

Original Article

Towards Robust Identification of Age-Sensitive Acoustic Features for Developmental Speech Analysis in Low Resource Bodo Language

Amkar Brahma¹, Manoj Kumar Deka², Pranchis Narzaree^{1,3}, Bijay Kumar Singh⁴

¹Computer Science and Technology, Bodoland University, Assam, India.

²Department of Information Technology, School of Technology, Assam Skill University, Assam, India.

³Computer Science, Kokrajhar University, Assam, India.

⁴Department of Computer Science, Assam University, Assam, India.

¹Corresponding Author : amkarbrahma9773@gmail.com

Received: 10 April 2026

Revised: 17 August 2026

Accepted: 21 August 2026

Published: 30 September 2026

Abstract - Children's speech exhibits wide variation because their vocal tract and articulatory system mature progressively with age, and thus, analysis of children's speech faces greater difficulty when compared to adults' speech. When the language is a low-resource language where only scarce child speech corpora are available and the development of speech has not been extensively explored, it becomes extremely difficult to build any sophisticated child speech technologies. This study investigates age-sensitive acoustic characteristics of child's speech in the low-resource Bodo language using Mel-Frequency Cepstral Coefficients (MFCC) and their dynamic characteristics. The children's speech recordings from various ages groups are analyzed, and the MFCC, mean, MFCC variance, del-MFCC mean, and del-MFCC variance are computed from each of the segments to find age-varying features. Correlation analysis, linear regression and one-way ANOVA were employed to identify the acoustically varying features as per age. From the experiments, it was found that deltamean3, deltamean2, mean3, mean6, and mean13 possess a high association with speech speaker age at all the statistical analyses tests. The developmental profile of selected acoustic features shows that both spectro-temporal characteristics of speech are growing into stability as children progress in their age to demonstrate maturity in the mechanism of speech production. The analysis and research using these age-sensitive acoustic features may provide valuable directions to developmental speech research and help to develop useful child-based speech applications for low-resource languages like Bodo.

Keywords - Child speech, Age-sensitive acoustic features, Low-resource language, Mel-Frequency Cepstral Coefficients (MFCC), Statistical analysis, Speech signal processing.

1. Introduction

Speech is the most common form of human interaction, and the characteristics of human speech continue to evolve during childhood, largely owing to physical maturation and cognitive growth. Children's speech exhibits considerable acoustic variation in comparison to adults' speech because the mechanisms of producing speech, including the vocal tract, articulators, and neural control, are still developing. Due to these changes, acoustic features such as speech fundamental frequency, formant structure, spectral properties and temporal structure of children's speech differ significantly from those of adults. Subsequently, traditional speech processing methods, often designed using adult speech, performed relatively poorly when tested on child speech, particularly on tasks such as automatic speech recognition, speaker age estimation, speech therapy and language assessment. Among the best-known acoustical representations used in speech

processing, Mel-Frequency Cepstral Coefficients (MFCCs) represent the spectrum and approximate the sound signal spectrum of the human ears. Features representing dynamic temporal information of speech, called delta features (e.g., deltas), are in general derived from static MFCC features and provide some additional information about the dynamic property of speech production.

A comprehensive investigation of combined acoustic and prosodic features is important for capturing developmental variations in children's speech and for characterizing acoustic measures which co vary with child age. Child speech is characterized by substantial variations in spectral and temporal acoustic and phonological features that change during the maturational period. Many of the above studies also found that age-based speech modeling helps to improve speech recognition rates on child speech, as a child's speech



can be discriminated by a child's speaking voice. Recently, studies attempted to use machine learning methods and combinations of acoustic and prosody features to classifier speakers' age. However, it is observed that the acoustic changes of children, such as development effects, and changes at particular age(s) can provide acoustic clues for determining age, in particular, when a large set of acoustic parameters has been investigated. Unfortunately few research on identifying statistically significant age-sensitive features of child speech is available for low-resource languages due to the scarcity of well-annotated speech corpora.

Bodo language is one of the constitutionally recognized languages of India, and is a representative low-resource speech technology language in the speech technology community. As far as it is known, there is no previous report in the literature about age-specific acoustic variability in children's Bodo-speaking, in the literature which is important to provide the child speech recognition as well as for language-specific speech analysis. Therefore, the development of a suitable Bodo age estimation approach based on features that have a significant relationship with the speaker's age becomes necessary for low-resource child speech processing.

The present work aims to perform a comparative study of MFCC-based spectral and dynamic acoustic features as a measure of identifying age-sensitive features of children in the Bodo speech sample. The statistic mean and variation of MFCCs along with delta MFCCs will be computed and analyzed against the speaker ages to identifying an age-correlated feature. To detect these features, standard signal analysis statistical methods like correlation coefficients, linear regression and one-way Analysis Of Variance (ANOVA) were applied on the features to quantify and find statistically significance age dependent features. Instead of performing classify based evaluation (e.g., accuracy as a measurement), the goal of current research is to demonstrate statistically how much each acoustic feature is related to the speaker's age.

