Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 1 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i01.57991

PAPER

Functional Validation of a Generative AI-Based Mobile App for Assessing Speech Difficulties in Children

Maritza Arones[1], Irma Aybar-Bellido[1], Willy Adauto-Medina[2] (✉), Santiago Rubiños-Jimenez[2], José Antonio Arévalo-Tuesta[3]

[1]Universidad Nacional San Luis Gonzaga, Ica, Perú

[2]Universidad Nacional Tecnológica de Lima Sur, Lima, Perú

[3]Universidad Nacional Federico Villarreal, Lima, Perú

wadauto@untels.edu.pe

ABSTRACT

Speech and voice development in childhood are essential for academic progress and social participation; difficulties in oral expression may impact learning and emotional health. Generative artificial intelligence (GAI) technologies integrated into interactive mobile applications offer new possibilities for the automated assessment of Spanish-speaking children's speech. This study functionally validates a mobile application that uses GAI to automatically assess children's speech fluency, pronunciation, and intonation, providing automated scoring, targeted feedback, and personalized recommendations. This Phase 1 functional validation, based on synthetic data, lays the groundwork for a four-phase framework aimed at guiding cross-national and multilingual research in artificial intelligence (AI)-supported speech evaluation. The methodology focused on a correlational analysis between the scores generated by the application and the acoustic indicators—number of pauses, pitch range, and spectral clarity—obtained from 160 samples. Descriptive and spectrographic analyses revealed mean decibel levels ranging from 15 to 33 dB, pitch ranges around 3.843 Hz, and spectral clarity between 0.033 and 0.036. It is concluded that this tool could contribute to the automated and multidimensional assessment and feedback of speech difficulties in Spanish-speaking children.

KEYWORDS

mobile application, generative artificial intelligence (GAI), speech difficulties, spectral analysis, functional validation

1 INTRODUCTION

During childhood, the development of speech and voice constitutes a fundamental pillar for communication, socialization, and learning, so that difficulties in oral expression can compromise both academic performance and the emotional and social integration of children [1]. These problems are usually caused by psycho-pedagogical, environmental, or specific language disorders, impacting key

Arones, M., Aybar-Bellido, I., Adauto-Medina, W., Rubiños-Jimenez, S., Arévalo-Tuesta, J. A. (2026). Functional Validation of a Generative AI-Based Mobile App for Assessing Speech Difficulties in Children. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(1), pp. 137–159. https://doi.org/10.3991/ijim.v20i01.57991

Article submitted 2025-08-01. Revision uploaded 2025-11-15. Final acceptance 2025-11-16.

dimensions such as fluidity, prosody, and intonation [2]. The opportunities to enrich oral language, especially in vocabulary and speed of nomination, are decisive for reading and communication development during childhood [3]. Phonological deficits, articulatory inaccuracy, and low prosodic awareness are usually associated with difficulties in pronunciation and oral expression [4]. Such alterations can be manifested in the form of errors in reading decoding, inappropriate pauses, or limited tonal variability, impacting comprehension and verbal expression [5]. Likewise, infantile dysphonia, characterized by alterations in timbre, intensity, or vocal height, can be influenced by physical, environmental, or psychological factors, extending throughout development if not attended to on time [6]. The literature warns that dysphonia and other infantile voice disorders are often underestimated, delaying diagnosis and increasing the risk of affecting psychosocial development and infant well-being [7]. Therefore, these vocal alterations demand validated strategies for their detection and evaluation to mitigate their consequences in the school and family environment [8], [9].

In response to these challenges, artificial intelligence (AI) has demonstrated a growing impact in solving real problems, particularly in audio classification and speaking in educational and health contexts; however, the development of robust child voice recognition systems faces specific challenges such as environmental noise, the scarcity of labeled data, and the variability inherent in children's voices [10]. The automatic analysis of voice signals by means of machine learning has allowed the early identification of children with linguistic alterations, making it possible to implement timely therapeutic interventions [11]. In addition, the integration of neurotechnology and AI in the school environment facilitates the detection of speech difficulties and the timely referral to specialists [12]. Strategies such as deep learning about voice spectrograms and data augmentation techniques increase accuracy in the detection of language disorders [13]. Likewise, AI applied to health has enhanced the personalization of interventions and the early detection of speech and language alterations, especially when integrated with augmentative communication technologies [14], [15].

