International Journal of Applied Engineering & Technology

STUDY OF MISHING LANGUAGE VOWELS USING FUNDAMENTAL FREQUENCY (PITCH) AND FORMANTS

S.K. Saikia, D.J. Borah and S. Kalita
Department of Computer Application, Mahapurusha Srimanta SankaradevaViswavidyalaya, Nagaon, Assam, India

ABSTRACT
Human speech, a remarkable feat of biological and linguistic complexity, relies heavily on theinterplay of various parameters. Analyzing two key acoustic features, fundamental frequency (pitch) and formants, plays a crucial role in understanding and interpreting spoken language. In this paper an attempt has been made to study the fundamental frequency (pitch) and formants of Mishing phonemes (vowels). Analysis of the Fourteen vowel speech samples in Mishing language is performed using fundamental frequency (F0) and formants (F1, F2, F3…). This study will help recognition of speech and speaker of Mishing Language, the language spoken by second largest tribes of Assam, India.

Keyword: Mishing Language, Fundamental Frequency, Formant, Speech Recognition

1. INTRODUCTION
Analysis and presentation of the speech signal in the frequency domain are of the great importance in studying the nature of speech signal and its acoustic properties [1]. Human speech, relies heavily on the precise coordination of various anatomical structures known as the vocal organs. These coordinated movements shape the airstream emanating from the lungs, giving rise to a spectrum of acoustic signals categorized as either voiced or unvoiced. Vowel sounds, the very essence of spoken language, fall definitively within the voiced category. Their distinctive characteristics arise from the concerted interplay of pitch and formants. Therefore, analyzing pitch and formants offers a window into the complex world of human speech. Formants are very important parameter to represent a vowel signal [2]. The vowels can be categorized by the temporal development of the formants. Each of the preferred resonating frequencies of the vocal tract (corresponding tothe relevant bump in the frequency response curve) is known as a formant. These are usually referred to as F1 indicating the first formant, F2 indicating the second formant, F3 indicating the third formant, etc.[3]. That is, by moving around the tongue body and the lips, the positionof the formants can be changed[4]. To achieve naturalness in human-machine oral communication, it is imperative to incorporate the full spectrum of speech characteristics. Automatic speech recognition (ASR) systems initiate their analysis with precisely this goal inmind. As fundamental pattern recognizers, the accuracy of ASR systems hinges heavily on their ability to effectively determine formant frequencies. These formants, crucial acoustic features, play a vital role in both speech recognition and speaker identification. While human listeners primarily rely on the overall formant pattern for signal interpretation, ASR systems require precise measurement of individual formant frequencies. Therefore, optimizing formant frequency determination algorithms remains an essential aspect in the ongoing pursuit of natural and robust human-machine speech interaction. The opening and closing of the vocal folds that occur during speaking break the air steam into chains of pulses. The rate of repetition of these pulses is the pitch and it defines the fundamental frequency of the speech signal [5]. In other words, the rate of vibrations of the vocal folds is the fundamental frequency of the voice. The frequency increases when the vocal folds are made taut. Relative differences in the fundamental frequency of the voice are utilized in all languages to study thevarious aspects of linguistic information conveyed by it [6]. Given a segment of a signal, the fundamental frequency estimation problem aims to identify the dominant repetitive frequency within that segment.

The Mishing language has a phonological system of twenty-nine phonemes, fifteen of whichare consonants and fourteen vowels [7]. Mishing language, in absence of its own script uses the Roman script for its lexicographical determinants. Therefore, there is a difference between the spoken form and the written form [8]. The vowels are represented in Table-1.