Major contributions:

- To analyse and determine statistically significant acoustic features, related to age, based on MFCC-based spectral and dynamic features in low resource Bodo language child speech.
- A comprehensive analyse of acoustic features based on correlation, linear regression, and the one-way ANOVA method to investigate statistical relationship.
- Analyse trends of the developmental patterns with respect to important acoustic features of development, and age-specific variability in child speech production.
- Delivers acoustic feature-set for constructing speech recognition systems, systems to estimate the age for children, and for forthcoming child-speech-based systems for low-resource Indian languages.

2. Related Work

2.1. Developmental Characteristics of Children's Speech

The differences observed in the children's speech compared to adults are associated with their growing vocal tracts, developing articulatory organs, and the time-varying development of speech motor control. The earliest findings of Lee et al. [1] demonstrated the systematic variation of acoustic attributes, both temporal and spectral ones, across childhood ages and confirmed that several parameters are progressively closing to those of adults. Also, Yildirim et al. [3] noticed that children of preschool ages have significantly higher F0, higher formants, and more speaker variability than adults because their vocal tracts are not mature enough. The spectral property of fricative consonants becomes more invariant over the course of age, as stated by Nissen and Fox [4].

2.2. Children's Speech Recognition and Acoustic Modelling

The mismatch between children's and adults' speech has long been a serious issue for ASR systems. Potamianos and Narayanan [2] observed that an ASR system built on adult speech provides very poorly results when tested on children's speech, due to age-based acoustic variation and showed that the acoustic modelling adapted to age dramatically improves the performance. In recent years, several machine learning techniques have been examined for the classification of age in children with a number of combinations of spectral, prosodic, and cepstral features [9].

In previous studies number of researcher attempts to study various acoustic feature combinations with various multiple classifiers [5]. Some research attempts focused on the LPCC feature along with the SVM technique for Indian children's speech recognition [6]. Also, a few studies have investigated enhancing the age group classification using the combination of feature selection method ANNs [10].

2.3. Acoustic Feature Analysis for Developmental Speech

Mel-Frequency Cepstral Coefficients (MFCCs) are one of the most common feature sets in speech processing. It is proven that MFCC and its time derivatives could encode spectral information related to the vocal tract, thus it's beneficial in exploring variations in the development of children's speech. It has been observed that formants, pitch, energy and LPCC features could be used well for different applications of speech analysis, such as predicting the age, detecting abnormalities and speaker identification in earlier work, thereby confirming that both static and dynamic acoustic features have significant values for explaining the development in the maturation of the speech system during different developmental stages.

2.4. Cross-Linguistic and Clinical Studies

There have been fewer, though key, cross-linguistic validation studies. Russian data from A. Grigorev et al., from children aged between 5-16, showed language universal

trends of decreasing vowel duration and pitch, with gender differences becoming significant around ages 13 to 16. The statistical approach used here confirmed the patterns not only in English-speaking subjects, but also in Russian [7]. Clinical applications have been established by E. Lyakso et al., who studied the voice characteristics of 300 children with differing developmental disorders between ages 4-16. Their comparative acoustic analysis indicated specific patterns associated with each condition, including significant differences for autism spectrum disorders (high pitch values and increased variation) and Down syndrome (minimal pitch and maximal vowel duration), indicating the potential to be used as diagnostic markers [8].

2.5. Research gap

Within the study of children's speech analysis, there has already been a considerable amount of research on the acoustic correlates for age estimation. There are still, nonetheless, quite a few important issues to explore. Several previous works focused on high-resource languages like English, whereas very little work has been published for low-resource Indian languages. Again, most of the previous studies are geared towards overall classification accuracy rather than finding the acoustical parameters that exhibit statistically significant developmental change. Combined analysis by applying correlation analysis, regression analysis and analysis of variance is used less for studying the relation between the age of the child and acoustical parameters. Finally, studies on the developmental changes in the speech of children in Bodo are still largely understudied due to the shortage of labeled corpora for Bodo.

2.6. Motivation of the Present Study

Driven by these constraints, this paper focuses on the age-related acoustic correlates of Bodo, a low-resource indigenous language of children, by extracting MFCC-based dynamic and spectral features. Unlike existing research, which prioritises class prediction, this study analytically extracts statistically validated correlates of speech development by applying correlational methods, linear regression and One-Way ANOVA. This provides insights into intra-speech development for this group of children and lends itself to effective system development for child speech-processing in low-resources scenario.

3. Materials and Methods

3.1. Acoustic Feature Extraction

In this work, the age-based changes in child speech are analyzed by MFCC-based static and dynamic features. The main reason for using MFCC's is that they capture most of the speech's spectral components, and the MFCC approximation to the critical-band spectral sensitivity of the ear appeared to be quite good. By employing the mean and variance of MFCC's and the mean and variance of delta MFCCs, static and dynamic changes can be represented.