Mobile applications, for their part, have been consolidated as key tools to support the evaluation and intervention of children with communication disorders, allowing the integration of voice recognition technologies and personalized feedback both in the clinical and educational settings [16], [17]. These applications offer flexibility, enable the personalization of exercises to reinforce linguistic skills, and promote inclusion in school contexts [18], [19]. At a functional level, the accessibility and communicative support provided by these applications are fundamental for the expression and understanding of children with speech difficulties [20], [21]. Technologies such as voice recognition in real-time support educational inclusion, transforming oral discourse into text for students with communication difficulties [22]. In addition, the use of smartphones to record and analyze the voice is objective and reliable, facilitating therapeutic follow-up and expanding intervention opportunities for families and professionals [23]. In addition to the use of these technologies in the monitoring and detection of speech problems, it is necessary to point out that the generation of synthetic voices through advanced architectures allows simulating-controlled attributes of child speech and validating diagnostic technologies in scenarios where the availability of real data is limited [24]. The use of statistical models has enabled the creation of useful synthetic databases for testing and training, although challenges persist in the perception of naturalness, age, or gender [25]. In addition, voice

conversion techniques allow the simulation of pathological voices, preserving the identity of the speaker, even with a few original samples, and facilitate the design of relevant clinical simulations [26], [27].

In this context, the main objective of the present study is to functionally validate a mobile application based on generative artificial intelligence (GAI) for the automatic evaluation of children's speech difficulties. The mobile application is designed to automatically analyze child voice recordings, assign scores to dimensions of fluency, pronunciation, and intonation, and deliver interactive visual feedback and tailored recommendations based on acoustic patterns. For this, an experimental design was developed that included the generation of 160 audio samples produced by 16 children's voices, each reading 10 short texts with intentional phonetic and fluid errors. The validation was carried out through the analysis of the correlation between the scores automatically granted by the application in the dimensions of fluids, pronunciation, intonation, and the acoustic indicators extracted from each audio, applying parametric statistics and spectral analysis. The proposed four-phase framework positions this work as an initial step in the validation of AI-based tools for children's speech, with the potential to guide future research in multilingual and cross-national contexts, contributing to global debates on mobile health and AI-assisted learning (refer to Section 2 for the conceptual framework). Therefore, the following research question (RQ) was posed: To what extent do the scores generated by the GAI-based mobile application (with respect to fluency, pronunciation, and intonation) correspond to the acoustic indicators derived from the children's voice recording?

2 CONCEPTUAL FRAMEWORK

This section explores two core conceptual axes that lay the theoretical and methodological groundwork for employing synthetic data in the initial validation of the mobile application at the heart of this study. By prioritizing controlled testing, this strategy paves the way for safely introducing real children's voices in later evaluation stages, ensuring both reliability and ethical responsibility.

2.1 Validation in mobile health applications

Validating mobile health applications calls for a phased progression that puts technical functionality, security, and usability first-long before involving real users. This method upholds quality standards while curbing risks, drawing from established global benchmarks such as the V3 Validation Framework and World Health Organization (WHO) guidelines, which champion transparency, privacy, and dependability in digital health tools [28]–[31]. Early on, it makes sense to lean on simulated data or controlled setups, advancing to human trials only once the app has shown a solid technical, ethical, and security footing. Complementary international frameworks, such as A-MARS, DECIDE-AI, mERA, CONSORT-AI, and SPIRIT-AI, outline essential criteria for quality, security, transparency, and methodological rigor in health apps [32]–[36]. Though they don't dictate a fixed sequence, these guidelines collectively point toward beginning in isolated environments with synthetic or simulated data. This gradual alignment with standards-prior to human involvement

or clinical testing-ultimately builds a trustworthy and secure tool that truly serves its users.

2.2 Synthetic data in the analysis of children's voices

Building automated tools to analyze children's speech isn't straightforward; it requires vast datasets and rigorous testing in safe, controlled spaces to lock in accuracy, safety, and real-world effectiveness before stepping into clinical settings [37]. Systems such as Tacotron-2, a text-to-speech (TTS) model, shine here with their lifelike naturalness and clarity, making them perfect for preliminary experiments that sidestep the need for actual voices right away [38]. Fresh research backs this up, showing how AI for speech therapy can kick off with simulated data to fine-tune precision and stability—keeping kids out of the loop until the tech is solid [39]. Take the Pronuntia system as a real-world example: it started validation with therapists and caregivers, focusing on rock-solid reliability before bringing in young participants [40]. Advanced models such as FastPitch take this further, crafting convincing child voices that power AI training and testing without tapping into real data, all while trimming ethical concerns and saving resources [41]. Studies in automatic speech recognition (ASR) echo this, revealing how early synthetic inputs ramp up model efficiency, cut dependency on genuine recordings, and hold onto accuracy as things scale [42]. Blending synthetic data with pre-trained models yields voices so realistic they're on par with natural speech in understanding, paving a secure path toward real-world use [43]–[44].

In essence, these elements frame synthetic data as a cornerstone for early-stage work—not just for honing the system's technical edge, but for upholding ethical standards by shielding vulnerable groups such as children from unproven tools [45]. Frameworks such as ACCEPT-AI reinforce this, aligning synthetic approaches with bedrock principles such as autonomy, beneficence, and data protection to foster thoughtful AI growth [46]. Ultimately, this mindful integration honors non-maleficence and privacy, letting innovation unfold responsibly without rushing exposure to real voices [47].