Vowels	Place of Articulation	Vowels	Place of Articulation
/o/	Back half open	/u:/	Back Close
/o:/	Back half open	/e/	Front half open
/a/	Central open	/e:/	Front half open
/a:/	Central open	/é/	Central half open
/i/	Front Close	/é:/	Central half open
/i:/	Front Close	/í/	Central Close
/u/	Back Close	/í:/	Central Close

Table-1: Mishing Vowels

The vowels are two categories- short vowels (Gomug Mukdeng in Mishing language) and long vowels (Gomug Mukyar in Mishing language). short vowels are /o/, /a/, /i/, /u/, /e/, /é/ and /í/. Long vowels are /o:/, /a:/, /i:/, /u:/, /e:/, /é:/, /í:/

2. DATA AND METHODOLOGY

Fundamental frequency (F0) and formants are both important aspects of sound. Fundamental frequency (F0), also known as pitch, plays a crucial role in various aspects of speech analysis,ranging from basic speech recognition to complex emotion detection. Formants, influenced by F0, shape vowel sounds. Accurate F0 determination helps differentiate vowels, critical for speech recognition accuracy. Individual voices have unique F0 ranges. Analysing F0 patterns helps identify speakers, especially in speaker verification systems. Formants and F0 are not completely independent. For example, smaller vocal tracts generally have higher F0 and higher formants. F0 can be affected by formants to some extent, especially when formants are close to harmonics of F0. Both F0 and formants are used in speech synthesisand recognition technologies.

The general problem of fundamental frequency estimation is to take a portion of signal and to find the dominant frequency of repetition. Thus, the difficulties that arises in the estimation of fundamental frequency are (i) all signals are not periodic, (ii) those are periodic may be changing in fundamental frequency over the time of interest, (iii) signals may be contaminatedwith noise, even with periodic signals of other fundamental frequencies, (iv) signals which areperiodic with interval T are also periodic with interval 2T, 3T etc., so we need to find the smallest periodic interval or the highest fundamental frequency, and (v) even signals of constant fundamental frequency may be changing in other ways over the interval of interest. [9].

The estimation of pitch, also known as fundamental frequency (F0), of vowel sounds primarily employs two distinct approaches: Fast Fourier Transform (FFT) and Autocorrelation. In this study we are using autocorrelation method. Calculating the pitch (F0) using autocorrelation involves some mathematical steps, but it's conceptually simpler than using FFT.

Calculate Autocorrelation: For a discrete signal x, the autocorrelation at lag τ is defined as,

$R(\tau) = \Sigma(x(n) * x(n + \tau)) / N$ where n iterates through all sample indices (0 to N-1), τ is the lag(time shift) between samples, N is the total number of samples. This essentially compares the signal with itself at different time shifts, revealing periodicities.

Locate Peak: The autocorrelation function $R(\tau)$ will typically have a strong peak at a lag corresponding to the period of the fundamental frequency. Identify the lag τ_peak corresponding to the highest absolute value of $R(\tau)$.

Convert Lag to Pitch: The fundamental frequency is the inverse of the peak lag,

$F0 = fs / \tau$_peak where fs is the sampling frequency in Hz.

To calculate the Formant, we are using Linear Predictive Coding (LPC) method. The Linear Prediction Coefficients (LPC) method is a popular approach for estimating formants in speechsignals. The linear predictive model is based on a mathematical approximation of the vocal tract [10].

First the speech signal is segmented into short frames (typically 20-30ms) and windowed (e.g.,Hamming window) to reduce spectral leakage. Then the autocorrelation function (ACF) of thewindowed frame is calculated. The ACF measures the correlation of the signal with itself at different time lags. The LPC coefficients are obtained, these coefficients represent the prediction filter thatminimizes the mean squared error between the actual signal and its predicted version based on past samples. The roots of the LPC polynomial (obtained from the coefficients) are calculated. These roots correspond to the poles of the prediction filter, which are located within the unit circle in the z-plane. Only the roots with positive imaginary parts and magnitudes less than 1 are considered formants. The formants are then converted from z-plane frequencies to actual frequencies in Hz using the following formula,

fn = (fs / 2 * pi) * arctangent(imag(zn) / real(zn))

where fn is the frequency of the nth formant, fs is the sampling frequency, zn is the nth rootof the LPC polynomial.