3.1.1. Mel-frequency Cepstral Coefficient (MFCC)

The speech was first pre-emphasised (to emphasize high frequencies and reduce noise in those areas of the spectrum), divided into segments of overlapping 25ms (50% overlap), windowed with a Hamming window (a commonly used windowing function to eliminate or minimise spectral leakage before taking a Fast Fourier Transform) and converted into frequency domain via FFT. This spectrum was then filtered using a series of triangular Mel-scale filters to achieve frequencies that imitate the frequency perception of the human ear. After these were calculated, take the log of the filter banks and finally take the Discrete Cosine Transform of this to produce the MFCC vectors. The Mel-frequency scale was calculated using Equation (1).

$$\text{Mel}(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

3.1.2. Delta Mel-frequency Cepstral Coefficient (Delta MFCC)

However, MFCCs present spectral features of speech in particular intervals, and they do not preserve the changes in speech between adjacent intervals. In order to encode articulatory movement dynamics of speech, Delta MFCC, which are the first-order differential of MFCCs, were computed. These time-varying features contain the articulatory dynamics of speech utterance. The Delta MFCC coefficients were computed using Equation (2).

$$\mathbf{d}_t = \frac{\sum_{n=1}^N \mathbf{n}(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^N n^2} \quad (2)$$

where, $\mathbf{d}_t$ is a delta coefficient from frame t computed in terms of the static coefficients c_{t+n} to c_{t-n} .

3.1.3. MFCC mean and MFCC variance

Let $c_t^{(k)}$ denote the k -th MFCC coefficient extracted from the t -th frame of a speech signal, with $t = 1, 2, \dots, T$. To represent each utterance in a fixed-dimensional space, the first-order statistical measures are computed across time. The mean of the k -th MFCC coefficient is given by

$$\mu_k = \frac{1}{T} \sum_{t=1}^T c_t^{(k)} \quad (3)$$

$$\sigma_k^2 = \frac{1}{T} \sum_{t=1}^T (c_t^{(k)} - \mu_k)^2 \quad (4)$$

3.1.4. Delta MFCC mean and Delta MFCC variance

Delta MFCCs are first-order time derivative values of MFCCs that are computed on the regression basis method and are used to obtain the dynamic characteristics of a speech signal. The delta feature for the k -th MFCC at frame t can be written as

$$\Delta c_t^{(k)} = \frac{\sum_{n=1}^N \mathbf{n}(c_{t+n}^{(k)} - c_{t-n}^{(k)})}{2 \sum_{n=1}^N n^2} \quad (5)$$

$$\overline{\Delta c_t^{(k)}} = \frac{1}{T} \sum_{t=1}^T \Delta c_t^{(k)} \quad (6)$$

$$\sigma_{\Delta, k}^2 = \frac{1}{T} \sum_{t=1}^T (\Delta c_t^{(k)} - \overline{\Delta c^{(k)}})^2 \quad (7)$$

3.2. Analysis of Statistical Features

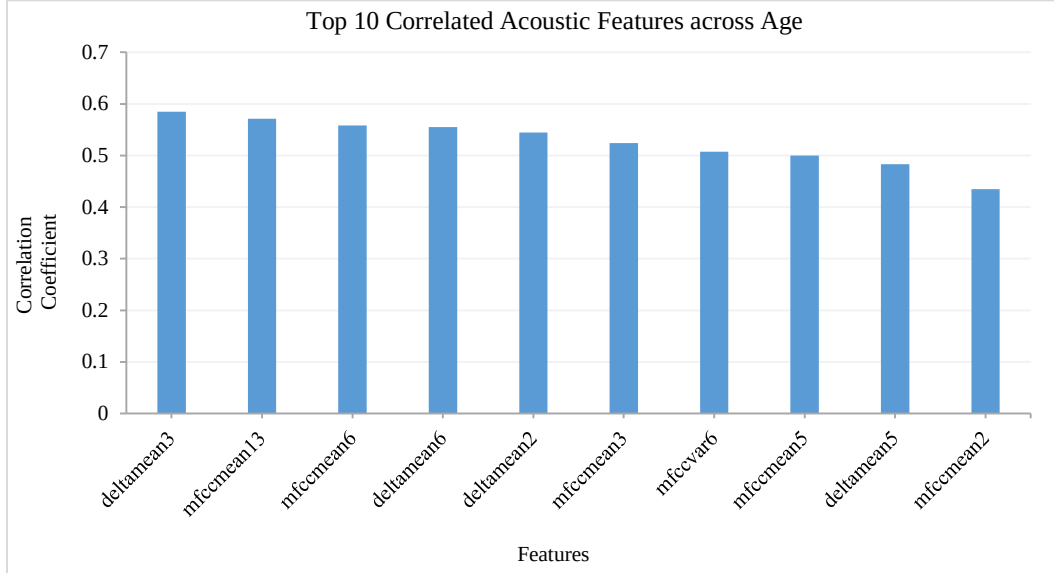


Fig. 1 Lists the ten most correlated acoustic characteristics with speaker's age: 'deltamean3' has the best positive relationship, about 0.58. Another characteristic with high positive correlation, 'mfccmean13' is about 0.57, then 'mfccmean6' with 0.56, followed by 'deltamean2' and 'mfccmean3'. These demonstrate that spectral features and temporal acoustic characteristics tend to follow the same trends

3.2.1. Correlation Analysis

The correlation analysis was carried out to check the correlations of these speech features and age in natural human speech. The stronger the values of the correlation coefficient between the features and age; indicated some relation in the developmental changes with human speech.

