JTIM : Jurnal Teknologi Informasi dan Multimedia

p-ISSN : 2715-2529
e-ISSN : 2684-9151
https://journal.sekawan-org.id/index.php/jtim

Acoustic Analysis on Cleft Lip Speech Signal

Siti Agrippina Alodia Yusuf [1]*, Nani Sulistianingsih [1] , Muhammad Imam Dinata [1], Syahroni Hidayat [2] and Joelianto Darmawan [1]

[1] Department of Information Systems and Technology, Universitas Muhammadiyah Mataram, Indonesia
[2] Department of Electrical Engineering, Universitas Negeri Semarang, Indonesia

* Correspondence: alodiaysf@gmail.com

Citation: Yusuf, S. A. A.; Sulistianingsih, N.; Dinata, M. I.;, Hidayat, S.; and Darmawan, J. 1 (**2025**). Acoustic Analysis on Cleft Lip Speech Signal. JTIM: Jurnal Teknologi Informasi Dan Multimedia, 7(4), 744-753. https://doi.org/10.35746/jtim.v7i4.766

Received: 11-06-2025
Revised: 11-07-2025
Accepted: 27-08-2025

Abstract: Cleft conditions significantly disrupt phonetic articulation, leading to hypernasality and irregular resonance characteristics. In this study, the formant analysis of normal and cleft speech is presented, with the aim of investigating acoustic differences in formant frequencies between cleft and normal speech using real-word utterances, focusing on the articulation of plosive consonants and resonance variability. The dataset consisted of 280 speech signals (140 cleft and 140 normal) uttering word /paku/. The speech signals were resampled to 16kHz and the silence in the speech was removed, next stage was followed by extracting the first three formants using the Burg algorithm. Statistical analysis revealed that the value of F1 and F2 in cleft speech were higher, alongside greater variability in formant distribution. Further analysis of plosive articulation highlighted irregular formant transition in cleft speech, indicating compromised intraoral pressure control. Additionally, a moderate negative correlation ($r = -0.423$, $p<0.001$) between F1 and F3 suggests a spectral pattern indicative of hypernasality. This finding underscores the potential of formant-based acoustic features as objective markers for early clinical assessment and provides a foundation for the development of diagnostic models in cleft speech research.

Keywords: cleft speech, hypernasality, formant analysis, acoustic features, speech signal;

1. Introduction

Cleft anomalies are among the most prevalent congenital conditions worldwide, commonly categorized as cleft lip (CL), cleft palate (CLP), or a combination of both (CL/P) [1]. A cleft palate is characterized by an opening in the roof of the mouth, forming an abnormal connection between the oral and nasal cavities [2]. In contrast, a cleft lip results from the incomplete fusion of the medial and nasal prominences with the maxillary prominence, leading to a separation in the upper lip [3]. Individuals affected by these conditions often experience a range of speech disorders, including hypernasality, hyponasality, consonant production errors, and voice disorders [4]. Hypernasality refers to an abnormal increase in nasal resonance during speech, typically caused by velopharyngeal insufficiency (VPI), where the velum fails to close properly, allowing air to escape through the nasal cavity during phonation [5]. On the other hand, hyponasality is marked by reduced nasal resonance, in which nasal sounds such as /m/ and /n/ are perceived as /b/ and /d/, often due to nasal obstruction or velopharyngeal dysfunction [6]. Nasal air emission, defined as the audible release of air through the nose during speech—particularly during the articulation of pressure consonants—is frequently associated with insufficient velopharyngeal closure [7]. Surgical intervention is often required early in life to restore anatomical structure, though multiple surgeries may be necessary depending on the cleft's severity [8]. Post-operative assessment by speech-language pathologists (SLPs) is crucial; however, limited availability of SLPs, particularly in rural areas, and the high cost of services pose substantial barriers to care.

Automatic speech recognition (ASR) is a technology designed to convert spoken language into text by processing acoustic signals [9], [10], and it is applied across various domains, including mobile interfaces, human-computer interaction, accent and language detection, voice biometrics, healthcare, and transcription systems. ASR systems typically consist of three core components: feature extraction, which identifies key features in the speech signal; acoustic modeling, which maps these features to phonetic elements; and classification, which matches the patterns to known linguistic units [10]. Despite advancements, ASR technologies are limited in their ability to process atypical speech, particularly cleft-affected speech, due to sparse dataset and high phonetic variability [11], [12]. Such atypical speech patterns, particularly in individual with cleft condition, are underpresented in current ASR research.

