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

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.: 
\[ K\dot{\dot{o}} \rightarrow K\dot{\dot{p}} + \{ @ \} + \dot{\dot{o}} \]

small duration

\[ K_{\dot{\dot{o}}} \rightarrow K\dot{p} + \{ @ \} + _{\dot{\dot{o}}} \]

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

<table>
<thead>
<tr>
<th></th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
<th>F3 (Hz)</th>
<th>F4 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aparna</td>
<td>1058</td>
<td>2120</td>
<td>2886</td>
<td>3978</td>
</tr>
<tr>
<td>Imran</td>
<td>450</td>
<td>1890</td>
<td>2775</td>
<td>3988</td>
</tr>
<tr>
<td>Suresh</td>
<td>254</td>
<td>2092</td>
<td>3776</td>
<td>4365</td>
</tr>
<tr>
<td>Swagatika</td>
<td>913</td>
<td>1987</td>
<td>2829</td>
<td>4032</td>
</tr>
</tbody>
</table>

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 ūaieKûeū 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.

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.

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.

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 cû, iûû etc. The duration for /@û/ is less for utterance of words like Zûô, Kûô 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...
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

\[
g_i(x) = \frac{1}{(2\pi)^{L/2}|W_i|^{1/2}} \exp \left[-\frac{1}{2} (x-m_i)^t W_i^{-1} (x-m_i)\right]
\]

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

\[
p_i g_i(x) \geq p_j g_j(x) \quad \text{for all } i \neq j
\]

Where \(p_i\) is the a priori probability that \(x\) belongs to \(i^{th}\) class.
We can decide class \(i\) if:

\[
d_i(x) = \frac{1}{2} (x-m_i)^t W_i^{-1} (x-m_i) + \frac{1}{2} \ln |W_i| - \ln p_i
\]

\[
\leq d_j(x) \quad \text{for all } i \neq j
\]

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

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

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

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

1Assistant Professor-Department of Computer Science and Engineering, Orissa Engineering College, Bhubaneswar, Odisha.
2Former Professor- P.G. Department of Computer Sc. and Application-Utkal University, Odisha.