassified & recognition rate is obtained 98.5% & 1.5% error rate. Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

11. REFERENCE

[1] Takayuki Arai and Steven Greenberg. “The temporal properties of spoken Japanese are similar to those of English.” Published in Eurospeech, Rhodes, Greece September 1997. ESCA.

[2] Richard Schwartz, Jack Klovstad, John Makhoul, and John Sorensen. A Preliminary design of a phonetic vocoder based on a diphone model. In ICASSP, Volume 1, Pages 32-35, Denver, Colorado, April 1980 IEEE

[3] M. Cravero, R. Pieraccini. And F. Raineri. “Definition and evaluation of phonetic units for speech recognition by hidden Markov Models.” In ICASSP, volume 3, pages 2235-2238, Tokyo, Japan, April 1986. IEEE.

[4] N. Rex Dixon and Harvey F. Silverman. “The 1976 modular acoustic processor.” IEEE Transactions of Acoustics, Speech and signal processing, ASSP-25(5):367-379, October 1977.

[5] Nelson Morgan, Herve Bourlard, Steve Greenberg, and hynek Hermansky. Stochastic Perceptual auditory-event-based models for speech recognition. In ICSLP, Pages 1943-1946, Yokohama, Japan, September 1994.

[6] L. Bahl, P. Cohen, A. Cole, F. Jelinek, B. Lewis, and R. mercer. “Further results on the recognition of a contineously read natural corpus.” In ICASSP, Volume 3, pages 872-875. Denver, Colorado, April 1980. IEEE.

[7] Osamu Fujimura. “Syllable as concatenated demisyllables and affixes.” Journal of the Acoustical Society of America, 59 (suppl. 1):S55,Spring 1976.

[8] Osamu Fujimura. “Syllable as a unit of speech recognition.” IEEE Transactions on Acoustics, Speech and signal processing, ASSP-23(1):82-87, February 1975.

[9] Gopalakrishna Anumanchipalli, Rahul Chitturi, "Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems"

[10] "Digital Signal Processing" By-P.Ramesh Babu Scitech Publications (India) PVT, LTD.

[11] "Fundamental of Speech Recognition" By-Lawrence Rabiner , Biing-Hwang Juang, Published by Pearson Education (Singapore) Pte. Ltd. Indian Branch.

[12] "Digital Signal Processing" A MATLAB based approach. By- Vinay K. Ingle, John G. Proakis.

[13] "Digital Signal Processing-Principles, Algorithms and Applications" John G. Proakis., Dimitris G. Manolakis.

[14] "Digital Signal Processing" by Farooq Husain.

[15] "Marathi Grammar Book" by Shripad Bhagwat.

[16] "A course in Phonetics and Spoken English"-J. Sethi, P.V. Dhamija

AUTHOR'S PROFILE

R. B. Shinde received the MSc(CS) degree from Dr. B. A. M. University, Aurangabad, in the year 2001. She is currently working as lecturer in the College of Computer Science and Information Technology, Latur, Maharashtra. She is leading to Ph.D degree in S.R.T.M.University, Nanded.

Dr. Vrushsen V. Pawar received MS, Ph.D.(Computer) Degree from Dept .CS & IT, Dr. B. A. M. University & PDF from ES, University of Cambridge, UK. Also Received MCA (SMU), MBA (VMU) degrees respectively. He has received prestigious fellowship from DST, UGRF (UGC), Sakaal foundation, ES London, ABC (USA) etc. He has published 90 and more research papers in reputed national international Journals & conferences. He has recognized Ph. D Guide from University of Pune, S. R. T. M. University & Singhaniya University (India). He is senior IEEE member and other reputed society member. Currently working as a Associate Professor in CS Dept of SRTMU, Nanded.