Recent studies have begun to address cleft speech processing, particularly focusing on intelligibility assessment and hypernasality detection [13], [14]. Song et al. [15] developed a method for estimating hypernasality by fine-tuning an ASR encoder on cleft-specific speech-to-text datasets, while Wang et al. [16] employed a Long Short-Term Memory-based Deep Recurrent Neural Network (LSTM-DRNN) to detect hypernasality in sustained vowels /a/, /i/, and /u/. While these methods demonstrate promise, they are primarily dependent on large, labeled datasets and complex deep-learning architectures, which may not be practical in low-resource or clinical environments where cleft data is scarce. Moreover, their focus on isolated vowels limits the ecological validity of their models for connected speech or phonetically complex words.

In contrast, this study uses formant-based acoustic analysis to characterize cleft speech. Specifically, we analyze the formant trajectories of the word */paku/*, which was deliberately selected due to its inclusion of two voiceless plosive consonants (/p/ and /k/), making it phonetically sensitive to articulatory impairments common in cleft conditions. Plosive consonants require complete intraoral pressure build-up and precise velopharyngeal closure, which are often compromised in individuals with cleft palate [17]. This makes */paku/* an effective probe for detecting articulatory instability and resonance deviation in cleft speech. Unlike deep learning-based methods, our approach emphasizes interpretable acoustic features of formant frequencies (F1, F2, F3), allowing for direct insights into the physiological basis of speech differences. Therefore, this research aims to perform an acoustic analysis of cleft speech by examining the statistical distribution and transition patterns of the first three formants in the word */paku/*. This study aims to investigate how anatomical irregularities in the vocal tract affect speech production in individuals with cleft conditions. Specifically, it focuses on the articulation of plosive consonants and the manifestation of hypernasality as acoustic deviations in formant structure.

2. Materials and Methods

This study employs a quantitative, comparative research approach to analyze acoustic differences between cleft and non-cleft speech. The investigation focuses on examining the statistical characteristics of the first three formant frequencies (F1, F2, F3) as key acoustic indicators. The methodology includes systematic stages of dataset acquisition, pre-processing, feature extraction, and statistical evaluation to ensure analytical rigor and reproducibility. The methods are provided in Figure 1.

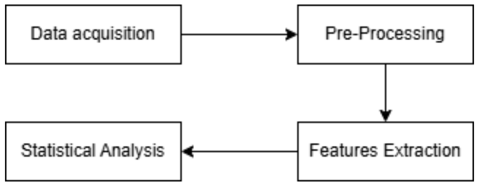

Figure 1. Methodology

2.1. *Dataset Acquisition*

Speech data were collected from individuals with cleft lip (CL), cleft palate (CP), or both (CL/P), including those who had undergone reparative surgery. Each participant was instructed to pronounce the word /paku/ repeatedly for 10 to 15 iterations.The word was chosen due to the inclusion of two voiceless plosives within a simple syllabic structure, thus making /paku/ phonetically ideal for revealing articulatory challenges in cleft speech. The dataset is balanced, consisting of 140 utterances each from speakers with cleft conditions and control (non-cleft) speakers. Recordings were captured using mobile phone applications in both open and closed environments, reflecting the variability and accessibility constraints commonly found in real-world clinical and remote settings. To mitigate environmental effects, energy-based silence trimming was applied, and only utterances with clear acoustic energy profiles and minimal background interference were included in the analysis. Speech durations ranged between 0.9 and 1.2 seconds per utterance.

2.2. *Pre-processing*

In this stage, all speech signals were standardized by applying resampling to 16 kHz. This step is essential for uniform frame segmentation and accurate spectral analysis. Following resampling, the recordings underwent energy-based silence removal to improve the reliability of the acoustic analysis. This technique identifies and eliminates non-speech segments—especially leading and trailing silences—based on the Root Mean Square (RMS) energy computed across overlapping frames (typically 20–30 ms). RMS values were converted to a decibel (dB) scale, and a threshold of 20 dB was applied to detect active speech regions [18]. This pre-processing step improves the signal-to-noise ratio and reduces the influence of environmental noise and silence on formant estimation.

