A Statistical Approach to Pitch-Based Feature Analysis for Galo Tone Recognition

[1]Bomken Kamdak and [2]Utpal Bhattacharjee*
Rajiv Gandhi University, Rono Hills, Doimukh, Arunachal Pradesh, India. Pin – 791112
[1]bomken.kamdak@rgu.ac.in [2]utpal.bhattacharjee@rgu.ac.in

Abstract:

Tone is a critical characteristic in any tonal language recognition system. The Galo language, spoken in central Arunachal Pradesh, is a tonal language with limited linguistic resources. This study presents a statistical analysis of features extracted from the pitch (F0) contour of speech signals of Galo tonal words. Multivariate analysis of variance (MANOVA), t-test and mutual information were employed to identify the optimal feature sets for Galo tone recognition. Furthermore, a comparative study of feature selection algorithms was conducted using a Support Vector Machine (SVM) based tone recognition system. The SVM classifier was trained and tested using both the complete set of extracted features and the subset selected by the feature selection algorithm. Results indicated that the feature selection algorithm enhanced the performance of the SVM-based tone recognizer by 4.32%, demonstrating a significant performance improvement.

DOI Number: 10.48047/nq.2020.18.12.NQ20251 **NeuroQuantology 2020; 18(12):160-175**

* Corresponding author
eISSN1303-5150

1. Introduction

Tone is a crucial characteristic in any tonal language recognition system. Despite significant progress in tone recognition for languages like Mandarin Chinese, precise tone recognition remains a major challenge in the recognition of tonal speech [1]. The lexical meaning of a tonal word is determined by its underlying tone, as tonal words are phonetically similar. The recognition of any tonal language heavily relies on tonal features, including the pitch of the spoken word. While speech recognition technology has advanced to the point where numerous software applications can recognize and translate speech between languages, there remains considerable scope for improvement in the domain of tonal language recognition, especially for low-resourced languages.

The Galo language is a low-resourced language spoken by the Galo tribe of the Tani branch of the Tibeto-Burman language family. It is primarily spoken in the central part of Arunachal Pradesh, located in the northeastern region of India [2]. Among the 26 major tribes in Arunachal Pradesh, the Galo is one of the prominent ones [3], and their language is tonal. The Galo population predominantly resides in districts such as West Siang, Leparada, Lower Siang, Upper Subansiri, and East Siang. The lack of a written script for the Tani languages increases their vulnerability to extinction. To address this, the Galo Welfare Society (GWS) has developed the Modified Roman Script (MRS), which adapts Roman letters to better represent the phonetic content of the Galo language [4]. The Galo language includes several dialects, such as Pugo (mainly spoken in the West Siang district), Lare (spoken in Leparada, Lower Siang and Upper Subansiri districts), Bogum, Lodu, and Karka (spoken in Leparada and West Siang districts). Arunachal Pradesh, a hub for many tonal languages, particularly those of the Tani tribes, offers significant opportunities for research in tonal speech processing. This study focuses on the tones of the Lare dialect of the Galo language. Tones are generally classified into Level tones and Contour tones. Level tones remain constant throughout the tone-bearing unit (TBU) and are further categorized as High, Low, and Middle based on pitch level. In contrast, Contour tones exhibit a clear variation from one pitch level to another within the syllabic boundary and can be further classified into rising and falling tones.

In Galo, there are seven vowels and seventeen consonants [5]. Table 1 lists the Galo vowels, while Table 2 lists the Galo consonants. The Lare dialect features two types of tones: High/Plain, which is characterized by a medium-high level tone and Low/Tense, which is characterized by a falling tone [6].

Table 1. Galo Vowels

Galo	A a	I i	U u	E e	O o	V v	W w
Phonetic	a	i	u	e	o	ə	ɨ

Table 2. Galo Consonants

Galo	K k	G g	Q q	C c	J j	X x	T t	D d	N n	P p	B b	M m	Y y	R r	L l	S s	H h
Phonetic	k	g	ŋ	tɕ	dʑ	ɲ	t	d	n	p	b	m	j	r	l	s/ɕ	h

The paper proposes a novel approach to feature selection using Multivariate analysis of variance (MANOVA), t-test and mutual information to detect the most important features for tone discrimination in the context of the Galo language. These statistical techniques offer a comprehensive analysis of the relationships among variables within two distinct groups determined by the two tones of Galo language, namely High/level and

Low/rising. They help in identifying the most significant features exhibiting variations across the groups, as well as both linear and non-linear associations of dependent variables (features) with the groups formed by the independent variable (tones). This process plays an important role in eliminating irrelevant and redundant features that may impair the efficiency of the model and increase the level of computational complexity.