3 MATERIALS AND METHODS

3.1 Material

For this study, a mobile application developed in Android Studio was used, specifically designed for the automatic evaluation of alterations in children's Spanish-language speech, with potential for use by teachers and language therapists in educational and clinical contexts. The architecture of the application integrates several Google Firebase services: "Firebase Authentication" for secure user access, "Firestore" for the structured management of records and student data, and "Firebase Storage" for the encrypted storage of collected and processed audio files. The core of intelligent processing resides in the incorporation of the Gemini 2.5 model (Google AI), responsible for analyzing audio, assigning scores, and generating personalized feedback based on indicators of fluency, pronunciation, and intonation. For automated learning and evaluation, the application implements a rubric adapted from the DELE instrument (Diploma of Spanish as a Foreign Language), allowing

the AI to score each sample on a scale of 1 to 4 according to the degree of alteration or performance observed. The mobile application presents a clear and interactive interface, structured into four main functional modules, as illustrated in Figure 1a:

- Management of students: Allows the registration, selection, and individualized monitoring of participants or students (see Figure 1b).
- Audio management: Facilitates the recording, loading, and organization of audio files, showing the evaluation status of each sample (see Figure 1c).
- Evaluation with AI: Performs automatic processing of each audio, issuing a score for fluency, pronunciation, and intonation, along with textual feedback and recommendations for improvement.
- Results and feedback: It presents the history of evaluations, the feedback generated, and a summary table of the scores obtained for each evaluated audio, allowing the longitudinal monitoring of progress.

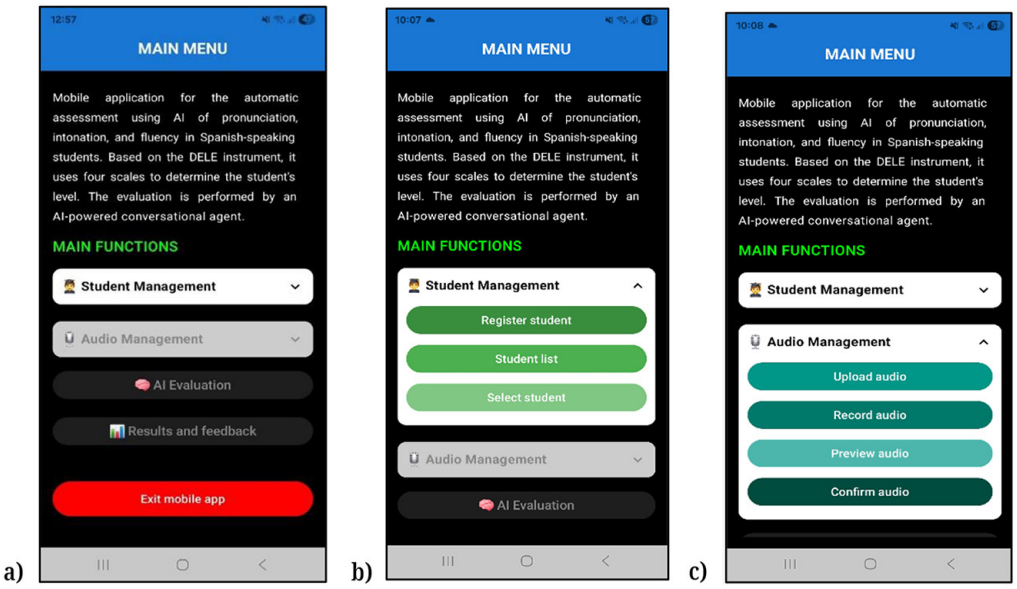

Fig. 1. Mobile application for the evaluation using GAI: (a) Main interface, (b) Student management interface, (c) Audio management interface

3.2 Method

This section describes the methodological procedures followed in the study, organized around three key components. First, the experimental design and analytical strategy applied to evaluate the performance of the generative AI-based mobile application during Phase 1 are detailed. Second, the methodological transparency section describes in detail the sequence of steps followed in the generation, validation, and acoustic analysis of the synthetic data, ensuring reproducibility and technical consistency of the study. Finally, the ethics statement specifies the inclusive use of non-identifiable synthetic audio and outlines the institutional approval and parental consent protocols planned for subsequent phases, in compliance with international standards for digital health research. Finally, the ethics statement specifies the inclusive use of non-identifiable synthetic audio and outlines the institutional approval

and parental consent protocols planned for subsequent phases, in compliance with international standards for digital health research.

Functional validation process: Before describing in detail the methodology used in the first phase, Figure 2 presents the proposed conceptual framework for the progressive validation of the mobile application to identify speech difficulties in children. This model contemplates four sequential phases that guarantee an orderly transition from controlled experimental environments to real-world mass deployment scenarios, ensuring the technical robustness, usability, clinical accuracy, and long-term sustainability of the system. However, this study is limited exclusively to Phase 1, which is the first step before involving real participants in subsequent phases.

