IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 4, Issue 2, September 2015)

Analysis of Vowels for Accent in Isolated Word in Speech Recognition for Oriya Language

[1]Sanjaya Kumar Dash
Sanjaya_145@rediff.com

[2]Prof. (Dr.) Sanghamitra Mohanty
Sangham1@rediffmail.com

Abstract: **Since years, various intelligible speech parameters have been on the process of evaluation and some are implemented to solve these real time challenges such as speech recognition and speaker identification system. Accent in which the person speaks depends upon the speaker and the way in which it is spoken. Apart from geographical parameters, the acoustic parameters for speech recognition and the intonation parameters of how one speaks, play important roles in accent analysis. This paper discusses some of the parameters in word levels that may be used for accent analysis for Oriya, an official language of the Republic of India.**

Keywords: **Acoustic Parameters, Oriya, Republic of India.**

I. Introduction

In the process of speech recognition "who"?, "what"? and "how"? are the queries to be handled "who" answers the speaker identification. "What" responses on, what is spoken where as "how" retorts the way speaker delivers. Accent is the variation in utterance due to geographical differences. It is seen that a person from eastern part of province speaks in a different accent than a person belonging to northern part of the province. A general technique to deal with the large variability in the acoustic realizations of the phonetic classes in speech is to partition the data as per a linguistically significant variable. In this work, accent dependent phonetic models are used both as an analysis tool for pronunciation variation and in the attempt to improve speech recognition system. The number of phonemes (linguistically distinct speech sounds) in a language is often a matter of judgment for words uttered.

Speaker identification is the process of automatically identifying who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers. This paper analyses the basic parameter of speech like formants of different isolated words. The calculation of mean formant of vowels for words spoken by different speakers in .wav format speech wave file was done and found to have certain values which can be used for speech recognition as well as for speaker identification. The work is done by using wave surfer.

II. Accent Analysis

The vowel sounds are the most reliable source to achieve high performance for recognizing a word. The well defined duration and spectrum of vowels are easily and reliably recognized in order to significantly contribute to recognize speech.

e.g.:

K`ò – Kþ + { @ } + `ò

small duration

K_ò – Kþ + { @ } + _ò

long duration

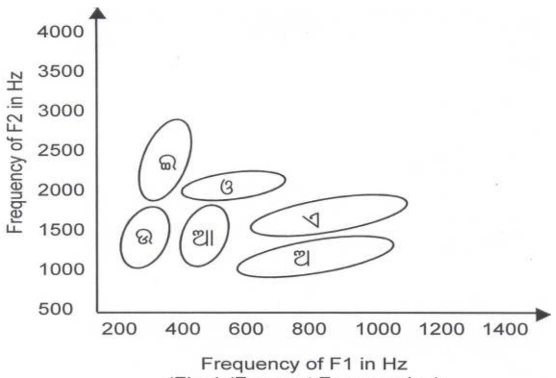
The Oriya vowels are long and are spectrally well defined. During speech production the movement of different articulators overlap in time for consecutive phonetic segments and interact with each other. As a consequence, the vocal tract configuration at any time is influenced by more than one phonetic segment. Due to this co articulation the same phoneme can have very different acoustic characteristics depending on the context in which it is uttered. We have analyzed the occurrence of vowels in ten Oriya words of four speakers belonging to different accent areas. Each Oriya word is uttered five times.

In the below Fig-1 we have the measured frequencies of first and second formants for different speakers for several vowels.

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 4, Issue 2, September 2015)



(Fig-1 (Formant Frequencies)

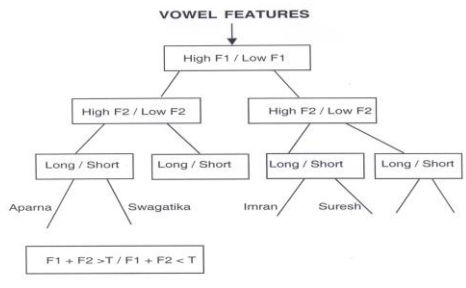
For the word ùaieKûeú the mean formant frequency for /G/ vowel of speaker Aparna is listed in below Table-1.

Table-1:

	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
Aparna	1058	2120	2886	3978
Imran	450	1890	2775	3988
Suresh	254	2092	3776	4365
Swagatika	913	1987	2829	4032

Accent Normalization is done to account for the variability in formants. Vowel classification is computed on the phonetic labeling procedure of an acoustic phonetic recognizer. We have detected three features over the vowel segment namely first formant F1, second formant F2 and duration of segment D to classify steady vowels of speakers on isolated words.

Taking the word ùaieKûeú the vowel /G/ distinguishes the speakers by taking above three features. The method uses a decision tree to make the decision. In the below Fig-2, where T is the Threshold value.



(Fig-2 - Decision Tree)

Other features such as Pitch, spectrum, power etc. play significant role for robust speech recognition systems to study accent variation. A sophisticated and robust distance measure [3, 4] can be used to develop speech recognition system.