2. Related Works

The task of speech recognition entails selecting an optimal set of feature vectors. In tonal languages, pitch (F0) related features are of paramount importance. Early studies employed statistical computations of the mean, variation, and range of pitch parameters to classify Mandarin tones [7][8]. Cheng and Yang [8] introduced an empirical method based on the turning points of the pitch contour to identify tones; however, these studies were restricted to speaker-dependent recognition. For speaker-independent recognition of tones, Hsu [9] proposed a method where the fundamental frequency contour of a syllable was divided into four equal segments. Tone identification was performed based on the normalized slopes of these segments, achieving a recognition accuracy of 95%. This study highlighted the importance of speaker pitch-based adjustment in speaker-independent tone recognition. Chen et al. [10] employed delta modulation of pitch sequences as feature sets and applied Hidden Markov Modeling (HMM) to Mandarin tone recognition. Their method utilized energies, means and slopes from three equally divided sub-segments of the voiced portion of the monosyllable, along with duration, forming a 10-dimensional feature vector that achieved an accuracy of 93.8% [11]. Pitch normalization was performed to avoid variations due to different speakers. In 1993, Tan Lee [12] proposed the use of supra-segmental features - namely, duration, relative pitch level and the rising rate of pitch within the utterance—for Cantonese tone recognition, achieving accuracies of 89% for single speakers and 87.6% for multiple speaker. Yang, W-J., et al. employed logarithmic pitch intervals and their first derivatives for Mandarin lexical tone recognition, attaining accuracies of 98.33% for speaker-dependent and 96.53% for speaker-independent recognition [13]. Fu, Qian-Jie, et al. utilized temporal features, including envelope (2-50 Hz), periodicity (50-500 Hz) and fine-structure (500-10,000 Hz) tonal envelope cues, reporting a recognition accuracy of 80% in a 500 Hz envelope filter condition for tone recognition [14]. Common tone feature vectors typically include pitch or log pitch and their first and second derivatives, Energy and duration [15, 16, 17]. Additionally, the coefficients of a Legendre decomposition of the pitch contour have been used as tone features [18].

Selecting an optimal set of feature vectors can be accomplished through various filter methods that perform statistical tests to rank features based on their test scores. When feature selection is conducted using mutual information, the performance of machine learning models improves in terms of both prediction accuracy and training time [19]. A greedy feature selection method utilizing mutual information was introduced, combining both feature-feature mutual information and feature-class mutual information to identify an optimal subset of features that minimizes redundancy and maximizes relevance among features [20]. This approach has been found to significantly enhance performance in terms of classification accuracy and execution time. Hegde et al. employed Fisher's ratio technique for feature selection within the Automatic Speech Recognition (ASR) task. This investigation aimed to determine whether a subset of Mel Frequency Cepstral Coefficients (MFCCs) could achieve classification accuracy comparable to, or better than, that obtained using the full set of 12 coefficients. By utilizing the Fisher's ratio measure, the study revealed that selecting eight MFCC coefficients based on the F-ratio criterion enhances classification

accuracy relative to using all twelve coefficients. This technique underscores the potential of efficiently selecting a subset of MFCC features for ASR tasks, thereby highlighting the significance of statistical feature selection in enhancing ASR performance [21].

3. Material and Method

3.1 Galo Tonal Database

A Galo tonal speech database was developed through the collection of speech data from 27 speakers, comprising 13 males and 14 females, who are members of the Galo tribe in Arunachal Pradesh and fall within the 20-50 years age range. The recordings were carried out in a controlled acoustic environment at a sampling frequency of 16 KHz and a mono resolution of 16 bits. Participants were presented with a phonemically diverse Galo script containing 20 monosyllabic and 20 disyllabic tonal words, along with their corresponding tonal variations. Each speaker's voice was captured 8 times over two sessions, separated by a minimum of one week. This process resulted in the creation of a comprehensive database comprising 17,280 distinct tonal words. Evaluation of the database involved 5 native Galo listeners and 3 phonetically trained individuals. Only speech samples that received a favorable rating from over 60% of the native listeners and a minimum of two experts were selected for further analysis. As a result, the final database encompasses 14,623 tonal words.

3.2 Feature Extraction

A first-order digital filter, defined by the transfer function $H(z)=1-0.95z^{-1}$ is employed to pre-emphasize the digitized speech. The pre-emphasized speech signal is subjected to an energy-based Voiced Activity Detection (VAD) algorithm with the objective of detecting the voiced segments in the speech signal while filtering out the leading and trailing silent intervals. The speech signal is segmented into frames of 30 milliseconds duration each with a one-third overlap between successive frames. The Fundamental Frequency (F0) is determined for each frame individually. To address speaker-related variations in F0 values, Z-score normalization has been applied to the extracted F0 values. This normalization process is accomplished using the equation.

$$x^* = \frac{x - \mu}{\sigma} \qquad (1)$$

where μ and σ represents the mean F0 and the standard deviation of the F0 values, respectively, computed over all frames. The following F0-related features are extracted from the normalized F0-contour.

i. Mean, Variance and Standard deviation of F0

For each utterance, the mean, variance, and standard deviation of F0 values have been calculated.