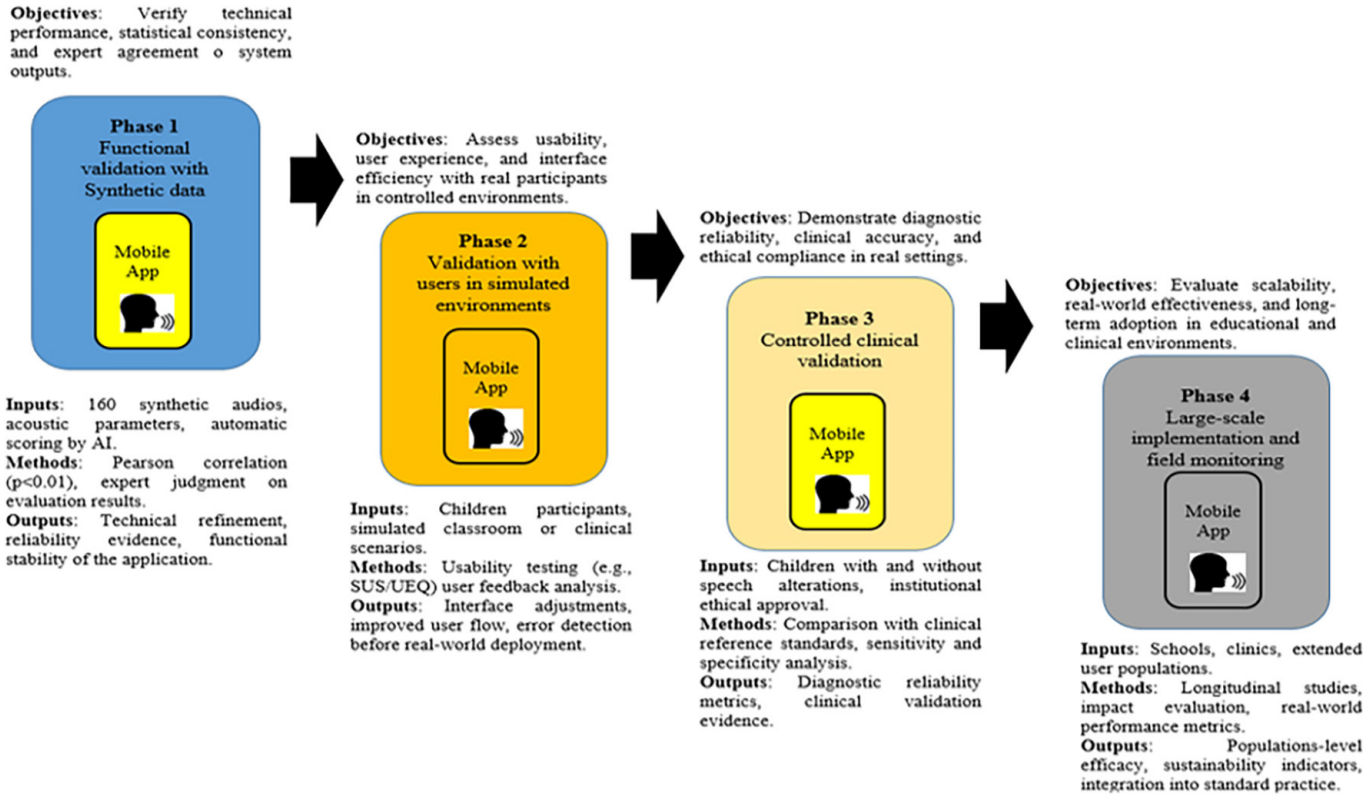

Fig. 2. Proposed conceptual framework for the progressive validation of the mobile application for identifying speech difficulties in children

Phase 1 – focuses on examining the technical performance, operational stability, and consistency of the results generated by the mobile application using synthetic data under controlled experimental conditions. The goal is to ensure the reliability of the automatic processing of acoustic parameters and the assignment of scores using AI before moving to direct human interaction contexts.

Phase 2 – aims to evaluate the user experience, usability, and efficiency of the system with real participants in simulated educational or clinical settings. Figure 3 shows the functional architecture and integration of environments for the first phase validation process.