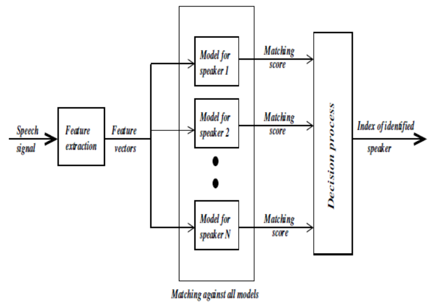


Figure-3: Decision process of speaker Identification

III. ACCENT DEPENDENT TRAINING

Parameters for Accent analysis can be evaluated by selecting people from different parts of Orissa to read different Oriya words. Various time domain as well as frequency domain features can be considered as few of the parameters, which are quantized to categories the accent.

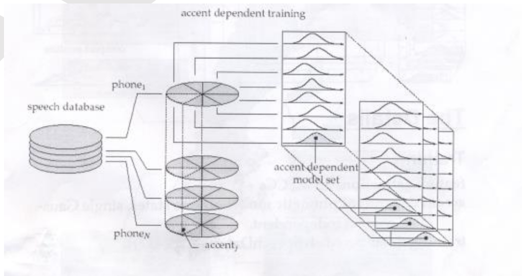


Fig-4 – Accent Dependent Training Model for speech recognition

The word level parameters include the length of the words. The occurrence of vowels in a word are important parameters in accent. In some of the cases it is observed that middle vowel (maAtrA) is pretty small compared to normal one. The length of the word mainly depends on the duration of vowels (mAtrA) within it. The duration of vowel /@û/ is more due to presence of nasal voice after its utterance for words like cjû^, iõiÚû^ etc. The duration for /@û/ is less for utterance of words like Z[û_ò, K\û_ò etc.

Applying Hamming window on speech segments the FFT transformation and LPC analysis is made for different predictor order like P = 8,12,16,20 and 128 point FFT. The LPC

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 4, Issue 2, September 2015)

model is very good for vowel sounds if the correct order, P, is used.

IV. Mathematical Formulations

Taking the joint probability density function of the measurements for the i^{th} speaker as a multi-dimensional Gaussian distribution[2] with mean m_i and covariance matrix W_i and x be an L-dimensional column vector representing the input pattern, the L-dimensional Gaussian density function for x is given as

$$g_i(x) = (2\pi)^{-1/2} |W_i|^{-1/2} \exp[-\tfrac{1}{2}(x-m_i)^t W_i^{-1} (x - m_i)]$$

The decision rule which minimizes the probability of error implies that measurement vector x will be assigned to class i if:

$$p_i g_i(x) \geq p_j g_j(x) \quad \text{for all i not equal to j}$$

Where p_i is the a prior probability that x belongs to i^{th} class.

We can decide class i if:

$$d_i(x) = \tfrac{1}{2}(x-m_i)^t W_i^{-1} (x - m_i) + \tfrac{1}{2} \ln |w_i| - \ln p_i \leq d_j(x) \text{ for all i not equal to j}$$

=> The measured distance $\mathbf{d_i} = (x - m_i)^t W_i^{-1} (x - m_i)$ is used

The decision rule uses $m_i = 1 / N_i \sum_{n=1}^{N_i} x_i(n)$

And $W_i = 1 / N_i \sum_{n=1}^{N_i} x_i(n) x_i^t(n) - m_i m_i^t$

V. Result

Taking 128 point FFT Transformation and LPC analysis, the peaks found for different predictor orders are substantially different for word uttered by speakers of different accent areas taking vowel occurrence into account. From the above analysis of four different speakers of different utterances, different distinguishing features are marked and are considered to be some of the effective parameters of accent analysis for speech recognition in case of Oriya.

VI. Conclusion

According to the result in Table-1 Aparna has the highest F1 formant value and Suresh has the lowest F1 formant value. The duration of utterance for Aparna is highest and the same for Suresh is the lowest. For other vowels the comparisons play the vital role for decision making of different words.This finding can be used for identification of speaker.

References

[1] Fundamentals of speech recognition By L. Rabiner and B.H. Juang

[2] Digital Processing of speech signals By L.R. Rabiner and R.W. Schafer.

[3] B.S. Atal, “Automatic recognition of speakers from their voices”. Proc. IEEE, Vol 64, No.4, PP 460-475 April 1976.

[4] B.S. Atal “Effectiveness of Linear prediction characteristics of the speech wave for Automatic speaker identification and verification,” J. Acoust. Soc. Am, Vol 55, PP. 1304-1312, June 1974.

[5] Mohanty S, Bhattacharya S., Bose S., Swain S., “An Approach to Parametric Base Mood Analysis in Oriya Speech Processing” proceedings of Frontier of Research in Speech and Music (FRSM)’, ITCSRA Kolkata, India (2005).

[1]Assistant Professor-Department of Computer Science and Engineering, Orissa Engineering College, Bhubaneswar, Odisha.

[2]Former Professor- P.G. Department of Computer Sc. and Application-Utkal University, Odisha

.