Formant Frequency and Cepstral Method Estimation of Galo Phonemes Using Acoustical Cues

Marpe Sora, Jyotismita Talukdar, and P. H. Talukdar

Abstract—**The main aim of this paper is study the acoustical cues of Galo language. The Galo is the major language of Arunachal Pradesh, the Northern state of India. The Galo tribe belongs to the Tibeto-Burman family of languages. The classification and Identification of different tribes/sub-tribes of Arunachal Pradesh can be made based on the "dialect" of the languages. There is marked variations of the same Galo Dilects with respect to locations and environment. These variations are noticed in terms of Prosody, Intonations, Pitch and durations etc. It is the aim of this paper to explore the acoustical features of this language based on Cepstral features.**

Index Terms—**Formant frequency, cepstral coefficient, galo language.**

I. Introduction

The *Galo* language is inherited from the Tibeto-Burman family of languages. This language is mostly spoken by the Galo people. of Arunachal Pradesh. The Galo is one of the major tribes of Arunachal Pradesh. Around 95% of Galo people learn Galo as a first language, although most are also bilingual and borrow frequently from Assamese, Hindi and English(The major languages of Indian subcontinent). In the Arunachal Pradesh there are total 25 major tribes and almost 110 sub- tribes. There is high degree of mutual intelligibility among the different languages of Arunachal Pradesh like language spoken by the Adis, Apatanis, Galos, the Hill Miris, the Nyishis and the Tagins. Moreover they share many characteristic features in their cultural code and trace their ancestry from a common forefather, namely Abotani. Hence, the language spoken by them can rightly be given generic name –Tani language[1]. The languages in Arunachal Pradesh can broadly be classified into two groups: namely *Abotani group* and Non-Abotani(Buddhism). The majority of the tribes of the state belong to Abo Tani group and Galo belong to Tani group[6]. These language groups are very close both in syntax and semantics. Galo people can understand each other when speaking with different Galo dilects .From region to region, village to village, and clan to clan, Galo people speak slightly differently in pronunciation and vocabulary. Sometimes differences are in pronunciation, sometimes in the actual words used, sometimes in the meaning of those words, and sometimes in the way they are used (i.e., the grammar). The major Galo dialects are *Pugo*, spoken around the district capital (Itanagar), *Aalo* and *Lare* spoken in the south of Aalo, and Subdialects are numerous, and often correspond to regional or clan groupings. The present study of the Galo language based on two major dialects i.e Pugo and Lare. The present script of Galo develop by Galo welfare

Society has gain acceptance among the speakers. Since vowel is the most distinguishing element of speech, feature vector extracted from the vowel have been considered for identifying the dialects. The Galo language has strong effects of intra-Tani contacts on the development of Tani languages. Galo tribe plays a major role as link language in the language scenario of Arunachal Pradesh. Galo tribe has contributed significantly in enriching the ethnic tradition in the state. The studies are more prominent today due to large changes in society. Galo script is based on Roman script it has seven vowels and seventeen consonant as segmental phonemes[5]. The super segmental phoneme is one tone to indicate high pitch point. In this study phonetic and phonemic variation of Galo vowel is examined in both time and frequency domain. An effort is made to study the vowel spectral characteristics corresponding to two samples sites-Aalo and Basar.

Manuscript received July 24, 2012; revised August 26, 2012.

Marpe Sora is with Department of Computer science and Engineering Rajiv, Gandhi University, Itanagar, CO 791112 INDIA (email: msorait@yahoo.co.in).