ii. F0 Parameters derived by curve fitting

To obtain the F0-related features, we fitted the normalized F0 values to a second-order polynomial. The time is scaled to the interval [0, 1]. The normalized pitch is represented by $f(t), 0 \leq t \leq 1$. To fit the second-order polynomial to the pitch contour, we find the best value for the coefficients c_0, c_1 and c_2 in the least square sense in the following equation:

$$f(t) \approx c_0 + c_1(t - 1/2) + c_2[(t - 1/2)^2] - 1/12 \qquad (2)$$

In this polynomial, the coefficients are multiplied by the first three coefficients of the Legendre polynomial, translating them to the interval [0,1]. The polynomials are orthogonal in the interval [0,1]. Therefore, c_0 is the mean of $f(t)$, c_1 represents the slope of $f(t)$ and c_2 is the second derivative of $f(t)$, which represents the curvature [22]. These parameters thus give important insight for distinguishing the two tones of the Galo language.

iii. Initial Pitch P_I, Final pitch P_F and rising index I_R

The normalized pitch values $\{p_1, p_2, \dots, p_N\}$ has been used to compute the Initial Pitch P_I, Final pitch P_F and rising index I_R using the following equations [23]:

$$P_I = (p_3 + p_4)/2 \quad (3)$$

$$P_F = (p_{N-2} + p_{N-3})/2 \quad (4)$$

$$I_R = k\frac{Max_{i=2}^{N-1}\{p_i\} - Min_{i=2}^{N-1}\{(p_i\}}{Max_{i=2}^{N-1}\{p_i\} + Min_{i=2}^{N-1}\{p_i\}} \quad (5)$$

where

$$k = \begin{cases} 1, & \arg Max_{i=2}^{N-1}\{p_i\} > \arg Min_{i=2}^{N-1}\{p_i\} \\ -1, & \arg Max_{i=2}^{N-1}\{p_i\} \leq \arg Min_{i=2}^{N-1}\{p_i\} \end{cases} \quad (6)$$

The first and the last segment of the pitch profile are not taken into consideration to reduce the possibility of error. The polarity of I_R indicates rise or drop in pitch level, magnitude gives the degree of such rise and drop. The initial and final pitch value is normalized to reduce the speaker dependencies using the following equations:

$$P_I = P_I/P_S \quad (7)$$

$$P_F = P_F/P_S \quad (8)$$

P_S is the intrinsic pitch value of a particular speaker.

iv. *F0 at 25% and 75%*

Additionally, F0 values are extracted at the intervals of 25% and 75% of the pitch contour. It gives insights about the change in pitch after one-fourth of onset and before one-fourth of offset.

3.3 Statistical Methods for Feature Selection

3.3.1 Multivariate analysis of variance (MANOVA)

MANOVA provides a robust framework for analysing the effects of independent variables on multiple dependent variables simultaneously. MANOVA tests the differences in the multivariate mean (centroid) of the dependent variable vectors across groups defined by the independent variable. It helps in identifying whether the mean differences among the groups or combinations of the dependent variables have statistical significance. The MANOVA is based on the following assumptions [24,25,26]

1. Multivariate Normality: The dependent variables follow a multivariate normal distribution within each group.
2. Homogeneity: The variance-covariance matrices of the dependent variables are equal across groups.

3. Independence: Observations are independent.

Therefore, before conducting a Multivariate Analysis of Variance (MANOVA), it is essential to check whether the assumptions hold. The validity of the results is established only if the assumptions hold. The MANOVA model can be expressed as follows [27]:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (9)$$

Where $Y_{ij}, 1 \leq i \leq k, 1 \leq j \leq p$, is the observation vector. Here, *k* is the number of groups, and *p* is the number of individual subjects within each group. μ is the overall mean of the dependent variables. τ_i is group effect of the i^{th} group, which is the deviation of the group mean vector from the overall mean vector given by

$$\tau_i = \mu_i - \mu \quad (10)$$

where μ_i is the mean of the dependent variable for the i^{th} vector. ε_{ij} is the vector of error terms for the j^{th} subject in the i^{th} group. Therefore, the eq (1) can be written as

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad (11)$$

In MANOVA, the null hypothesis is that mean vectors are the same across all groups, whereas the alternative hypothesis is that at least one mean vector is different. Several statistics have been used to determine the significance of the group differences. Wilks' Lambda (Λ) is a test static used in MANOVA to assess whether there is any significant difference in the mean vector of the dependent variable across groups. It is given by [28]:

$$\Lambda = \frac{|E|}{|E + H|} \quad (12)$$

Where E is the error matrix that represents the variation within the group, it is computed by summing the squared deviation of the individual observation from the group mean. H is the hypothesis matrix that represents the deviation between groups. It is computed by summing the squared of the deviations of the group means from the overall mean. The Wilks' Lambda is obtained by dividing the determinate of the error matrix by the determinate of the sum of the Error matrix and the Hypothesis matrix. The Wikis' Lambda takes the value from 0 to 1. A value close to 0 means the mean vectors are significantly different across the groups, and the independent variable strongly influences the dependent variables. A Wikis' Lambda value close to 1 indicates that the group means are not significantly different; that is, the independent variable has very little effect on the dependent variables.