The study was performed by recording the 14 vowels spoken by 6 male and 4 female speakers (age group 23-54 years). All speakers are Mishing people and they speak Mishing as their native language. All of them are educated and in teaching profession. Recording was done using a Sony digital recorder in a quite environment to minimise the ambient noises. Sample rate taken 44100 Hz, 16 bit wav format. Analysis was done using Matlab R2019a with audio toolbox, PRAAT 6.1.51 and Google Colab.

3. RESULTS AND DISCUSSION:

The mean fundamental frequencies of Mishing vowels uttered by 6 male and 4 female speakers are estimated in this study and listed in Table2.

Vowels	Mean Fundamental Frequency (F0) in Hz	
	Male	Female
/o/	201.93	323.82
/o:/	102.19	310.52
/a/	196.06	238.37
/a:/	110.84	190.57
/i/	174.21	258.75
/i:/	170.49	205.64
/u/	185.75	290.63
/u:/	163.81	210.35
/e/	174.77	255.76
/e:/	140.50	205.68
/é/	192.04	269.37
/é:/	162.95	174.59
/í/	197.00	273.57
/í:/	105.39	189.56

Table-2: Mean Fundamental Frequency (F0) of Mishing vowels

An examination of Table 2 reveals that the values of pitch, also known as fundamental frequency (F0), for female informants consistently exceed those of male informants. Typically, the pitch or fundamental frequency ranges from 80Hz to 160Hz for male speakers and from 140Hz to400Hz for female speakers [11]. In adult, generally the length of vocal folds in male is more than thatof female counterpart [12]. This observation aligns with the findings presented by many previous studies.

Figure-1 Shows visualization of Mishing Male and Female Vowel Utterances throughFundamental Frequency.

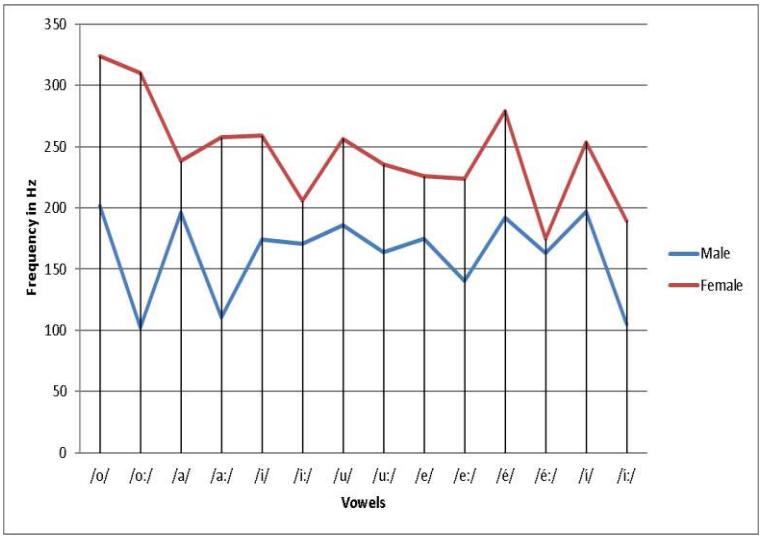

Figure-1: Visualization of Mishing Male and Female Vowel

The estimated result of mean formant frequency of fourteen vowels of Mishing language (6 nos. of Male Speaker) are listed in table 3.

Formant Frequency / Vowels	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)
/o/	225.99	1171.78	3203.92	4374.08
/o:/	529.30	865.92	2557.47	4315.46
/a/	1094.51	1593.54	3371.33	4266.60
/a:/	910.63	1251.44	2782.65	5312.80
/i/	268.68	2494.61	2994.13	4247.88
/i:/	317.39	2553.60	3013.05	4156.06
/u/	311.41	853.08	2506.89	4118.97
/u:/	437.00	896.00	2527.60	3980.49
/e/	520.86	1849.29	2627.11	4285.34
/e:/	483.26	2042.05	2759.16	3601.94
/é/	278.93	1524.22	2640.08	4470.41
/é:/	530.88	1227.93	2755.67	3856.90
/í/	1179.55	1716.73	2735.48	3900.38
/í:/	314.04	1306.57	2656.06	3837.13

Table-3: Estimated result of mean formant frequency of fourteen vowels

The formants are calculated upto F4. These estimated results yield the following observation, For F1, front close articulations /i/ and /o:/, central half open articulation /é/ , back close articulation /u/, back half open articulation /o/ have higher formant then their corresponding long vowels (Gomuk Mukyar in Mishing language) due to their short duration.