3.2.2. Regression Analysis

The regression analysis was applied to analyze which features could contribute to age prediction effectively. The coefficient of determination, R , was taken as the assessment to determine how well a feature can model the information, and for develop of regression analysis, larger values of R indicate that the features are helpful.

$$R^2 = \frac{\sum (y_i^2 - \bar{y}_i)^2}{\sum (y_i^2 - \bar{y})^2} \quad (8)$$

3.2.3. Analysis of Variance (ANOVA)

One-way ANOVA was performed for the average score on each feature based on basis of different ages so that it would be evaluated for any difference using a statistical approach. The F-Statistics and corresponding p values have recorded for all the tests. Features having p value < 0.05 are selected to make up the analysis.

$$\text{Total Sum of Squares}(SST) = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (9)$$

$$\text{Between Group Sum of Squares}(SSB) = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2 \quad (10)$$

Within Group Sum of Squares (SSW) =

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij}^2 - \bar{y}_j)^2 \quad (11)$$

$$MSB = \frac{SSB}{k-1} \quad (12)$$

$$MSW = \frac{SSW}{n-k} \quad (13)$$

$$F = \frac{MSB}{MSW} \quad (14)$$

3.3. Feature Ranking

Correlation and regression and ANOVA indicated the importance and predictability of acoustical feature which leads to their scoring and accordingly the best ranked acoustic features that might indicate changes in child's voice were filtered.

4. Results and Discussion

4.1. Correlation Analysis

Correlation analysis was executed with every individual audio feature to examine association with speaker age, and Pearson correlation coefficients were obtained to describe and analyze the correlations as well as their directions between individual audio features and age.

From the chosen dataset of audio features, those audio features with the highest absolute value of correlation coefficients were considered most significantly reflective of the developmental variations among childhood speakers.

The strong influence of Delta MFCC features within the best features likely means that speech variation is more consistent in the speech signals of the older kids. Children generally tend to exhibit larger articulatory variation because of the immaturity of their speech motor system. This variation decreases in older children as the fine muscular control matures and the speech production processes become more consistent. This enables Delta MFCC features to act as an ideal marker for speech aging. Similarly, many static MFCC features correlate positively with the age. MFCC parameters carry the spectral envelope of speech and also reflect resonatory features of the speech. During development, there are changes in vocal tract size during the growth of kids. This also changes the resonatory features, the MFCC coefficients exhibit a steady variation during growth period.

Our finding agrees with observations in terms of language development in developing children reported in Lee et al. [1] and Potamianos and Narayanan [2]. Our study, which seems to follow these language developments in the Bodo language, which is still considered a low-resource language, is very promising.

4.2. Regression Analysis

We performed a linear regression analysis to evaluate individual acoustic features as predictors of speaker age and report the corresponding coefficient of determination R^2 for each feature, which measures the percentage of variance in speaker age accounted for by that acoustic feature. Features with higher R^2 values were considered more descriptive of differences attributable to development in childhood.