Pillai's trace is a multivariate test statistic used in the MANOVA to test the hypothesis that mean vectors of dependent variables across different groups are equal. It is one of the most robust test static, especially when the MANOVA assumptions are significantly violated. Mathematically, Pillai's trace is defined as [29]:

$$V = \sum_{i=1}^{S} \frac{\lambda_i}{1 + \lambda_i} \quad (13)$$

Where $\lambda_i$are the eigenvalues of the matrix $E^{-1}H$. S is the number of significant canonical correlations or eigenvalues, which is the smallest of the number of dependent variables or the number of groups minus 1. Pillai's trace takes the value from 0 to S, which is the smallest of the number of dependent variables or a number of groups minus one. A larger Pillai's trace indicates that the group mean vectors are significantly different from each other; that is, the independent variable has a strong influence on the dependent variables, whereas a Pillai's trace close to 0 indicates that the influence of the independent variable on the dependent variables is insignificant.
Hotelling's trace, also called Hotelling-Lawley trace is another statistic used in MANOVA to test the hypothesis. It is particularly useful in assessing the overall multivariate effect of the independent variable over the dependent variables. The Hotelling trace is mathematically represented as follows [30]:

$$T^2 = \sum_{i=1}^{n} \lambda_i \qquad (14)$$

Where λ_i are the eigenvalues of the matrix $E^{-1}H$ and *n* is the number of dependent variables. The Hotelling's trace takes the value from 0 to ∞. A large Hotelling's trace value indicates a stronger multivalued influence of the independent variable on the dependent variables.
Roy's Largest Root is a test statistic used in MANOVA to determine whether there is any significant difference among the group mean vectors of the dependent variables across different groups. Roy's Largest Root is particularly sensitive to the largest eigenvalue of the matrix formed by the ratio of the hypothesis matrix and the error matrix. Roy's Largest Root is defined as [32]:

$$\theta = max(\lambda_1, \lambda_2, \ldots, \lambda_n) \qquad (15)$$

Where λ_i are the eigenvalues of the matrix $E^{-1}H$ and *n* is the number of dependent variables. Roy's Largest Root takes a value from 0 to ∞. A larger value of the statistic suggests that the group means are significantly different across different groups and the independent variable has a stronger influence on the dependent variables.

The MANOVA results indicate whether there are any significant differences in the combined dependent variables between the two groups. To reveal which dependent variables contribute to these differences, we have conducted t-test and Mutual Information test.

3.3.2 Post Hoc Test

MANOVA provides information regarding the significant difference in at least one group mean. However, it does not specify which group is different. Subsequently, a Post Hoc test is carried out following MANOVA, particularly when it indicates the presence of such difference, to identify the specific groups that exhibit significant differences among themselves. In the present study, Tukey's Honestly Significant Difference (HSD) post hoc test was conducted to calculate the minimum difference between group means using the following formula [24]:

$$HSB = q_{A,\alpha}\sqrt{\frac{MS_{within}}{S}} \qquad (16)$$

Where $q_{A,\alpha}$ is the critical value from the studentized range of distribution, which depends on the significance level (α), the number of groups A and degree of freedom N-A, where N is the total number of observations. *S* is the number of observations in each group. The means square within the group is computed using the formula [24]:

$$MS_{within} = \frac{SS_{within}}{df_{within}} \quad (17)$$

Where SS_{within} is the sum of squares within the group and df_{within} is the degree of freedom within the group. For each dependent variable, Tukey's HSD performs a pair-wise comparison between all possible pairs of group means and computes a value known as HSD, which is compared to a critical value derived from the Studentized range distribution. It also computes the p-values for each pair-wise comparison. If the p-value is less than the significant level, often represented by α, we conclude that the group means are significantly different.

3.3.3 T-test

The t-test is a statistical test used to determine if there is a significant difference between the means of two groups. There are three different types of t-test, these include :

a) One sample t-test is conducted to compare the mean of a sample with a known reference mean.
b) Independent sample t-test (Two sample t-test) is conducted when we want to compare the means of two independent groups or samples. The objective is to find whether there is any significant difference between these two groups.
c) Paired sample t-test (Dependent T-test) is conducted when we want to compare the means of the same group of dependent variables at two different points of time under to different conditions.

In the present study, we conducted paired sample t-test to identify the most distinguishing features for the discrimination of two Galo tones. The t-test statistic for paired sample t-test is computed as follows [32]:

$$t = \frac{\overline{D}}{S_D/\sqrt{n}} \quad (18)$$

where $\overline{D}$ is the mean of the differences between paired observations, S_D is the standard deviation of the differences *n* and is the number of pairs. A greater t-value indicates a more considerable difference among the means, while a smaller t-value suggests a more subdued distinction between the means. Moreover, the t-value decreases as the dispersion of the mean values increases as greater variability in the data hinders the identification of a significant distinction among the means.

3.3.4 Mutual Information

Mutual Information (MI) quantifies the level of mutual dependence between two variables. It assesses the quantity of information obtained from one variable about the other variable. In contrast to correlation, which evaluates linear relationships, mutual information encompasses both linear and nonlinear dependencies. In mutual information-based feature selection, the target is to identify and retain the features that have the highest dependency with the target variable. Mutual information between the target variable and each feature is computed using the following formula [33]:

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right) \quad (19)$$

Where $p(x,y)$ is the joint probability of feature X and target Y, and p(x) and p(y) are the marginal probability distributions of X and Y, respectively. The Mutual Information takes the value from 0 to E_{min}, where E_{min} is the minimum of the entropy among the feature X and target Y.