2.3. *Feature Extraction*

In this stage, the Burg algorithm was employed to extract the formants. This linear prediction method models the speech signal with an autoregressive (AR) model, which posits that the speech signal can be expressed as a linear combination of its previous values. The algorithm estimates the parameters of the predictive filter, which are then used to compute the linear prediction spectrum. The peaks in this spectrum correspond to the formant frequencies [19], [20]. The algorithm divides the audio signal into short, overlapping frames to account for the quasi-stationary nature of speech. For each frame, the Burg method recursively minimizes both forward and backward prediction errors to estimate the LPC coefficients [21]. Unlike other LPC methods, Burg guarantees filter stability and avoids spectral leakage, making it ideal for formant tracking [22]. In this study, we set the Burg method for formant extraction using Praat's default configuration via the Parselmouth library, with specific adjustments to the maximum number of formants and the maximum formant frequency. The number of formants was limited to five, with the maximum formant frequency was set to 5000 Hz, which is commonly suitable for adult speech analysis [23]. Other parameters such as the analysis time step (0.01 seconds), window length (0.025 seconds), and pre-emphasis frequency (50 Hz)—were also retained as defaults, because they are optimized for capturing quasi-stationary characteristics of speech frames. These settings allow for consistent temporal and spectral resolution, ensuring that the formant trajectories of cleft and non-cleft speech samples are comparable and analytically robust.

2.4. Statistical Analysis

To assess acoustic differences between cleft and non-cleft speech, the extracted formant frequencies (F1, F2, and F3) were analyzed using a structured statistical framework. This evaluation consisted of four main components: descriptive statistics, statistical significance testing, formant transition analysis of plosive consonants, and correlation analysis related to hypernasality. First, descriptive statistics were calculated to summarize the central tendency and variability of each formant, including measures such as mean, standard deviation, minimum, maximum, and interquartile range. The Shapiro–Wilk test was applied to examine the normality of the formant distributions. Based on this assessment, a non-parametric approach was selected. Specifically, the Mann–Whitney U test was employed to compare formant frequency distributions between cleft and non-cleft groups due to its robustness in handling non-normally distributed data.

In addition to static formant values, formant transitions during plosive consonant articulation were examined. This analysis focused on the voiceless plosives /p/ and /k/ in the word /paku/. Five utterances were randomly selected from each group, and variability in F2 and F3 was measured using standard deviation and range, allowing for quantification of articulatory consistency during plosive production. To investigate acoustic patterns associated with hypernasality, a correlation analysis was conducted within the cleft speech group. The linear relationship between F1 and F3 was analyzed using Pearson's correlation coefficient, which served as a proxy measure for detecting spectral alterations related to nasal resonance coupling.

3. Results

This section presents the results of the acoustic analysis, structured around descriptive statistics, statistical significance testing, phonetic features of plosive consonants, and evidence of hypernasality based on formant relationships. The analysis was conducted on 280 speech signals (140 cleft and 140 normal), resampled to 16 kHz, and pre-processed for silence removal.

3.1. Statistical Description of Formants

Table 1 summarizes the central tendencies and variability of the first three formants (F1, F2, F3) for cleft and normal speech samples.

Table 1. Statistical descriptions of formant frequencies (Hz) by Condition

Formant	Statistic	Normal (Hz)	Cleft (Hz)
F1	Mean	584.25	638.63
	Std Dev	56.39	95.20
	Min	461.91	465.76
	25% Quartile	534.75	568.39
	50% Quartile	584.86	640.63
	75% Quartile	625.58	674.54
	Max	699.62	1058.77
F2	Mean	1178.20	1329.14
	Std Dev	133.60	269.17
	Min	921.55	930.71
	25% Quartile	1085.63	1124.63
	50% Quartile	1168.76	1322.59
	75% Quartile	1199.06	1486.66
	Max	1593.78	2038.38
F3	Mean	2260.72	2250.39

Formant	Statistic	Normal (Hz)	Cleft (Hz)
	Std Dev	187.44	331.12
	Min	1896.34	1820.01
	25% Quartile	2126.66	1990.36
	50% Quartile	2198.67	2157.69
	75% Quartile	2408.05	2614.41
	Max	2843.11	2940.42

On average, cleft speech exhibited higher F1 and F2 values compared to normal speech, with a greater standard deviation across all three formants, indicating more acoustic variability. Notably, the mean F1 for cleft speech was 638.63 Hz compared to 584.25 Hz for normal speech, while F2 was 1329.14 Hz in cleft speech versus 1178.20 Hz in normal speech. F3 values, however, showed a marginal difference (2250.39 Hz in cleft vs 2260.72 Hz in normal), though variability was markedly higher in the cleft group. Boxplots of the formants distribution (Figure 2) illustrate the broader spread and presence of high-frequency outliers, particularly in the F1 and F2 distributions of cleft speech.