For F2, all central articulation short vowels (/í/,/é/,/a/) having higher formant values then their long vowels (/í:/, /é:/, /a:/). All front articulation short vowels (/i/, /e/) having lower formant values then their long vowels (/i:/, /e:/).

For F3, all front articulation short vowels (/i/, /e/) having lower formant values then their longvowels (/i:/, /e:/).

For F4, except central open articulation all other articulation having higher formant values forshort vowels.

Figure 2 illustrates the waveform and fundamental frequency (pitch) of Mishing Vowelsample /o/ (male speaker) and /e/ (female speaker) respectively. The blue line represents the

Pitch of the Vowel Utterance.

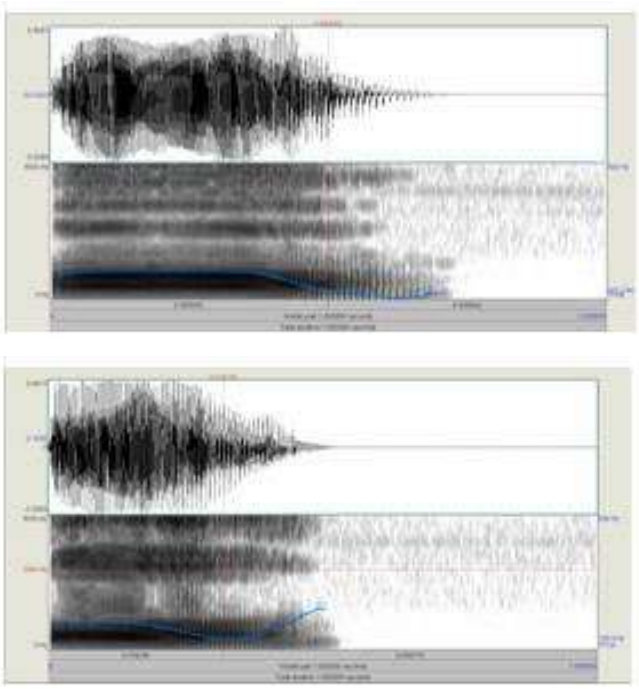

Figure 2: Fundamental frequency (pitch) of Mishing Vowelsample /o/ (male speaker) and /e/ (female speaker)

Figure 3 illustrates the formant of Mishing Vowel sample /é/ (male speaker) and /í/ (female speaker). The figures having 5 red dotted lines that represent formant F1 to F5 (bottom to top).

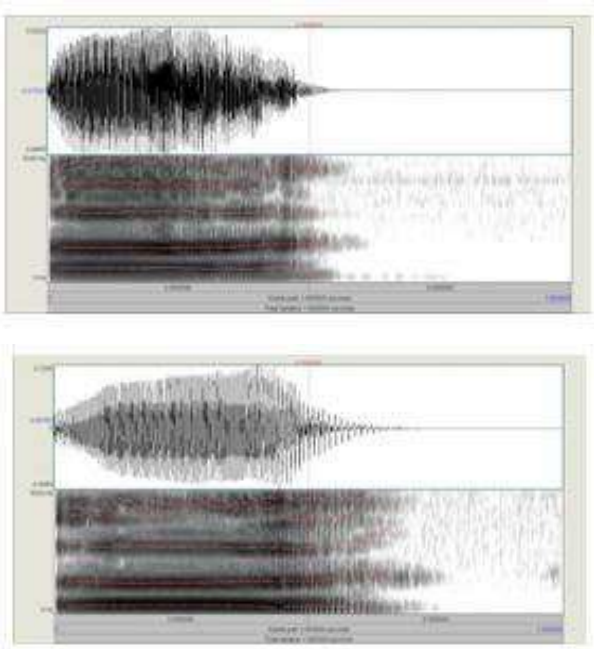

Figure 3: Formant of Mishing Vowel sample /é/ (male speaker) and /í/ (femalespeaker).