4. Experiment and Result

The experiments reported in this paper utilized the Galo Tonal Database, detailed in section 3.1. The database consists of 14,623 tonal words. To extract the pitch contour, the speech signal is divided into overlapping frames of 30 milliseconds duration with one-third overlapping. Hamming window is applied to each frame to minimize the discontinuity at the edges. The pitch value has been extracted from each frame using an autocorrection-based pitch extract algorithm. The pitch value extracted may include zero values for invoiced and silent segments. The cubic Spline Interpolation function has been used to estimate pitch values for the frames where the pitch value was 0, resulting in a continuous pitch contour. A moving average filter is applied to the spline interpolated pitch contour to smooth out the fluctuations while preserving the overall pattern of change.

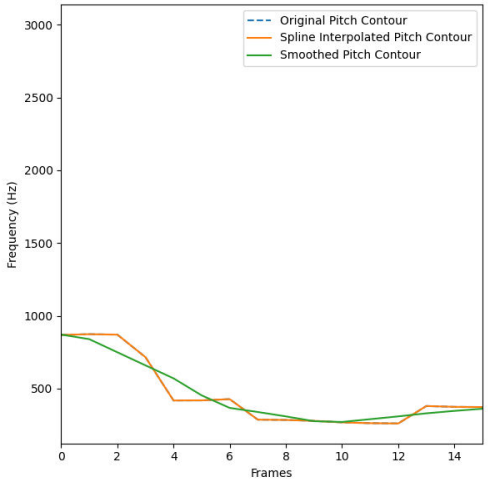


Fig.1 Pitch Contour for the Galo word [*aii*] with Interpolation and Smoothing

The duration, mean F0 (F0-mean), F0 standard deviation (F0-std), F0 median (F0-median), initial pitch (P_I), final pitch (P_F), the pitch at 25% (P_{25}), the pitch at 75% (P_{75}) and Rising Index (I_R) are extracted from the smoothed F0 contour. Further, a second-order polynomial equation has been fitted to the pitch contour, and its coefficients c_0, c_1 and c_2, representing the intercept, rate of change and curvature of the pitch contour, respectively, have been extracted. Thus, we get a 12-dimensional feature vector derived from the pitch contour. In this paper, we have analyzed each feature for their suitability in distinguishing two tones High/Plain and Low/Falling of Galo language.

In the first experiment, we employed a MANOVA test to investigate whether there were any significant differences in the group means of the feature vector (dependent variable) based on tone. The test statistics of MANOVA have been computed using various criteria including Wilks' Lambda, Pillai's trace, Hotelling-Lawley trace and Roy's greatest root. The results of the experiment are shown in table-1

Table. 1: MANOVA Test Statistics for Evaluating Differences in Group Means due to Tone of Galo Language

Test Statistics	**Value**	**Numerator Degree of Freedom**	**Denominator Degree of Freedom**	**F-Value**	**Pr>F**
Wilks' Lambda	0.2648	12	14,611	103.1667	<0.0001
Pillai's trace	0.73521	12	14,611	103.1667	<0.0001
Hotelling-Lawley trace	2.7758	12	14,611	103.1667	<0.0001
Roy's greatest root	2.7758	12	14,611	103.1667	<0.0001

The Wilks' Lambda measures the proportion of dependent variables whose group mean differences are not explained by the independent variable. A smaller Wilks' Lambda value of 0.2648 suggests substantial differences among the groups defined by tone. The F-value of 103.1667 and p-value (Pr>F) less than 0.0001 indicate that these differences are statistically significant.

Pillai's trace, which measures the effect size, indicates the portion of the dependent variables which is affected by the change in the tone. Here, the value of 0.73521 indicates a strong effect of tone on the dependent variables. The associated F-value and p-value indicate that this effect is statistically significant.

Hotelling-Lawley trace, which accumulates the differences between group centroids, has a value of 2.7758. This indicates substantial differences among the groups due to changes in tone. The F-value and p-value further confirm that these differences are statistically significant.

Roy's greatest root, which measures the largest eigenvalue of the hypothesis matrix, has a value of 2.7758, suggesting that the largest dimension of group difference is substantial. The statistical significance of this difference is indicated by the p-value, confirming that there are significant differences in the group due to the change in tone.

All four multivariate test statistics consistently indicate the presence of statistically significant variances among the groups with respect to the tones of Galo Language. A p-value below 0.0001 indicates a high level of statistical significance, implying that tone exerts a substantial influence on the dependent variables.

After finding significant differences among the group means with MANOVA, we conducted Tukey's Honestly Significant Difference (HSD) post-hoc test to find the features that contribute significantly to these differences. The results of the experiment are shown in Table 2.