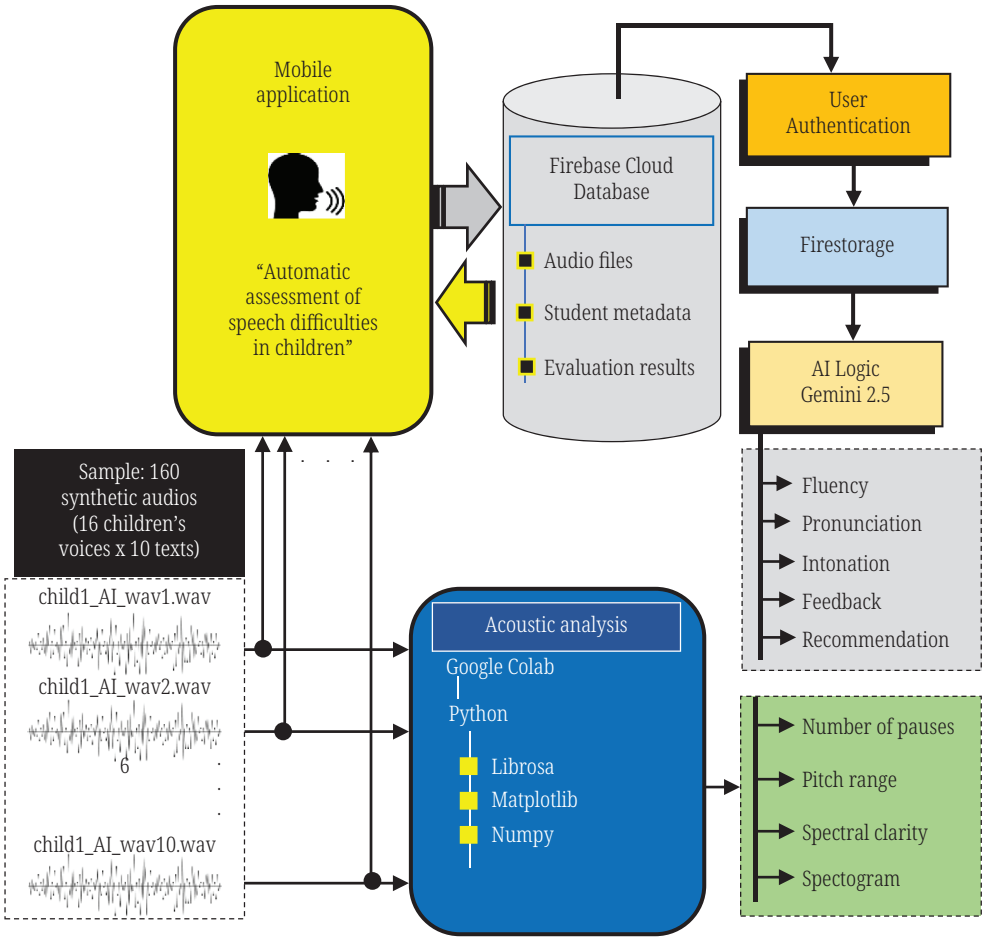

Fig. 3. Functional architecture and integration of environments for the first phase validation process

This stage allows for the detection of potential interface limitations, optimization of interaction, and ensuring that the application is intuitive and secure prior to clinical validation. In Phase 3, the mobile application will be evaluated in real-life clinical settings with children, following strict ethical protocols approved by institutional committees. The goal is to verify diagnostic accuracy and validity through comparisons with reference standards and sensitivity and specificity metrics. Finally, Phase 4 contemplates the adoption of the application in real-life educational and clinical settings, with longitudinal studies to assess its impact, sustainability, and scalability, as well as its potential integration into routine professional practice.

Methodological transparency of Phase 1: In this study, validation focused exclusively on Phase 1, using a set of 160 synthetic audio clips generated from 16 AI-created children's voices. Each voice was crafted with ten texts that intentionally incorporated phonetic errors, pauses, and intonation variations to simulate a representative range of speech difficulties in Spanish-speaking children. The audio clips, with an average duration of 18 to 24 seconds, were processed by the mobile application and statistically analyzed using normality tests and Spearman correlations ($p < 0.01$) to corroborate the consistency of the results obtained. The flowchart of the Phase 1 performance evaluation process is shown in Figure 4, which summarizes the sequential steps and key metrics for greater reproducibility.