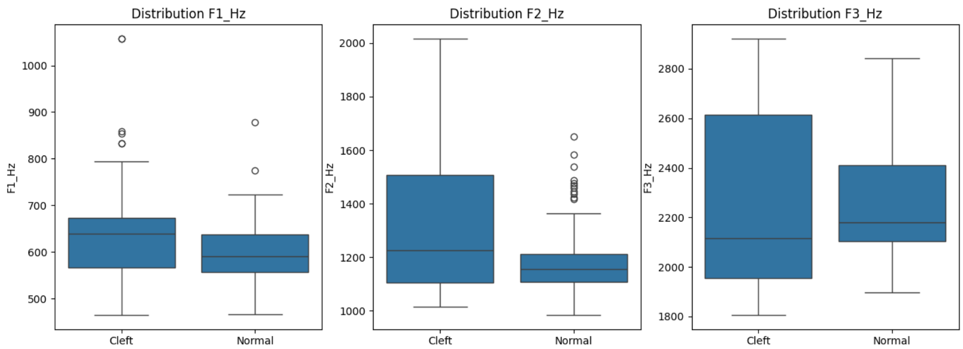

Figure 2. Formants distribution (F1, F2, F3) for cleft and normal speech

Figure 2 emphasizes these differences, showing that in F1, the cleft speech signal exhibits a higher median, a broader interquartile range, and a prominent outlier around 1050 Hz, which indicates a more dispersed and elevated distribution. Additionally, the cleft signal in F2 is characterized by a taller box plot, reflecting greater variability, with an extended upper whisker surpassing 2000 Hz. On the other hand, the normal speech signal is more confined, with several outliers appearing above the upper quartile. In F3, both signals share a similar median, though cleft signals display a broader range. Individuals with cleft conditions often exhibit significant anatomical variation in the vocal tract, including differences in the shape and size of the oral and nasal cavities. These variations can lead to inconsistent resonance characteristic, resulting in broader formant distributions [24], [25]. Velopharyngeal insufficiency, in which the soft palate fails to close properly against the back of the throat during speech affect the acoustic properties of speech, contributing to variability in formant frequencies [24], [25], [26].

The observed variability in formant distributions aligns with findings by Nikitha et al. [27], who examined vowel space area (VSA) in children with cleft lip and palate and reported a progressive reduction in VSA with increasing hypernasality severity. Although their study used VSA as a global acoustic metric derived from F1 and F2, the increased dispersion observed in the present study's formant distributions, particularly in F1 and F2, similarly reflects reduced articulatory precision and stability. Both analyzes indicate that cleft-affected speech is characterized by imprecise vowel articulation due to underlying anatomical and physiological disruptions. Moreover, while the study employed sustained vowels and CVCV sequences in controlled phonetic contexts, this study analyzed a real-word utterance (/paku/), incorporating bilabial and plosives in a naturally coarticulated speech context. Despite the relatively lower articulatory demands expected in bilabial contexts such as /p/, the current results showed significant variability in formant distributions even within this phonetic environment. This suggests that natural speech may reveal articulatory instability more clearly than isolated segments. Additionally, both studies link the findings to common physiological mechanisms, namely velopharyngeal insufficiency, oral-nasal coupling anomalies, and compensatory articulation strategies, which collectively impact formant production. Therefore, the broader formant variability observed in this study supports and extends previous study's conclusions by demonstrating similar articulatory-acoustic consequences within a connected speech context.

3.2. Statistical Significance Testing

To evaluate differences in formant frequencies between cleft and normal speech, the Shapiro-Wild test was first conducted to assess normality. All formant distribution in both speaker groups showed significant deviations from normality ($p<0.05$), therefore the Mann-Whitney U test was applied for group comparisons. The results are shown in Table 2.

Table 2. Significance test result

Formant	p-Value	Significance
F1	0.0000270	Significance
F2	0.00188	Significance
F3	0.0114	Significance

Examining the table, it is evident that there are significant differences across all formants. The p-Values for F1 and F2 reflect highly significant values, suggesting notable disparities in acoustic distribution likely stemming from variations in articulation and resonance due to the cleft's morphological structure. Although F3 also displays a statistically significant difference, the magnitude of this difference is quite modest, necessitating mild but observable differences in resonance behavior. In summary, the findings indicate that individuals with a cleft exhibit altered articulation, leading to distinctive patterns in vocal acoustics.