P. H. Talukdar is with the Department of Instrumentation & USIC, Gandhi University, Itanagar, CO 791112, India.

II. Formant Estimation for Vowel Reorganization

The formant estimation is based on digital resonator technique the entire frequency range divided into fixed number of segment, each segment representing a formant frequency[3]. Here for differentiation we have use the five point differentiation equation given in Equation (1). The interval should be chosen such that it is smaller than the value of the RR interval. The corresponding value of the differentiated wave at this point is calculated and a conditional loop is initiated to find out whether it is negative or positive. This step is the very important as the rest of the analysis depends on it. Depending upon what the value comes out to be, the lowest point to the left or to the right of the peak is calculated. This point is the R point.

$$f'(x) \approx \frac{-f(x+h)+8f(x+h)-8(x-h)+f(x-2h)}{2h} \quad (1)$$

The highest point to the left of the R value in the absolute differentiated wave is stored in the memory as the temporary point. From this point as a reference the lowest point to the left of this point is again calculated, to get the point. For the detection of the S point on the wave the procedure is the same as that for the wave detection. Here too the proper detection of the R peak is very important. After getting the R peak point, he highest point to the right of the R value in the absolute

DOI: 10.7763/IJIEE.2013.V3.258

differentiated wave is found out as temporary S point. From this point as a reference the lowest point to the right of this point is again calculated as the S point [2].

III. Feature Extraction Formant Method

The speech data (16 kHz sampling frequency) is used in this study. Only native speakers were selected. The quality of speech/voice was tested by the Galo phonetician. The recording duration was about 10 hours.

The formant estimation has received considerable attention in speech analysis and recognition during first few years [4]. In fact formant frequency estimation could be useful for various applications due to important role in determining the phonetic content as well as the close relation to vocal tract geometry. The present object is evaluate Galo Dialect is different from Lare speaker and Pugo. To evaluate the LPC method for the estimation of three formant frequencies for both Pugo and Lare Male and Female Speakers. The frequencies of first formant are denoted F1, F2 & F3 contain sufficient information for recognition of vowel.

The resolution and distinction of formant frequencies is clear in 3rd formant frequency. Among the mean formant frequencies considered, the range of frequencies of vowel.

Thus, it is observed that from the mean variation we can differentiate the dialect variation of Galo speaker of Arunachal Pradesh.

TABLE I: Variation Formant Frequencies (in KHz) of Galo Vowels for Aalo(Pugo) and Basar(Lare)Male and Female.

GV	Sampled Area	F1		F2		F3	
		Male	Female	Male	Female	Male	Female
|a|	Aalo Pugo	3.7998	2.2577	3.2234	1.7554	1.6469	1.7554
	Basar Lare	4.7710	4.2059	4.1510	4.1715	2.2600	4.1715
|i|	Aalo Pugo	11.978	1.7871	5.3503	1.3815	2.7553	1.3815
	Basar Lare	4.5172	4.8878	2.4322	0.9289	2.4322	0.7463
|u|	Aalo Pugo	7.7283	9.6015	4.0712	7.6021	2.6175	3.6790
	Basar Lare	4.4258	5.7557	3.7143	3.3200	3.7143	3.3200
|e|	Aalo Pugo	7.7892	1.0295	2.3768	0.7417	2.3768	0.5606
	Basar Lare	7.5013	10.355	3.8034	4.1923	3.8034	2.4029
|o|	Aalo Pugo	6.4488	2.6405	6.3886	1.8479	6.3605	1.4259
	Basar Lare	5.3528	7.2697	5.1698	4.1392	3.9844	3.0602
|v|	Aalo Pugo	4.7133	5.5823	3.8448	3.9549	3.1360	2.4089
	Basar Lare	6.7387	9.3168	4.3021	5.0999	4.1069	3.2500
|w|	Aalo Pugo	5.7099	7.4609	2.5870	3.8228	2.3414	3.8083
	Basar Lare	7.6980	5.4646	4.2699	4.4327	3.3392	3.7084

IV. Central Measure for Vowel Recognization.

In the recorded spectra of vowels, each speech spectra corresponding to each of seven vowels from two sampled sites is divided into 15 frames (1 frame =250 samples), and each frame giving 15 cepstral coefficients. In present study dialect, Galo dialects taken from two sites of west siang district namely Aalo and Basar. Each vowel spectra is sampled at the sampling rate 8000 samples per second and 15 frames. For each frame 15 cepstral coefficient has been calculated from given equation (3).

TABLE II(A): Formant Frequencies (in KHz) of Galo Vowels for Aalo(Pugo) and Basar(Lare)Male and Female.