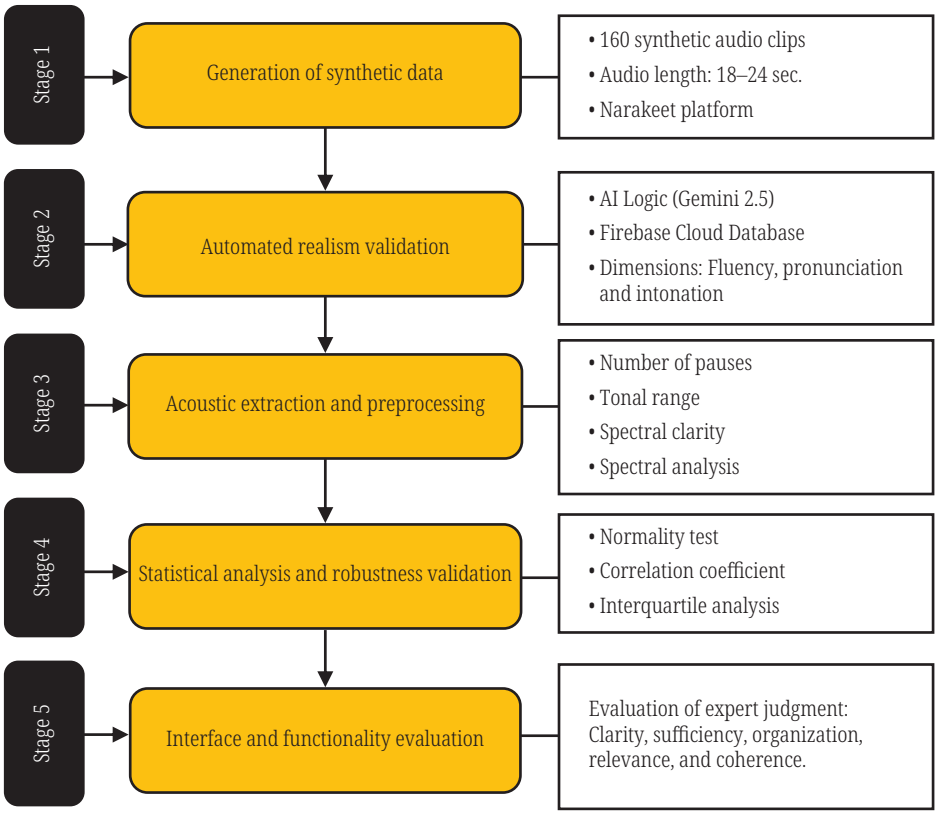

Fig. 4. Workflow of the Phase 1 performance evaluation process

In Stage 1 (Synthetic Data Generation), the voices were created using the Narakeet platform, which converts text into natural speech using neural network-based speech synthesis services such as Google Cloud TTS and Amazon Polly. Each text included controlled errors (omissions, misplaced pauses, and prosodic variations) to produce realistic child-like speech.

In Stage 2 (Automated Realism Validation), the samples were processed by the AI Logic module (Gemini 2.5), integrated with Firebase Cloud Database, which generated automatic scores in three dimensions: fluency, pronunciation, and intonation. These scores served as an indirect measure of linguistic realism and coherence, as signals with obvious synthetic errors tended to obtain low scores, while more natural ones had higher scores. To validate sample realism heuristically, preliminary acoustic checks were conducted using metrics such as pitch range (Fundamental-Frequency [Fo] bandwidth ≈ 3700–4100 Hz, reflecting spectral extent rather than mean Fo) and harmonics-to-noise ratio (HNR ≈ 20–25 dB) via *librosa* in Google Colab. Only those samples that met these reference ranges, derived from benchmarks in the literature for child voice synthesis, were retained. This confirmed that the Narakeet output's perceptual naturalness is suitable for simulating child variability and avoids robotic artifacts, as demonstrated in synthetic voice generation for pediatric applications. Both studies reported comparable HNR and pitch patterns for synthetic child speech, supporting their use as validation benchmarks. This ensured the clips mimicked authentic prosody while retaining controlled errors for assessment, consistent with the spectrogram variability observed in Figure 7 (e.g., frequency density and decibel distribution reflecting pause-induced amplitude drops), enhancing ecological plausibility under controlled Phase 1 conditions.

In Stage 3 (acoustic extraction and preprocessing), the validated audio samples were analyzed in Google Colab using Python libraries (Librosa, Matplotlib, NumPy) to extract acoustic indicators of pause count, pitch range, and spectral clarity, after amplitude normalization and silent segment removal. Additionally, visual spectral analyses were performed to illustrate signal variability and complement quantitative interpretation. The selection of acoustic indicators responds to their established relevance in pediatric speech assessment: the number of pauses reflects fluency and temporal flow, capturing disruptions in verbal continuity [11]; pitch range captures articulatory stability and phonetic variations, indicating intonation control and expressive range [5]. These metrics were prioritized for their non-invasiveness, computational feasibility in mobile contexts, and alignment with DELE rubric dimensions, enabling direct correlations with GAI outputs—as evidenced in Table 1 (e.g., mean pauses 34.95, pitch range 3,843.85 Hz, spectral clarity 0.034) and Figure 7 showing amplitude patterns for 41, 26, and 35 pauses, respectively.

Stage 4 (Statistical Analysis and Robustness Validation) focused on verifying the reliability of the data and the sensitivity of the AI model. To do this, before performing correlations, the normality of the variables was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. The results indicated that the scores generated by the mobile app (fluency, pronunciation, and intonation) did not follow a normal distribution ($p < 0.01$), while some acoustic indicators did follow a normal distribution. Additionally, a descriptive analysis of the interquartile range (IQR) was performed to address potential outliers and assess dispersion. The descriptive analysis showed that the scores generated by the mobile application showed moderate variability in fluency (IQR = 2) and slight variability in intonation (IQR = 1), while the dispersion in pronunciation was minimal (IQR = 0), indicating homogeneity in the evaluation of this indicator. These values reflect that the AI model was able to differentiate difficulty levels primarily in fluency, maintaining stability in the pronunciation rating. Likewise, the acoustic indicators processed in Google Colab showed controlled dispersion patterns consistent with the study objectives. The number of pauses showed the greatest variability (IQR = 11), demonstrating substantial differences in fluency between the synthetic audios. The word rate per five seconds (IQR = 1.91) and tonal range (IQR = 40.6) showed moderate dispersion, reflecting slight heterogeneity in articulation rate and intonation. Spectral clarity (IQR = 0.004) remained stable, confirming consistency in the quality of the synthesis. Overall, these results corroborate that the signals included sufficient variability to evaluate the model's sensitivity without compromising the homogeneity of the data set. For this reason, Spearman's correlation coefficient (ρ) was applied, which is appropriate for variables with non-normal distributions and monotonic relationships. These coefficients ($p < 0.01$) were calculated between the scores generated by AI and the acoustic indicators, obtaining significant correlations in fluency, intonation, and pronunciation.