3.3. Plosive Consonant Analysis

The word /paku/ contains voiceless plosive /p/ and /k/, which require precise intraoral pressure control. Variability in formant transitions during plosive production was analyzed using the standard deviation and range of F2 and F3. Figure 2 demonstrated the formant transition of F2 and F3 where five random samples from each group were taken.

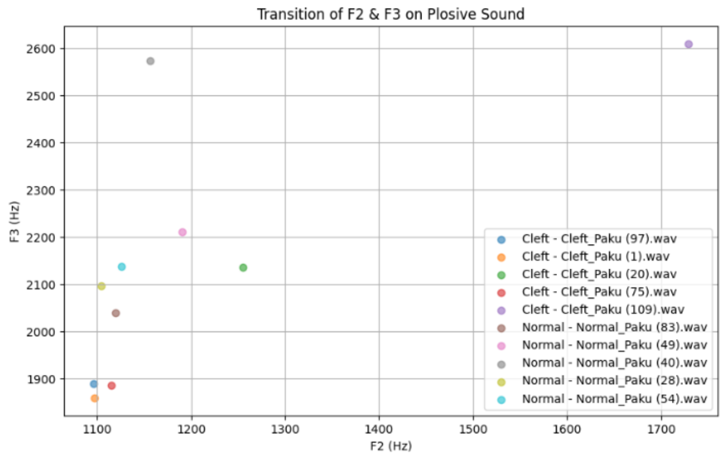

Figure 3. F2 and F3 transition trajectories across five utterances per group showing variability during plosive production

As demonstrated by the formant transition in Figure 3, the distribution in individuals with a cleft condition was broader and more irregular than in those without. This irregularity suggests difficulties in articulation particularly in producing plosive sounds, which require greater intraoral control. In F2 transition variability for cleft speech, the standard deviation and range are 269.17 Hz and 1002.04 Hz respectively. Meanwhile, in normal speech, the standard deviation and range are 133.60 Hz and 667.04 Hz. Furthermore, the F3 transition variability on cleft and normal speech is 331.12 Hz and 1117.77 Hz, and 187.44 Hz and 944.79 Hz respectively. These results clearly indicate that cleft speech exhibits significantly greater variation in both F2 and F3 transitions. The findings align with [28], stating that there is differences in F2 and F3 formant values during the production of voice plosives in children with cleft. Although the current study involves adults, the persistence of such acoustic anomalies suggests that cleft-related articulatory deficits may extend into adulthood, particularly in the absence of effective early intervention. Moreover, structural issues such as dental anomalies in cleft patients can lead to altered speech production which may manifest as irregular formant transition [29], [30]. This finding is further supported by [31], demonstrating that anatomical variation in the vocal tract contribute to variability in formant frequencies. Their study confirms that differences in the oral cavity can drive dispersion in F3 patterns. Such anatomical differences are particularly pertinent to the cleft population, where malformations of the oral cavity are frequently observed. Conversely, in the normal group, the formant transition appears consistent, suggesting regular and efficient plosive sound production. Consistent with the finding of [31], the stability in transitions are likely supported by typical anatomical development, which allows for more predictable acoustic outcomes during plosive production. Therefore, the present study highlights how structural deviations in cleft speech affect both articulatory precision and the acoustic cues necessary for clear speech perception.

3.4. Hypernasality Correlation Analysis

Typically, changes in F1, particularly an increase in its amplitude and bandwidth, are indicative of hypernasality [32], [33]. This rise occurs because the nasal cavity is coupled with the oral cavity, resulting in additional nasal formants and anti-formant appearing within the speech spectrum [34], [35]. Additionally, the spectral traits of hypernasality involve troughs between F2 and F3, attributed to the enhanced acoustic coupling of the nasal cavity, which influences the spectral envelope and formant configuration. Therefore, the correlation between F1 and F3 was examined as a proxy for hypernasality. We calculated Pearson's correlation coefficient between F1 and F3 for the cleft speech group. The calculation yielded a moderate negative correlation ($r = -0.423$, $p<0.001$) as shown in Figure 4 which aligns with prior findings that associate this spectral behavior with hypernasality resonance patterns. This moderate inverse realtionship provides a quantifiable acoustic correlate to hypernasality, supportin gthe development of formant-based detection frameworks.