Table. 2: Tukey's HSD Post-Hoc Comparisons Between High/Level and Low/Rising Tone Groups of Galo Language

Dependent Variable	Mean Difference	p-adj	95% CI Lower	95% CI Upper	Significant (p < .05)
F0-std	19.7302	<0.0001	17.7666	21.6939	Yes
R_I	0.1369	<0.0001	0.1267	0.147	Yes
c_2	-69.8926	<0.0001	-85.3322	-54.453	Yes
P_{75}	36.1785	0.0001	17.8843	54.4728	Yes
P_F	36.4591	0.0001	17.8681	55.0502	Yes
c_0	26.0376	0.0009	10.6549	41.4203	Yes
F0-median	27.1602	0.0033	9.067	45.2534	Yes
Duration	-0.0315	0.0034	-0.0525	-0.0105	Yes
F0-mean	18.4449	0.0419	0.6764	36.2134	Yes
P_I	-15.0597	0.0798	-31.9172	1.7977	No
c_1	1.5461	0.813	-11.2873	14.3795	No
P_{25}	0.6971	0.9374	-16.7243	18.1184	No

The result of Tukey's HSB test indicates a significant difference in the value of the features (dependent variables) across the groups formed by the two Galo tones high/level and low/rising. Significant differences have been observed for the features F0-std, R_I, c_2, P_{75}, P_F, c_0, F0-median, Duration and F0-mean. However, no significant difference has been observed for P_I, c_1 and P_{25}. The High/Level tone group has significantly higher F0-std, R_I, P_{75}, P_F, c_0, F0-median and F0-mean compared to the Low/Rising tone group, whereas c_2 and Duration has significantly higher values for the Low/Rising tone group compared to the High/Level tone group. These findings support our hypothesis that features derived from the pitch contour differ significantly across tones, reflecting the manifestation of distinct phonetic characteristics across each tone.

In order to ascertain the statistical significance of the features in distinguishing between the tones and to demonstrate that the observed alterations are not attributable to random variation, a paired sample t-test was performed on the features extracted from the F0-contour. The results of the experiment are presented in Table 3 and Fig-2, the -log10(p-value) transformation is used to visualize the importance of the top 5 features in discriminating tones in the Galo language.

Table. 3: Summary of Statistical Analysis for Features for Tone Discrimination using T-test for Galo Language

Feature	**t-statistic**	**p-value**	**Significance**
R_I	25.7764	< 0.0001	Highly Significant
F0-std	18.5338	< 0.0001	Highly Significant
c_2	-8.6194	< 0.0001	Highly Significant
c_0	4.5493	< 0.0001	Highly Significant
P_F	3.3413	< 0.01	Significant

P_{75}	3.2685	< 0.01	Significant
P_I	-3.1417	< 0.01	Significant
Duration	-2.9646	< 0.01	Significant
F0-median	2.2387	< 0.05	Marginally Significant
F0-mean	1.2443	≥ 0.05	Not Significant
P_{25}	-0.9560	≥ 0.05	Not Significant
c_1	-0.2563	≥ 0.05	Not Significant

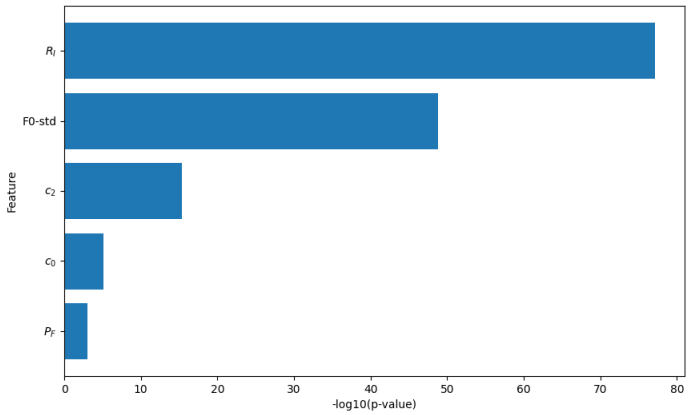


Fig.2: Top 5 Features for Tone Discrimination in Galo Language (T-test Results)

The results of the T-test indicate that the Rising Index (R_I), F0-std, c_2 and c_0 exhibit high significance in distinguishing tones within the Galo language. Similarly, features like P_F, P_{75}, P_I and Duration also hold considerable importance in tone differentiation. However, features such as F0-mean, P_{25}, and c_1 have less significance and possess less relevance within this context of tone discrimination. These findings establish a robust groundwork for the development of more precise and effective models for tone identification, linguistic exploration and potential utilization in speech technology for the Galo language.

The Mutual Information (MI) test, which quantifies the degree of relationship between individual features and the target variable, has been performed to identify the pertinent features for tone discrimination. The experimental results can be found in table 4, while the visualization of the top 5 features exhibiting higher MI values is presented in Fig-3.