Galo Vowel	Sampled Area	Range Male	Range Female	Mean	Std.Dev
|a|	Aalo(Pugo)	2.1529	0.5023	2.8900	1.1145
	Basar(Lare)	2.5110	0.0344	3.7273	1.3080
|i|	Aalo(Pugo)	9.2229	0.4056	6.6946	4.7561
	Basar(Lare)	2.0850	4.1415	3.1272	1.2038
|u|	Aalo(Pugo)	5.1108	5.9225	4.8057	2.6334
	Basar(Lare)	0.7115	2.4357	3.9515	0.4108

(A)

TABLE II(B): Formant Frequencies (in KHz) of Galo Vowels for Aalo(Pugo) and Basar(Lare)Male and Female.

Galo Vowel	Sampled Area	Range Male	Range Female	Mean	Std.Dev
|e|	Aalo(Pugo)	5.4124	0.4689	4.1809	3.1249
	Basar(Lare)	3.6979	7.9521	5.0360	2.1350
|o|	Aalo(Pugo)	0.0883	1.2146	6.3993	0.0451
	Basar(Lare)	1.3684	4.2095	4.8357	0.7429
|v|	Aalo(Pugo)	1.5773	3.1734	3.8980	0.7900
	Basar(Lare)	2.6318	6.0668	5.0492	1.4664
|w|	Aalo(Pugo)	3.3685	3.6526	3.5461	1.8779
	Basar(Lare)	4.3588	1.7562	5.1024	2.2955

(B)

The basic idea behind the LPC model is that a given speech signal at a time n, s[n], can be examined approximately as a linear combination of the past p speech samples[7] such that

$$S(n) \approx a_1 s(n-1) + a_2 s(n-2) + ... + a_p s(n-p) \quad (2)$$

The equation (2) can further be transformed by including an excitation term Gu(n) as:

$$S[n] = \sum_{i=1}^{p} a_i s[n-i] + GU(n) \quad (3)$$

where $a_i (i = 1,2,3....p)$ are coefficients and assumed to be constant over the speech analysis frame (n) is the normalized excitation and G is the gain of excitation. If

$$\hat{S}[n] = \sum_{k=1}^{p} a_k . s[n-k] \quad (4)$$

The error $e[n]$ in estimating $\hat{S}[n]$ from $S[n]$ is given by

$$e[n] = S[n] - [n] \quad (5)$$

The total squared error is then;

$$E \sum_n e[n] = \sum_n \left[s[n] - \sum_{k=1}^{p} a_k . s[n-k] \right]^2 \quad (6)$$

The minimum total square error E_p is obtained after expanding equation (6) substituting:

$$\left.\begin{aligned} &\sum_{k=1}^{p} a_i \sum_n s[n-k].s[n-i] = \sum_n s[n].s[n-i] \text{ and} \\ &\frac{\partial E}{\partial_{ai}} = 0 \quad 1 < i < p \end{aligned}\right\} \quad (7)$$

Thus $E_p = \sum_n s^2[n] + \sum_{k=1}^{p} a_k \sum_n s[n].s[n-k]$ (8)

The smaller the value of E_p, the better would be the estimated $\hat{S}[n]$ from $S[n]$ .To determine the [n], it is required to determine the predictor co-efficient $\{a_i\}$ over a short segment of speech signal occurring around n. Now, LPC cepstral co-efficient c[i] are computed from p^{th} order, Liner predictor coefficients are a[i] computed by following equation.

$$c[1] = a[1]$$

$$c[n] = a[n] \sum_{m=1}^{n-1} \frac{m}{n} a[m].c[n-m], 2 \le n \le p$$

$$c[n] = \sum_{m=1}^{p} \frac{n-m}{n} a[m].c[n-m], n > p$$

(9)