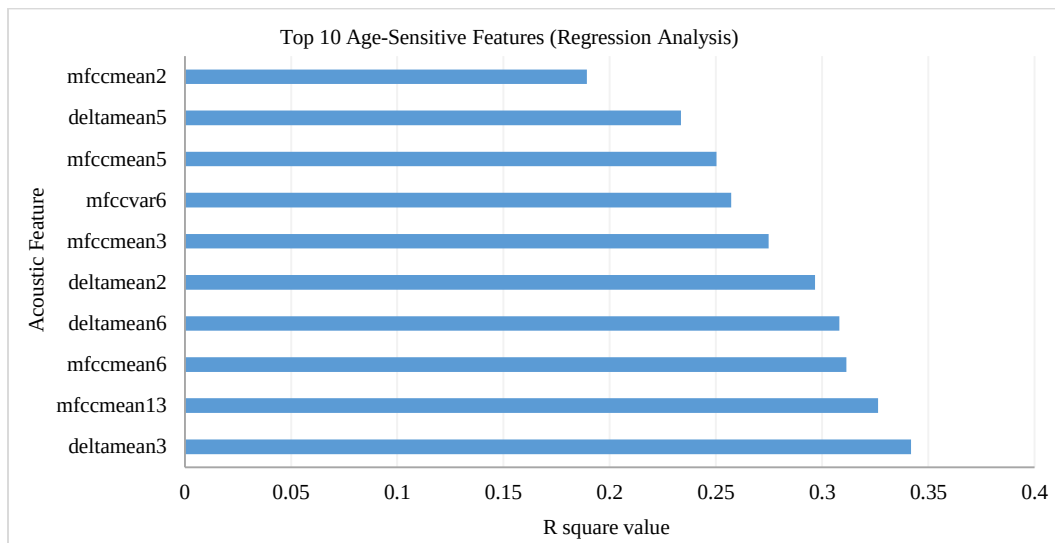


Fig. 2 Illustrates the ten acoustic features with the highest coefficient of determination. Among all the extracted features, **deltamean3** achieved the highest R^2 value (approximately 0.34), indicating that this feature alone explains nearly 34% of the observed variation in speaker age. The next most informative features were **mfccmean13** ($R^2 \approx 0.33$), **mfccmean6** ($R^2 \approx 0.31$), **deltamean2**, and **mfccmean3**, all of which demonstrated substantial predictive capability

The results of regression analysis show the systematic change of the spectral and dynamical acoustic features as children's age grows. Delta MFCC reflects the variation of the current frames within speech and can be used to interpret the stability of the articulatory control. Previous work has proved that in children, the change in motor speech control, the improvement in the integration of movement is accompanied by more consistent dynamic acoustic features of speech. Therefore, the strong correlation of Delta MFCC with age indicates its predictive ability.

In addition, the Delta mean acoustic features may also represent articulatory motor control. In addition, strong importance of the MFCC mean features demonstrates the gradual change of spectral features through childhood mainly attributed to child physical development including the increase in the vocal tract length and vocal cavity length and better motor speech control in the oral region.

As such, the MFCC-based spectral features provide essential developmental speech information for children. These findings agree with earlier studies, which indicate that acoustic characteristics change and also show the significant importance of cepstral features for developmental speech research in children. The influence of this acoustic parameter is reinforced by the predictive power of a statistical model that has some additional capability over other researchers, who relied on a classification task and its accuracy only.

Delta mean acoustic characteristics namely **deltamean3**, **deltamean2** and **mfcc mean** acoustic features namely **mfccmean3**, **mfccmean6** and **mfccmean13** appear to show very little consistency yet predominantly provide large explanatory power to analyses carried out. These features were thereby identified as age-sensitive features of a child speech and might be applicable for applications in building Child speech recognition software and developmental studies.

4.3. ANOVA Analysis

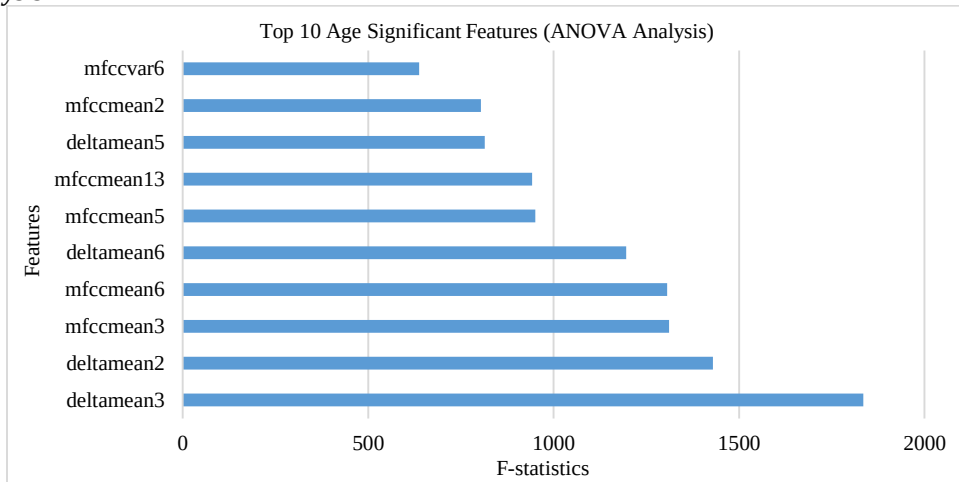


Fig. 3 Top 10 acoustic features in descending order of F-statistic value. Feature deltamean3 had the highest value of the F-statistic, whereas other top 10 most discriminant features in descending order were as follows: deltamean2, mfccmean3, mfccmean6, mfccmean5, mfccmean13 and mfccmean12 with F values = 1834.97 and 1429.51, 1311.09, 1162.23, 963.33 and 889.27, respectively. All p-value is less than 0.05

In an effort to see whether our features from the extract could distinguish the effect of age group differences between our data set, one-way ANOVA was conducted. One-way ANOVA is used to understand how much the variance between the group's means correlates with the variance within each one of the groups. High F-value with a significant p-value led us to conclude that they correlate very well with age, as well as an established marker.