Table 4: Summary of Mutual Information Test Results for Tonal Discrimination in Galo Language

Feature	**MI Score**
R_I	0.4602
F0-std	0.3737
c_2	0.1732
P_{75}	0.1296
c_0	0.0991

P_F	0.0865
P_I	0.0725
P_{25}	0.0585
F0-median	0.0530
F0-mean	0.0514
c_1	0.04187
Duration	0.0201

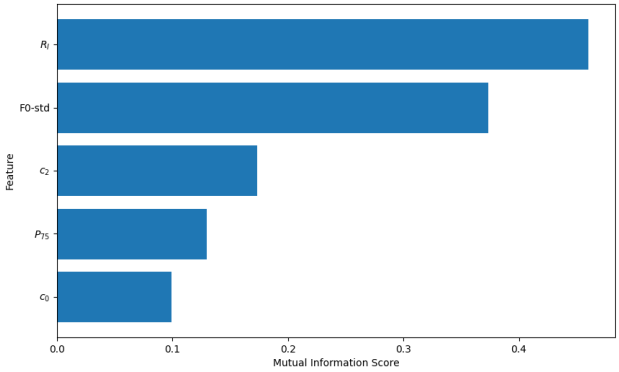


Fig. 3: Top 5 Most Significant Features for Tonal Discrimination Identified by Mutual Information Analysis

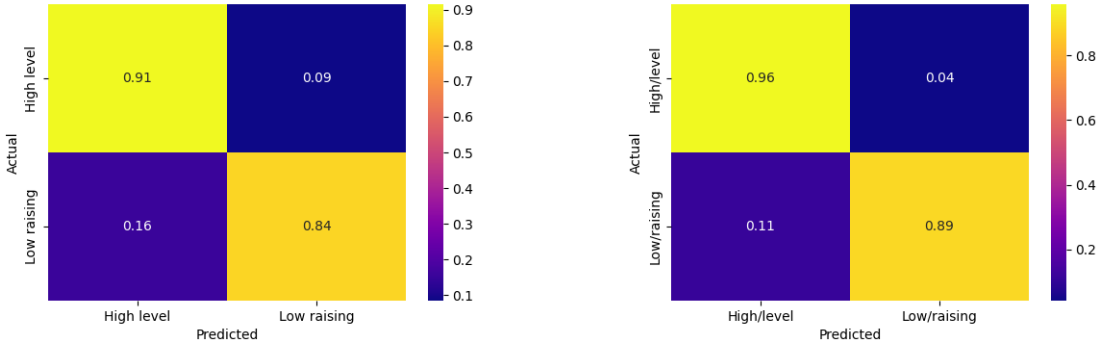
The results of the MI test clearly delineate that R_I and F0-std as the most prominent features for Galo tone discrimination. These features exhibit the highest level of reliance on tonal discrimination, thus proving to be essential for understanding and distinguishing various tones. Features like c_2 and P_{75} also contribute moderately, whereas others demonstrate limited or minimal significance. This observation establishes a distinct hierarchy of feature importance, offering a baseline for further exploration and practical implementations in tone recognition within the Galo language.

In the above experiments, it has been observed that the Rising Index (R_I), Standard Deviation of F0 (F0-std) and curvature of the pitch contour extracted by second-order polynomial fitting (c_2) consistently occupy the first three positions across all tests. These were followed by the pitch at 75% (P_{75}), final pitch (P_F) and intercept coefficient (c_0) of the curve fitting polynomial. It is consistent across all tests. Therefore, these six features have been selected for further analysis of the Galo tone.

Initially, the SVM was trained and tested using all 12 features derived from the pitch contour and its performance was evaluated. In the next step, the SVM was trained and tested using only the selected features. The performance of the SVM classifier has been evaluated using the metrics accuracy, precision, recall and F1-score, both before and after the application of the feature selection method. The results of the experiments, both before and after the application of the feature selection method, are presented in Table 5, with the corresponding confusion matrices in Fig 4.

Table 5: Performance metrics of SVM classifier

Metric	**Before Feature Selection (in %)**	**After Feature Selection (in %)**
Accuracy	88.04	92.39
Precision	88.18	92.56
Recall	88.04	92.39
F1-score	88.02	92.37



(a) Before feature selection (b) After feature selection

Fig 4: Confusion matrix for the SVM Classifier

The SVM-based recognizer effectively validated the efficiency of the feature selection method. It has been observed that after feature selection, the accuracy of the SVM-based tone recognizer increased by 4.35% under the same operational conditions. This validation using an SVM-based method highlights the strength of the feature selection method and its applicability in further linguistic analysis.

5. Conclusion

In this study, we explored the effectiveness of F0-contour based features for Galo tone recognition. The analysis focused on statistical methods such as MANOVA with post hoc test Tukey's HSD, t-tests, and Mutual Information to identify the most significant features contributing to tone differentiation. The features identified as most significant were rising index (R_I), F0 standard deviation (F0-std), the pitch value at 75% (P_{75}), final pitch (P_F) and intercept and curvature features extracted by fitting the F0-contour into a second-order polynomial, namely c_0 and c_2.

The performance of the features are further ascertained using Support Vector Machine (SVM) based Galo tone recognizer. The results demonstrated a 4.35% improvement in tone recognition accuracy when reduced feature set is used.

This paper underscores the important role of feature selection in tone detection, offering a robust foundation for future research in the domains of speech technology, computational phonetics and linguistics with a specific focus on Galo speech recognition in particular and tonal speech recognition in general. By using advanced statistical methods and machine learning algorithms, it is possible to attain greater precision and reliability in tone classification, thereby enriching the overall understanding and documentation of low resource tonal languages like Galo language of Arunachal Pradesh.