4. CONCLUSION

This paper proposes a novel method for analysis of Mishing language vowels, relying solely on fundamental frequency (F0) and formant analysis (F1, F2, F3, F4) as the foundation for its investigation. The exploration of fundamental frequency and formants has provided valuable insights into the intricate mechanisms of speech production and perception. F0, representing perceived pitch, carries emotional cues and speaker identity information, while formants act as acoustic fingerprints of vowels, shaping the sounds we hear. While significant progress has been made, several avenues remain for future exploration. Integrating advanced machine learning techniques with acoustic analysis holds promise for even more robust and nuanced understanding of speech. Further research on cross-linguistic variations and the interaction between acoustic parameters and other linguistic features can provide deeper insights into the complexities of human communication. In conclusion, the analysis of fundamental frequency and formants remains a important part of speech recognition, paving the way for advancements in various fields related to language, communication, and technology.

REFERENCES

[1] Prica, B., and Sinisallic. (2010). Recognition of Vowels in Continuous Speech by Using Formants. Facta Universitatis (NIS), SER. ELEC. ENERG. vol. 23, no. 3, 379-393.

[2] Ghosh, T., Saha, S., and Ferdous, A. H. M. Iftekharul. (2016). Formant Analysis of Bangla Vowel for Automatic Speech Recognition. Signal & Image Processing (SIPIJ), Vol.7, No.5. DOI 10.5121/sipij.2016.7501.

[3] Alotaibi, Y., and Hussain, A. (2010). Comparative Analysis of Arabic Vowels using Formants and an Automatic Speech Recognition System. International Journal of Signal Processing, Image Processing and Pattern Recognition, 3,11-22.

[4] Deller, J., John R., Proakis, J., and Hansen, J. H. (1993). Discrete Time Processing of Speech Signal. Prentice Hall PTR.

[5] Zhao, W.W., and Ogunfunmi, T. (1999). Formant and Pitch Detection Using Time-Frequency Distribution. International Journal of Speech Technology 3, 35–49. DOI 10.1023/A :1009626826626

[6] Choudhury. S, A statistical analysis of speaker dependent independent pattern congruity of assamese and bodo phonemes. Ph.D. Thesis, Gauhati University, India.

[7] Taid, T. R. (2010). Mising Gompir Kumsung. Anundoram Borooah Institute of Language, Art and Culture, Assam. ISBN 978-81-910016-0-0

[8] Rehman, R. and Hazarika, G. C. (2014). Analysis and Recognition of Vowels in SHAI`YANG MIRI Language using Formants. International Journal of Computer Applications (USA), Volume 89 / Number 2, 89. 10.5120/15472-4155.

[9] Deka, M. K., Kalita, S.K., and Sarma S. K. (2010). A Comparative Study for Identification of Sex of the Speaker With Reference to Bodo Vowels. Int. J. Open Problems Compt. Math., Vol. 3, No. 5, ISSN 1998-6262.

[10] Sarma, M., and Sarma, K. K. (2012). Formant frequency estimation of phonemes of Assamese speech. 2nd National Conference on Computational Intelligence and Signal Processing (CISP), Guwahati, India, 119-125, DOI: 10.1109/NCCISP.2012.6189691

[11] Rabiner, L. R., Levinson, S.E., Rosenberg, A.E., and Wilpon, J.G. (1979). IEEE Trans Acoustics, Speech, Signal Proc, ASSP-27, 336.

[12] Boro, J. (2015). Fundamental Frequency Analysis of Bodo Vowels. International Journal of Engineering and Technical Research, Volume-3, Issue-9, 2454-4698, ISSN 2321-0869.