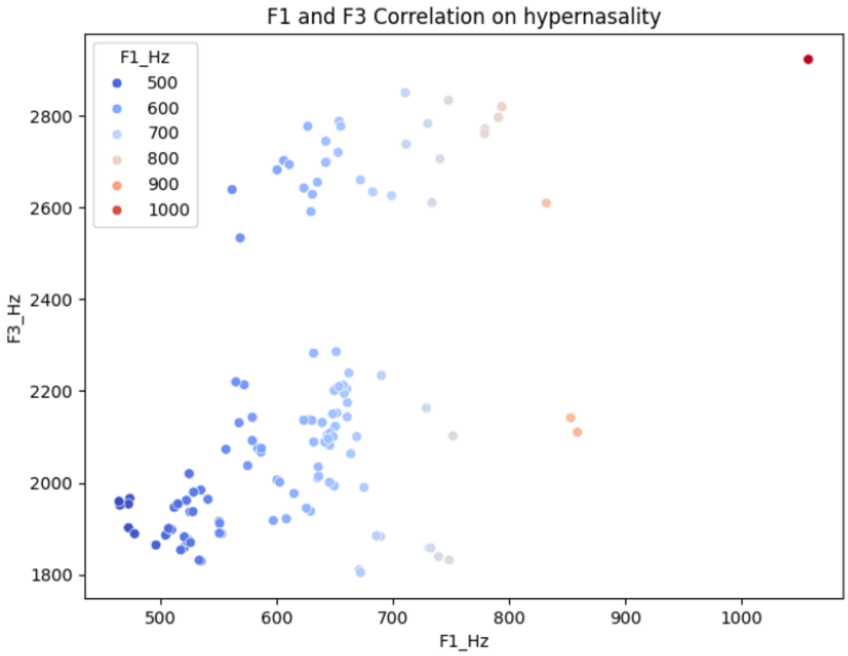

Figure 4. Scatterplot showing F1-F3 correlation in cleft speech group

4. Conclusions

In this research, we investigated the formant frequencies of the word /paku/ as produced by speakers with cleft conditions. The experiments demonstrated that a cleft significantly impacts vowel formant features, affecting the articulation of plosive consonants and modifying resonance, which can result in hypernasality. This study offers an acoustic perspective on how anatomical abnormalities in the oral cavity influence speech quality and may serve as a foundation for future research, particularly in developing formant-based methods for robust vocal tract length normalization.

References

[1] F. A. Putri, M. Pattamatta, S. E. S. Anita, and T. Maulina, "The Global Occurrences of Cleft Lip and Palate in Pediatric Patients and Their Association with Demographic Factors: A Narrative Review," *Children*, vol. 11, no. 3, p. 322, 2024, https://doi.org/10.3390/children11030322.

[2] J. Qian, F. Fu, X. Liu, L. He, H. Yin, and H. Zhang, "The analysis and detection of hypernasality based on a formant extraction algorithm," in *Journal of Physics: Conference Series*, IOP Publishing, 2017, p. 012082. https://doi.org/10.1088/1742-6596/887/1/012082.

[3] S. Mzezewa, K. Hamese, and T. A. B. Mashego, "Neonatal cleft lip repair in babies with breastfeeding difficulties at Polokwane Mankweng Hospital Complex," *South African Journal of Child Health*, vol. 8, no. 4, pp. 137–140, Nov. 2014, https://doi.org/10.7196/SAJCH.693.

[4] K. Sireesha, A. K. Dubey, D. Govind, S. K, and S. V. Gangashetty, "Variational mode decomposition based features for detection of hypernasality in cleft palate speech," *Biomedical Signal Processing and Control*, vol. 97, p. 106689, 2024, https://doi.org/10.1016/j.bspc.2024.106689.

[5] H. A. Carvajal-Castaño and J. R. Orozco-Arroyave, "Articulation Analysis in the Speech of Children with Cleft Lip and Palate," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11896 LNCS, pp. 575–585, 2019, https://doi.org/10.1007/978-3-030-33904-3_54.

[6] D. Sell, V. Pereira, Y. Wren, and J. Russell, "Speech Disorders Related to Cleft Palate and Velopharyngeal Dysfunction," in *The Handbook of Language and Speech Disorders*, Wiley, 2021, pp. 468–494. https://doi.org/10.1002/9781119606987.ch21.

[7] E. Sundström, S. Boyce, and L. Oren, "Effects of velopharyngeal openings on flow characteristics of nasal emission," *Biomechanics and Modeling in Mechanobiology*, vol. 19, no. 5, pp. 1447–1459, Oct. 2020, https://doi.org/10.1007/s10237-019-01280-9.