Large Delta MFCC feature values indicate highly dynamic ranges throughout childhood related to speech motor development, articulatory coupling, and an increase in utterance stability. Children tend to move the mouth more variably than their older counterparts. Consequently, the Delta MFCC features do succeed in capturing developmental shifts in speech movement.

Several MFCC mean values did likewise, with the values of some of these MFCC means differing in statistically significant amounts for different age groups. The MFCCs represent the vocal tract's resonance features as conveyed within the spectrum, and their shape has been shown to have changing characteristics based on one's physical changes during one's growth, changing overall length and shape of the oral cavity during speech. Therefore, it is possible that these significant differences should have occurred just due to the process of normal maturation in human development. The ANOVA test our group made on the results showed support towards our previous study on which had identified some of the vocal acoustic variables change in the different ages. Even though other study calculated simply accuracy on classification across acoustic measurements; there has been an additional evidence in our study to what is expected when one measures classification and what is it for acoustic set and statistics: An actual account on how different acoustics affect what one is measuring. The outcome of the present ANOVA

test confirms that deltamean3, deltamean2, mfccmean3, mfccmean6 and mfccmean13 were significantly stable, as no statistical significance is found.

4.4. Developmental Trends of Age-Sensitive Acoustic Features

In order to identify potential changes in usage behavior of selected acoustic features with increasing age, the top-rated features obtained from the correlation analysis, the regression analysis, and the ANOVA were analyzed for all age groups. The Further, the developmental trend analysis in our experiment showed a clear trend with respect to aging for both spectral and dynamic acoustic characteristics. The features, namely deltamean3 and deltamean2, clearly exhibit a decreasing trend across the age for children from immature age to immature children to mature children. This could indicate that in speech production, younger children have greater variety or variability in producing words, while older children produce speech with greater control of articulatory movements and with fewer variabilities. It may be attributed that temporal acoustic variability decreases because there may be gradually changing in fine-tuned motor control for speech, neuromuscular function and greater controllability with regards to articulatory configuration during speaking activity. The MFCC mean features, that is, mfccmean3, mfccmean6 and mfccmean13, show quite a difference in developmental trends from immature children to mature. These features may be associated with vocal tract acoustic characteristics in terms of vocal tract resonances. In response to the change in vocal-tract size and the growing oral cavity of a child, and the improvement of the articulator during childhood to maturity, leads to changing of the frequency slope of the speech envelope. This causes differences in MFCCs related to age. Moreover, these acoustical characteristic show quite a same patterns when analyzed by multiple statistical models of different procedures, meaning the trends and relationships we

get are valid because they consist among correlation analysis, regression analysis, analysis of variance and developmental trend analysis which strengthen evidence to the selected developmental age sensitive acoustic feature, rather than the statistical variability. Previous researches reported similar developmental trends: spectral parameters as well as temporal characteristics of the speech features evolve in direction of adult-like speech as the age of child increases, however most studies addressed high resource language and speech classification based on age group of immature/mature children. This research provided a clear, detailed developmental analysis of Bodo, which is low resource

language and contributed to understanding the speech developmental characteristics and to support technology designing related to this low-resource language. From the conclusion of developmental trend analysis; we can be concluded that the acoustic characteristics of deltamean3, deltamean2, mfccmean3, mfccmean6 and mfccmean13 are stable, age sensitive features across various statistical analyses that can be applicable in age estimation for speakers of the Bodo language, child speech recognition technology, speech monitoring, evaluation of speech disorders for children of various ages as well as the design of speech processing system on a low resource language.

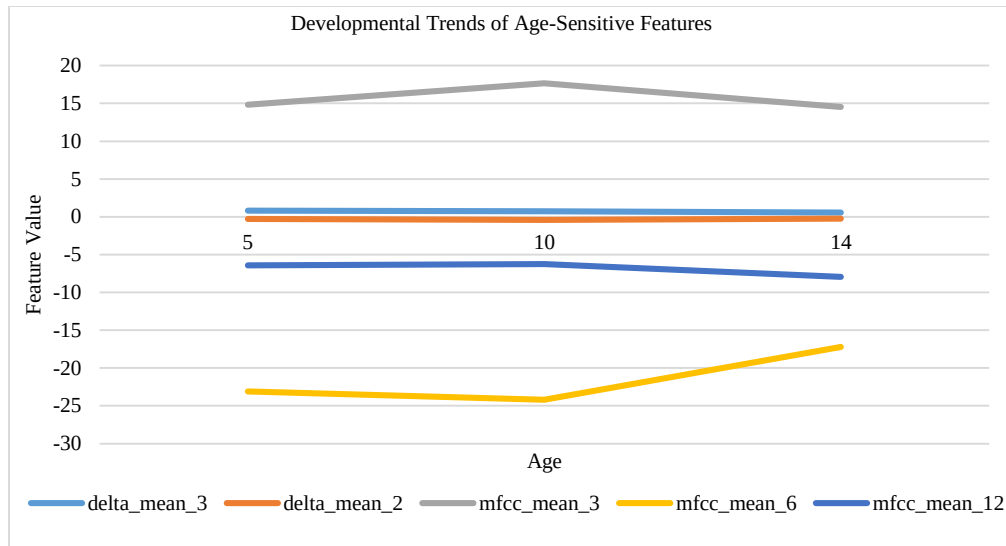


Fig. 4 Illustrates the developmental trends of the most significant acoustic features, namely deltamean3, deltamean2, mfccmean3, mfccmean6, and mfccmean13