In Stage 5 (Interface and Functionality Evaluation), a panel of 24 experts in linguistics and educational technology evaluated the application using an instrument structured in five dimensions: clarity, adequacy, organization, relevance, and coherence, to validate its practical applicability. The results showed moderately high levels of satisfaction in all dimensions, with minor recommendations for improvements to the interface and user orientation. Although ecological validity remains limited due to the exclusive use of synthetic data, future phases 2 to 4 will incorporate real-life child recordings, institutional ethics approval, and parental consent, in accordance with international standards of ethics in digital health.

Ethics Statement: This study did not involve human participants. All analyses were conducted using non-identifiable synthetic audio samples generated using

AI-based speech synthesis. No personal or clinical data were processed. Future validation phases (Phases 2–4) will include real children's voices and will strictly adhere to institutional ethical review procedures, informed parental consent, and data protection regulations, in full compliance with the V3 Validation Framework and WHO digital ethics standards.

4 RESULTS

The mobile application demonstrated its core functionality through the automatic evaluation and feedback of children's speech, as illustrated in the application's interfaces. Firstly, the results and feedback module allows users to view the historical records of audios evaluated by the AI. For each selected audio, the app displays the assigned fluency score, alongside textual feedback and a personalized recommendation to support improvement (see Figure 5a). This interface enables longitudinal monitoring of each participant's performance, providing both quantitative scores and qualitative feedback in an accessible format. Secondly, the application presents specific feedback for the intonation dimension (see Figure 5b).

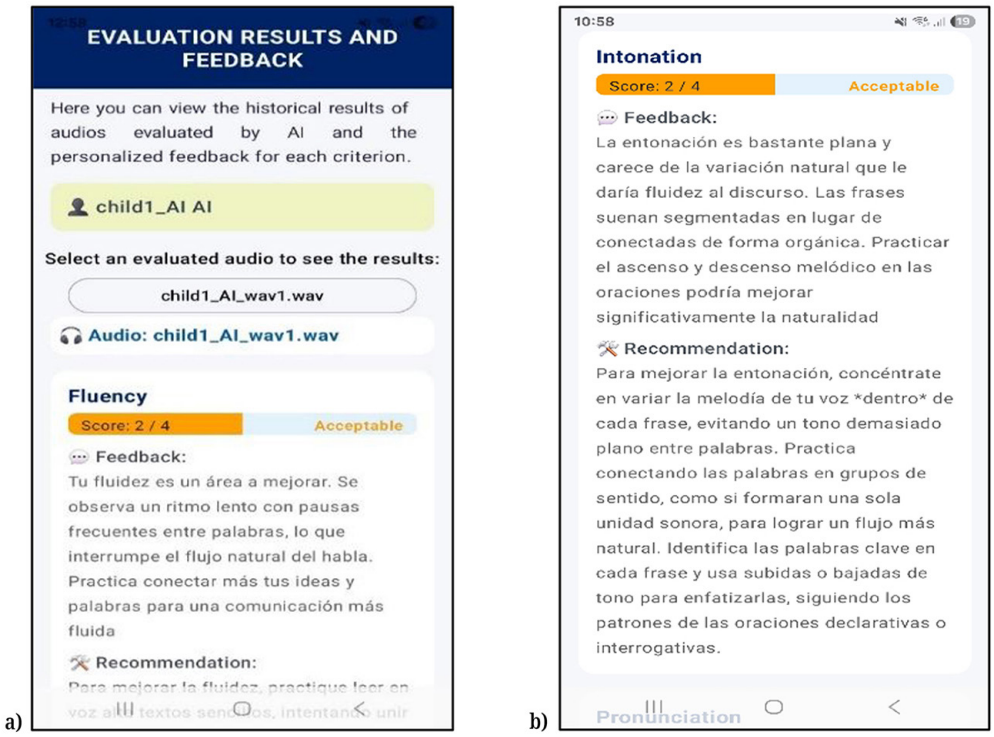

Fig. 5. Evidence of the automatic assessment in the mobile application: (a) Fluency and (b) Intonation

After automatic processing, users receive not only a numerical score for intonation but also detailed observations about the expressiveness, prosodic variation, and naturalness of the speech. Personalized recommendations are included to guide practice and improvement, addressing the particular prosodic challenges detected by the AI. Thirdly, for the pronunciation dimension, the mobile application generates an automatic score, feedback, and specific recommendations based on the AI's analysis of the audio (see Figure 6a). Users receive explicit advice focused on

articulatory aspects and pronunciation accuracy, tailored to the errors or strengths identified in each recording. This detailed interface supports targeted interventions for pronunciation improvement, promoting individualized language development. Finally, the audio evaluation history interface (see Figure 6b) summarizes the scores assigned for fluency, pronunciation, and intonation across all evaluated audio files for a given student. This consolidated table facilitates comprehensive progress tracking, allowing teachers, therapists, or families to visualize performance trends and monitor the impact of interventions over time.