[8] A. Zhang, R. E. Pyon, K. Chen, and A. Y. Lin, "Speech Analysis of Patients with Cleft Palate Using Artificial Intelligence Techniques: A Systematic Review," *FACE*, vol. 4, no. 3, pp. 327–337, 2023, https://doi.org/10.1177/27325016231187985.

[9] A. Adjila, M. Ahfir, and D. Ziadi, “Proceedings - 2021 International Conference on Information Systems and Advanced Technologies, ICISAT 2021,” in *11th International Conference on Information Systems and Advanced Technologies (ICISAT)*, Virtual, Online: IEEE, Dec. 2021. https://doi.org/10.1109/ICISAT54145.2021.9678476.

[10] S. R and R. C.S, “A data-driven weighted LP method for formant estimation,” in *Proceedings of the 4th IEEE Conference on Information and Communication Technology (CICT 2020)*, Chennai, India: IEEE, Dec. 2020. https://doi.org/10.1109/CICT51604.2020.9312110.

[11] K. Kajihara, S. Izumi, S. Yoshida, Y. Yano, H. Kawaguchi, and M. Yoshimoto, “Hardware Implementation of Autoregressive Model Estimation Using Burg’s Method for Low-Energy Spectral Analysis,” in *2018 IEEE Workshop on Signal Processing Systems (SiPS)*, Cape Town, South Africa: IEEE, Oct. 2018. https://doi.org/10.1109/SiPS.2018.8598315.

[12] R. Gupta, T. Chaspari, J. Kim, N. Kumar, D. Bone, and S. Narayanan, “Pathological speech processing: State-of-the-art, current challenges, and future directions,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 6470–6474. https://doi.org/10.1109/ICASSP.2016.7472923.

[13] Y. Kozaki-Yamaguchi, N. Suzuki, Y. Fujita, H. Yoshimasu, M. Akagi, and T. Amagasa, “Perception of Hypernasality and its Physical Correlates,” *Oral Science International*, vol. 2, no. 1. pp. 21–35, 2005. https://doi.org/10.11277/osi.2.21.

[14] J. Weerasinghe, J. Sato, and K. Kawaguchi, “Spectral evaluation of hypernasality in children with repaired cleft palate,” *Asian Journal of Oral and Maxillofacial Surgery*, vol. 18, no. 3. pp. 191–201, 2006. https://doi.org/10.1016/S0915-6992(06)80017-8.

[15] A. K. Dubey, S. R. M. Prasanna, and S. Dandapat, “Detection and assessment of hypernasality in repaired cleft palate speech using vocal tract and residual features,” *The Journal of the Acoustical Society of America*, vol. 146, no. 6, pp. 4211–4223, Dec. 2019, https://doi.org/10.1121/1.5134433.

[16] P. Vijayalakshmi and M. R. Reddy, “Detection of hypernasality using statistical pattern classifiers,” *9th European Conference on Speech Communication and Technology*. pp. 701–704, 2005.

[17] M. L. Ng, E. T.-S. Tong, and K. M. Yu, “Articulatory Contact Pressure during Bilabial Plosive Production in Esophageal and Tracheoesophageal Speech,” *Folia Phoniatr Logop*, vol. 71, no. 1, pp. 1–6, 2019, https://doi.org/10.1159/000493344.

[18] D.-J. Kim, “Formant detection technique for the phonocardiogram spectra using the 1st and 2nd derivatives,” *Transactions of the Korean Institute of Electrical Engineers*, vol. 64, no. 11. pp. 1605–1610, 2015. https://doi.org/10.5370/KIEE.2015.64.11.1605.

[19] D.-J. Kim, “A Study on Acoustic Analysis of Stethoscope Signal using the Burg Algorithm,” *Transactions of the Korean Institute of Electrical Engineers*, vol. 70, no. 1. pp. 236–242, 2021. doi: https://doi.org/10.5370/KIEE.2021.70.1.236.

[20] G. Suciu, S. Segărceanu, A. Negoiță, and D. A. Trufin, “SPEECH RECOGNITION SYSTEM,” *eLearning and Software for Education Conference*. pp. 203–210, 2021. https://doi.org/10.12753/2066-026X-21-095.

[21] J. Ujwala Rekha, K. Shahu Chatrapati, and A. Vinaya Babu, “A studyon speech processing,” *Advances in Intelligent Systems and Computing*, vol. 435. pp. 209–226, 2016. https://doi.org/10.1007/978-81-322-2757-1_22.