References

1. Huang, Hao, and Z. H. U. Jie. "Discriminative tonal feature extraction method in mandarin speech recognition." *The Journal of China Universities of Posts and Telecommunications* 14.4 (2007): 126-130.
2. Sun, Jackson T-S. "Tani languages." *The Sino-Tibetan Languages* (2003): 456-66.
3. Abraham, P. Arunachal Languages with Special Reference to Tani. Tribal Studies-Emerging Frontiers of Knowlege. 2007:181.
4. Post, Mark William. *A grammar of Galo*. Diss. La Trobe University, 2007.
5. Rwbaa, Toomoo, et al. "Galo-English Dictionary", Galo Welfare Society, 2009.
6. Post, Mark W. "Tones in Tani languages: A fieldworker's guide." Northeast Indian Linguistic Society Eighth International Conference in Guwahati, Assam, India, 2014.
7. Cheng, C-C and B. R. U. C. E. Sherwood. "Technical aspects of computer-assisted instruction in Chinese." *The Tsing Hua Journal of Chinese Studies* 35 (1982).
8. Chang, K. C., & Yang, C. C. (1986). A real-time pitch extraction and four-tone recognition system for mandarin speech. Journal of the Chinese Institute of Engineers, 9(1), 37-49.
9. Hsu, H. T. *Tone recognition of Mandarin word speech*. Diss. Masters thesis, Dept. of Electrical Engineering at Tsing Hua University, 1985.
10. Chen, Xi-Xian, et al. "A hidden Markov model applied to Chinese four-tone recognition." *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 12. IEEE, 1987.
11. Chang, P-C., S-W. Sun, and S-H. Chen. "Mandarin tone recognition by multi-layer perceptron." *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1990.
12. Lee, Tan, et al. "An NN based tone classifier for Cantonese." *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*. Vol. 1. IEEE, 1993.
13. Yang, W-J., et al. "Hidden Markov model for Mandarin lexical tone recognition." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36.7 (1988): 988-992.
14. Fu, Qian-Jie, et al. "Importance of tonal envelope cues in Chinese speech recognition." *The Journal of the Acoustical Society of America* 104.1 (1998): 505-510.
15. Chen, C.J., et al. New methods in continuous Mandarin speech recognition. In Proc. EUROSPEECH, Rhodos, 1997, pp. 1543-1546.
16. Lin, C.H., et al. Frameworks for recognition of Mandarin syllables with tones using sub-syllabic units. In Speech Communication, 18 (1996), Elsevier Science B.V., pp. 175-190.
17. Wang, Z.Y., et al. Methods towards the very large vocabulary Chinese speech recognition. In Proc. EUROSPEECH, Madrid, 1995, pp. 215-216.
18. Wang, C. A study of tones and tempo in continuous Mandarin digit strings and their application in telephone quality speech recognition. In Proc. fCSLP98, Vol. 3, Sydney, 1998, pp. 695-698.
19. Sulaiman, M. A. and J. Labadin, "Feature selection based on mutual information," *2015 9th International Conference on IT in Asia (CITA)*, Sarawak, Malaysia, 2015, pp.
20. Hoque, N. D.K. Bhattacharyya, J.K. Kalita, MIFS-ND: A mutual information-based feature selection method, Expert Systems with Applications, Volume 41, Issue 14, 2014,
Pages 6371-6385.
21. Hegde, S., K. K. Achary, and Surendra Shetty. "Feature selection using Fisher's ratio technique for automatic speech recognition." *arXiv preprint arXiv:1505.03239* (2015).
22. Tupper, P, Leung, K., Y. Wang, A. Jongman & J. A. Sereno. Characterizing the distinctive acoustic cues of Mandarin tones. The Journal of the Acoustical Society of America, 147(4), 2000, 2570-2580.

23. Tan Lee, P. C. Ching, L. W. Chan and B. Mak, "An NN based tone classifier for Cantonese," *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, Nagoya, Japan, 1993, pp. 287-290 vol.1, doi: 10.1109/IJCNN.1993.713914.
24. Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE Publications.
25. Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis (6th ed.). Pearson Education.
26. Rencher, A. C. (2002). Methods of multivariate analysis (2nd ed.). Wiley.
27. Everitt, B., & Hothorn, T. (2011). An introduction to applied multivariate analysis with R. Springer.
28. Warne, R. T. (2014). Statistics for the social sciences: A general linear model approach. Cambridge University Press.
29. Anderson, T. W. (2003). An introduction to multivariate statistical analysis (3rd ed.). Wiley.
30. Mertler, C. A., & Vannatta, R. A. (2013). Advanced and multivariate statistical methods (5th ed.). Routledge.
31. Huberty, C. J., & Olejnik, S. (2006). Applied MANOVA and discriminant analysis (2nd ed.). Wiley.
32. Kalpić, D., Hlupić, N., & Lovrić, M. (2011). Student's t-Tests. In M. Lovric (Ed.), International Encyclopedia of Statistical Science (pp. 1569-1571). Springer.
33. Walters-Williams, Janett, and Yan Li. "Estimation of mutual information: A survey." *Rough Sets and Knowledge Technology: 4th International Conference, RSKT 2009, Gold Coast, Australia, July 14-16, 2009. Proceedings 4*. Springer Berlin Heidelberg, 2009.