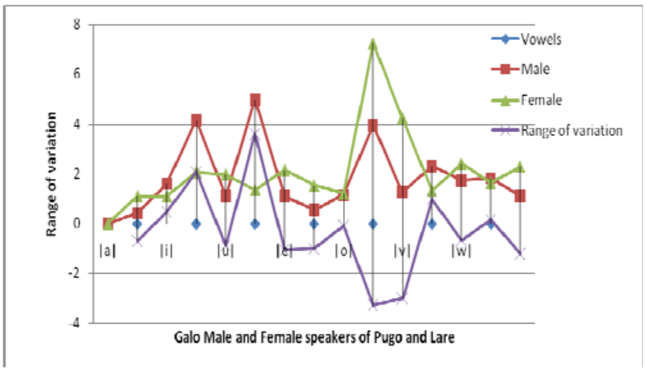

Fig. 1. Cepstral coefficient Range of variation of galo male and female

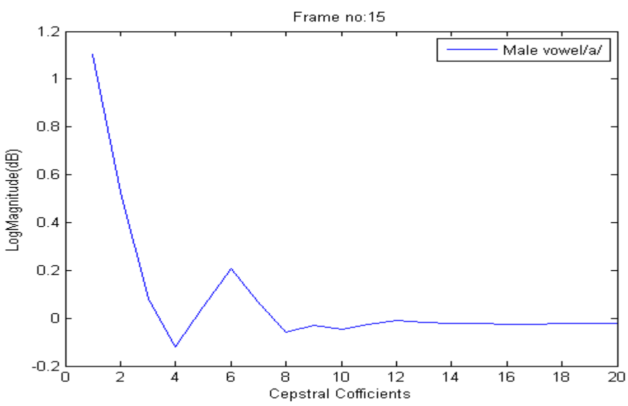

Fig. 1.2.Cepstral coefficients of male vowels/a/ Pugo

TABLE III(A): SPECTRAL FEATURES OF FEMALE VOWELS WITH C_{MIN} AND C_{MAX}

Vowel name	Sample area	Cepstral Coefficients(C)		Range of variation
		Cmin	Cmax	
|a|	Aalo Basar	-0.0626 -0.0301	1.0450 1.0763	1.1076 1.1064
|i|	Aalo Basar	-0.7395 -0.8143	1.3339 1.1712	2.0734 1.9855
|u|	Aalo Basar	-0.2971 -0.6348	1.0689 1.5333	1.366 2.1681
|e|	Aalo Basar	-0.4783 -0.0619	1.0592 1.1625	1.5375 1.2244
|o|	Aalo Basar	-2.5338 -1.3218	4.7004 2.9293	7.2342 4.2511
|v|	Aalo Basar	-0.2540 -0.8392	1.0810 1.5701	1.335 2.4093
|w|	Aalo Basar	-0.5631 -0.8180	1.0943 1.4844	1.6574 2.3024

(A)

TABLE III(B): SPECTRAL FEATURES OF MALE VOWELS WITH C_{MIN} AND C_{MAX}

Vowel name	Sample area	Cepstral Coefficients(C)		Range of variation
		Cmin	Cmax	
|a|	Aalo	0.6948	1.1041	0.4093
	Basar	-0.3334	1.2601	1.5935
|i|	Aalo Basar	-2.5448 -0.0156	1.6077 1.0961	4.1525 1.1117
|u|	Aalo Basar	-3.6401 -0.0301	1.3603 1.0734	5.0004 1.1035
|e|	Aalo Basar	0.6016 -0.0586	1.1457 1.0885	0.5441 1.1471
|o|	Aalo Basar	-2.3764 -0.1670	1.5790 1.0890	3.9554 1.256
|v|	Aalo Basar	-0.7111 -0.6460	1.6182 1.0955	2.3293 1.7415
|w|	Aalo Basar	-0.7040 -0.0121	1.1216 1.0867	1.8256 1.0988

(B)

TABLE IV: DIFFERENCE OF CEPSTRAL COEFFICIENT FOR MALE AND FEMALE SPEAKERS

Vowels	Male Speaker	Female Speaker	Range of variation
|a|	0.4093 1.5935	1.1076 1.1064	-0.6983 0.4871
|i|	4.1525 1.1117	2.0734 1.9855	2.0791 -0.8738
|u|	5.0004 1.1035	1.366 2.1681	3.6344 -1.0646
|e|	0.5441 1.1471	1.5375 1.2244	-0.9934 -0.0773
|o|	3.9554 1.256	7.2342 4.2511	-3.2788 -2.9951
|v|	2.3293 1.7415	1.335 2.4093	0.9943 -0.6678
|w|	1.8256 1.0988	1.6574 2.3024	0.1682 -1.2036

V. CONCLUSION

The cepstral envelope computed was generated based on the separation of the excitation and vocal tract resonances.