[22] K. Babu Rao, B. Mopuru, M. Jawarneh, J. Luis-Arias, S.-S. M. Ajibade, and P. Prabhu, “Automatic speech recognition design modeling,” *Conversational Artificial Intelligence*. Wiley, 2024. https://doi.org/10.1002/9781394200801.ch22.

[23] M. H. Ali *et al.*, “Harris Hawks Sparse Auto-Encoder Networks for Automatic Speech Recognition System,” *Applied Sciences (Switzerland)*, vol. 12, no. 3. 2022. https://doi.org/10.3390/app12031091.

[24] M. Scipioni, M. Gerosa, D. Giuliani, E. Nöth, and A. Maier, “Intelligibility assessment in children with cleft lip and palate in Italian and German,” *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. pp. 967–970, 2009.

[25] A. Schulz, T. Bocklet, U. Eysholdt, C. Bohr, M. Döllinger, and A. Ziethe, “Validation of an automatic speech analysis in children with isolated cleft palate; [Validierung einer automatischen Analyse der Sprechproben von Kindern mit isolierter Gaumenspalte],” *HNO*, vol. 62, no. 7. pp. 525–529, 2014. https://doi.org/10.1007/s00106-013-2825-x.

[26] K. Song *et al.*, “Improving Hypernasality Estimation with Automatic Speech Recognition in Cleft Palate Speech,” in *Interspeech 2022*, ISCA, Sep. 2022, pp. 4820–4824. https://doi.org/10.21437/Interspeech.2022-438.

[27] N. K., S. Kalita, C. M. Vikram, M. Pushpavathi, and S. R. M. Prasanna, “Hypernasality Severity Analysis in Cleft Lip and Palate Speech Using Vowel Space Area,” in *Interspeech 2017*, ISCA, Aug. 2017, pp. 1829–1833. https://doi.org/10.21437/Interspeech.2017-1245.

[28] X. Wang, S. Yang, M. Tang, H. Yin, H. Huang, and L. He, “HypernasalityNet: Deep recurrent neural network for automatic hypernasality detection,” *International Journal of Medical Informatics*, vol. 129. pp. 1–12, 2019. https://doi.org/10.1016/j.ijmedinf.2019.05.023.

[29] L. Yang, Y. Mu, Y. Zhai, and R. Chen, “Impaired speech input and output processing abilities in children with cleft palate speech disorder,” *International Journal of Language and Communication Disorders*, vol. 59, no. 5. pp. 1906–1922, 2024. https://doi.org/10.1111/1460-6984.13037.

[30] Q. Lou, X. Wang, L. Jiang, G. Wang, Y. Chen, and Q. Liu, “Subjective and Objective Evaluation of Speech in Adult Patients with Unrepaired Cleft Palate,” *Journal of Craniofacial Surgery*, vol. 33, no. 5. pp. E528–E532, 2022. https://doi.org/10.1097/SCS.0000000000008567.

[31] H. K. Vorperian *et al.*, “Developmental Sexual Dimorphism of the Oral and Pharyngeal Portions of the Vocal Tract: An Imaging Study,” *J Speech Lang Hear Res*, vol. 54, no. 4, pp. 995–1010, Aug. 2011, https://doi.org/10.1044/1092-4388(2010/10-0097).

[32] M. Pushpavathi, A. K. Abraham, S. R. Mahadeva Prasanna, and K. S. Girish, "Perceptual judgments of resonance, speech understandability, and speech acceptability in children with repaired cleft palate across words and sentences," *Lecture Notes in Mechanical Engineering*. pp. 75–83, 2021. https://doi.org/10.1007/978-981-15-5776-7_7.

[33] L. Ristovska, Z. Jachova, L. Spasov, and T. Balova, "Acoustic characteristics of Macedonian vowels," *Journal of Special Education and Rehabilitation*, vol. 19, no. 3–4. pp. 40–50, 2018. https://doi.org/10.19057/jser.2019.39.

[34] M. Béchet, F. Hirsch, C. Fauth, and R. Sock, "Consonantal space area in Children with a Cleft Palate An acoustic Study," *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012*, vol. 1. pp. 58–61, 2012.

[35] E. Jiménez-Castellanos, A. Carmona, C. J. Catalina-Herrera, and J. Jiménez-Castellanos, "Oral anatomical defects associated with cleft palate and cleft lip," *European Journal of Anatomy*, vol. 8, no. 3. pp. 137–141, 2004.