4.5. Comparison with Existing Studies

Table 1. Comparison with existing studies

Reference	Language / Dataset	Acoustic Features	Methodology	Major Findings
Lee et al. [1]	English	Temporal and spectral features	Statistical analysis	Demonstrated developmental changes in children's speech acoustics.
Potamianos and Narayanan [2]	English	Acoustic features	Age-adaptive ASR	Improved children's speech recognition using age-specific acoustic models.
Gautam and Singh [5]	English	Fundamental frequency, formants, LPC, duration	Machine learning classifiers	Demonstrated age prediction using multiple acoustic features.
Aggarwal and Singh [6]	Indian children's speech	LPCC	SVM and ANN	Achieved effective age classification using LPCC features.
Kumari et al. [10]	Children's speech	MFCC, Chroma and spectral features	ANN	Reported high age-classification accuracy.
Present Study	Low-resource Bodo children's speech	MFCC mean, MFCC variance, Delta MFCC mean and Delta MFCC variance	Correlation analysis, Linear regression and One-way ANOVA	Identified statistically significant age-sensitive acoustic features (deltamean3, deltamean2, mfccmean3, mfccmean6 and mfccmean13) associated with children's speech development.

As illustrated in Table 1, a substantial number of prior works dealt with the acoustic based prediction of age, classification of human age and classification based on various acoustic feature, etc. However, this proposed work considers statistic determination of age-relevant acoustics feature over development. Moreover, in this investigation, we consider voice features for child speech development and also consider one low-resource language, Bodo. Finally, combining with the proposed methods, including Pearson's correlation, linear regression, and one-way ANOVA, provides supplementary support towards a more stable set of developmental acoustics features as well as a statistical foundation of child speech identification, speaker age classification, and speech evaluation systems.

5. Conclusion

In this study, the age relevant acoustic features of the speech of children (in the low-resource Bodo language) are extracted using Mel-Frequency Cepstral Coefficient (MFCC) spectral and dynamic features. Several statistical methods have been developed for relevant age selection of features from acoustic speech signals Pearson correlation coefficient and one-way ANOVA and regression are used for age assessment, which yields outstanding consistent results, i.e., the features for delta power delta delta step and mfcc (e.g., meanmfcc3, meanmfcc6, meanmfcc13) reveal high degree of correlation and significantly high predictive power (regression coefficient); while one way ANOVA shows significantly differences (F -statistic) for each feature's age review. Through the correlations, the features that correlate with developmental changes were identified.

It has been proven and shown that these acoustical indices are correlates of those changes that take place in speech production through a wide range of levels of analysis. The developmental trend analysis showed clear systematic changes in both spectrum and dynamic acoustic features according to the change of age of children and this trend indicated the fact that the delta MFCCs features showed decreased temporal variability as the child ages, which suggests the improvement of articulatory stability and motor skill for speech. It also suggested that the MFCCs mean features showed that resonance varies gradually with the development of a child's development with vocal tract growth. For example, this work made an attempt to present some of the age-sensitive features found for the Bodo language, which can have direct implications for developing models of developmental age verification and child speech. Compared with other earlier research focusing mainly on speech

recognition systems or classification of individual utterances for speech, this study is unique as it examines individual acoustic feature variability in-development via various statistical methods. It provides a scientific basis for the selection of age-relevant acoustic parameters for several speech application domains, such as age estimation for kids, developmental speech verification, or designing robust Automatic Speech Recognition (ASR) or children's voice synthesis. This paper will contribute towards the development of speech technologies that work best with the development changes. This study demonstrates that MFCC-based spectral features are well-suited to distinguish variations in age among children's speech, as they are sensitive to the underlying developmental modifications in child speech dynamics. Moreover, the strategy presented here could be easily applied to other data-limited languages, thereby providing a standard procedure to utilize speech in diverse applications.

Limitations and Future Work

Presently, there have been important contributions of this research in the direction of analysing developmental speech; However, there are some shortcoming which need to be acknowledged while pursuing future research: (1) This analysis was only conducted on three age groups and therefore does not accurately characterize the ongoing smooth speech development of a child through all their childhood stages. Additional age groups, more speakers and recording of speech from children at other ages will lead to thorough knowledge about the continuous trend in developmental acoustics features variation over different stages. (2) This research relied mostly on MFCC-derived spectral and dynamical acoustic features. Complementary acoustic and prosodic features, including fundamental frequency, formants, energy, spectral descriptors, and characteristics of voice quality, would enable a thorough description of the development of children's speech. (3) This study aims to improve future work by combining these statistical analyses result with supervised learning (ML/DL algorithms), so that an automated system can be built for the recognition of child's speech, age estimation of young children, and assessment of developmental speech features. (4) The speech corpora would eventually consist of many dialects and more Indian low-resources languages in future work in order to improve the reusability and reliability.

Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

References

- [1] Sungbok Lee, Alexandros Potamianos, and Shrikanth Narayanan, "Acoustics of Children's Speech: Developmental Changes of Temporal and Spectral Parameters," *The Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455-1468, 1999. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] A. Potamianos, and S. Narayanan, "Robust Recognition of Children's Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 603-616, 2003. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [3] Serdar Yildirim et al., "Acoustic Analysis of Preschool Children's Speech," *Proceedings of the International Congresses of Phonetic Sciences (ICPhS)*, pp. 949-952, 2003. [[Google Scholar](#)]
- [4] Shawn L. Nissen, and Robert Allen Fox, "Acoustic and Spectral Characteristics of Young Children's Fricative Productions: A Developmental Perspective," *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2570-2578, 2005. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Sumanlata Gautam, and Latika Singh, "Developmental Pattern Analysis and Age Prediction by Extracting Speech Features and Applying Various Classification Techniques," *International Conference on Computing, Communication and Automation*, Greater Noida, India, pp. 83-87, 2015. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Gaurav Aggarwal, and Latika Singh, "Age Classification with LPCC Features using SVM and ANN," *Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS*, vol. 40, pp. 399-408, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Alexey Grigorev, Olga Frolova and Elena Lyakso, "Acoustic Features of Speech of Typically Developing Children Aged 5-16 Years," *Artificial Intelligence and Natural Language: 7th International Conference, AINL 2018*, St. Petersburg, Russia, vol. 930, pp. 152-163, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Elena Lyakso, Olga Frolova, and Aleksandr Nikolaev, "Voice and Speech Features as a Diagnostic Symptom," *Psychological Applications Conference and Trends*, vol. 202, pp. 359-363, 2021. [[Google Scholar](#)]
- [9] Elena Lyakso et al., "Speech Features of 13-15Year-Old Children with Autism Spectrum Disorders," *Speech and Computer: 22nd International Conference, SPECOM 2020*, St. Petersburg, Russia, vol. 12335, pp. 291-303, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Vishakha Kumari, Abhijit Sinha, and Hemant Kumar Kathania, "Role of Acoustics and Prosodic Features for Children's Age Classification," *2024 International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, India, pp. 1-5, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] JunHwi Moon et al., "CSAF: Child-like Speech Augmentation Framework through Adult Speech Data Modulation," *KSII Transactions on Internet and Information Systems*, vol. 19, no. 11, pp. 4119-4137, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Abhijit Sinha, "Effect of Speech Modification on Wav2Vec2 Models for Children Speech Recognition," *2024 International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, India, pp. 1-5, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Sonal Yadav et al., "A Review of Feature Extraction and Classification Techniques in Speech Recognition," *SN Computer Science*, vol. 4, no. 6, pp. 1-13, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Pegah Ghahremani et al., "A Pitch Extraction Algorithm Tuned for Automatic Speech Recognition," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, pp. 2494-2498, 2014. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Jyoti Guglani, and A.N. Mishra, "Automatic speech Recognition System with Pitch Dependent Features for Punjabi Language on Kaldi Toolkit," *Applied Acoustics*, vol. 167, pp. 1-3, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Chang-Sheng Yang, and Hideki Kasuya, "Uniform and non-Uniform Normalization of Vocal Tracts Measured by MRI Across Male, Female and Child Subjects," *IEICE Transactions on Information and Systems*, vol. 78, no. 6, pp. 732-737, 1995. [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Si-Ioi Ng, Cymie Wing-Yee Ng, and Tan Lee, "A Study on using Duration and Formant Features in Automatic Detection of Speech Sound Disorder in Children," *Proceedings Interspeech*, pp. 4643-4647, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Samreen Naeem et al., "An Unsupervised Machine Learning Algorithms: Comprehensive Review," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 911-921, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Muhammad Usama et al., "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges," *IEEE Access*, vol. 7, pp. 65579-65615, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Thomas Drugman et al., "Traditional Machine Learning for Pitch Detection," *IEEE Signal Process Letters*, vol. 25, no. 11, pp. 1745-1749, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] M.K. Benkaddour, "CNN based Features Extraction for Age Estimation and Gender Classification," *Informatica*, vol. 45, no. 5, pp. 697-707, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Okko Räsänen, and Daniil Kocharov, "Age-Dependent Analysis and Stochastic Generation of Child-Directed Speech," *arXiv preprint arXiv*, vol. 13, pp. 1-7, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]