Vowel data were collected for 10 speakers from each sex, and were analyzed using the formant estimation and cepstral method which is believed to reflect vocal tract resonances. Significant variations among the speakers were observed for all the acoustic measures (Cepstral and formant frequencies). The data collected for each gender were compared. In the present study formant frequencies for male speakers were higher than those of female in Pugo and Lare especially in the formants F1 and F2.Formant F3 were found to be sensitive to locate. In order to confirm the accuracy of our algorithm, the standard deviation has been computed. Results show that a wide range in the estimated values of formant frequencies has been obtained. This confirms the limitation of the use of cepstral method to the estimation of formants especially at high frequencies.

The range of variation of Cepstral variations for Galo Pugo and Galo Lare male is found within the range of $5.00>Cepstral_{Pugo}>0.409and1.714>Cepstral_{Lare}>1.0988$ respectively. The range of variation for female is found $7.2342>Cepstral_{Pugo}>1.1076$ and $4.2511>Cepstral_{Lare}$ $i>1.1064$ i.e. the variation of Cepstral analysis for Pugo words is more (Male-4.5947; Female-6.1266) with respect to the Lare words (Male-0.6152; Female-3.1447) i.e., the former is stable as compared to the latter. This observation may be helpful in sex determination for both Pugo and Lare speakers of Galo dialects.

References

[1] G. Agom, *Galo Welfare Society*, Itanagar.

[2] S. Majumdar, S. Dhar, A. Shina, and A.Roy, "A hybrid wavelet and time plane based method for QT interval measurement in ECG signal," *ICSP 2008 proceeding*, 2008.

[3] L. Willing and H. Ney, "Formant Estimation for Speech Recognition," *IEEE Transaction on speech and Audio processing*, vol. 6, no. 1 January, 1998.

[4] M. Baudry, P. Deleglise, and J. C. Friedmann, "Speech analysis in the time domain using syntactic technics an attempt to formalize the description of phonemes using acoustical cues".

[5] M. William, *A script for Galo language*, September, 2007.

[6] P. T. Abraham, *A* Grammar *of Nyishi language*.

[7] P. H. Talukdar, U. Bhattacharjee, C. K. Goswami, and J. Barman, "The cepstral measure of Boro Vowels through LPC-Analysis," *Journal of the computer society of India*, vol. 34 no. 1, pp. 56-68, Jan-March 2004.

Marpe Sora obtained B. Tech form NERIST and M.Tech from Tezpur University. Presently he is Assistant professor in Rajiv Gandhi University, Department of Computer Science and Engineering Arunachal Pradesh. His main research interests include signal and speech processing and Data mining.

Joytismita Talukdhar obtained B. Tech form Guwahati and perusing Master in Computer Science from Asian Institute of Technology Bangkok Thailand. Her main research interests include signal and speech processing.

Prof. P. H. Talukdar is with Dept. of Instrumentation & USIC, GU (2006 to till date) Reader, Dept. of Instrumentation & USIC, GU (1996 to 2006) Assistant Professor, Dept. of MCA, Assam Engg. College (1991 to 1996) Assistant Professor, Dept. of Electronics & Telecommunication Engineering., Assam Engineering College (1986 to 1991) Lecturer, Dept. Of Physics, Cotton College, Guwahati (1982 to 1986). Lecturer, Dept. of Physics, Bajali College (1981 to 1982) Lecturer, Dept. of Physics, D K College, Mirza (1979 to 1981) Scientific Assistant, Defence R & D Organization, Dehradun (1976 to1979).He is author of six technical books. His main research interests include signal and speech processing.