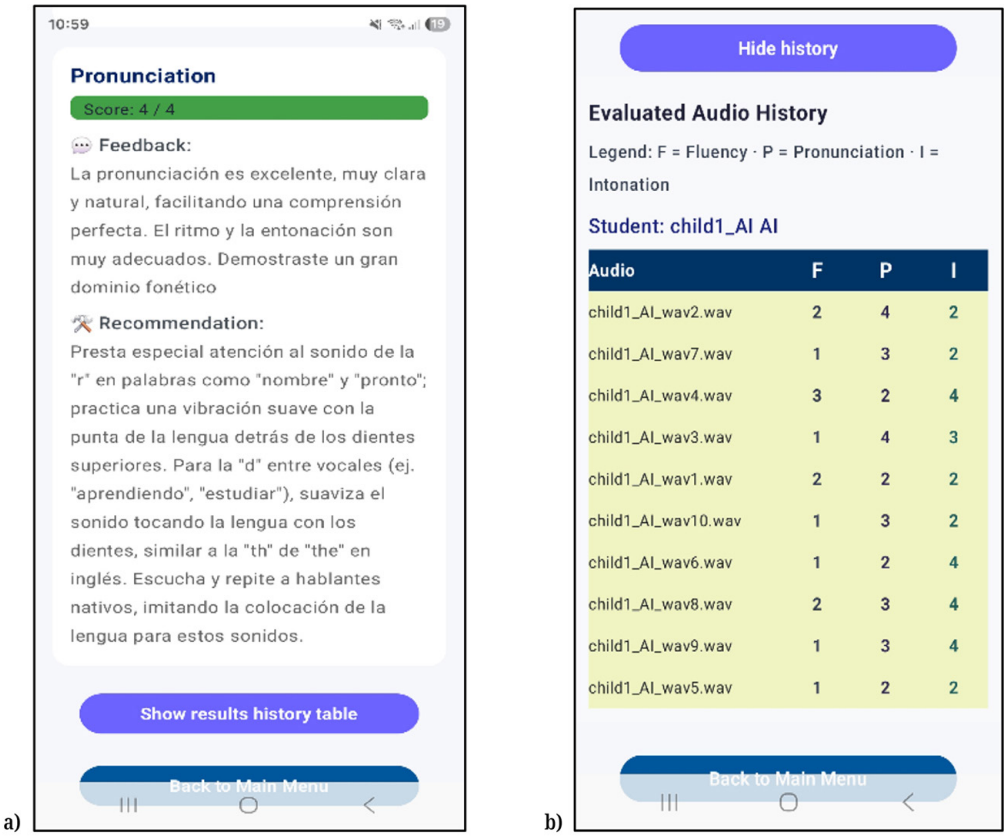

Fig. 6. Evidence of the automatic assessment in the mobile application: (a) Pronunciation and (b) Audio evaluation history

4.1 Descriptive characterization of the data collected from the validation indicators

The descriptive results obtained show the variability and heterogeneity in the performance of synthetic children's voice samples, intentionally created to represent multiple difficulty levels in speech parameters. In this study, a total of 160 evaluations were processed, offering a controlled and representative view of errors and correct productions, simulating the conditions expected in real-life automatic voice disorder detection. The scores calculated by the GAI for the indicator's fluency, pronunciation, and intonation ranged between 2.3 and 2.5 on a 1–4 scale, with standard deviations from 0.6 to 0.7. This pattern confirms that the dataset included both samples with severe errors and others with acceptable performance, ensuring variability for analysis. Regarding fluency-related acoustic indicators, the average number of pauses

per audio reached 34.95, while the mean word rate every five seconds was 10.726, revealing heterogeneous fluency profiles designed to test the application's sensitivity to different difficulty levels. For pronunciation, the pitch range showed minimal variability, suggesting that the synthetic audios preserved relative stability in articulation parameters and enabled the distinction between clearer and less accurate pronunciations. Regarding intonation, spectral clarity ranged from 0.031 to 0.041, reflecting subtle variations in the quality of the synthesized prosody and confirming that the samples offered sufficient variability to evaluate the application's discriminatory capacity. Table 1 shows the results of the descriptive analysis of the validation indicators of the audio evaluation obtained from both the mobile application and the acoustic analysis.

Table 1. Descriptive analysis of the synthetic audio evaluation indicators

Validation Indicators	Mean	Standard Deviation	Minimum	Maximum
Fluency score	1.96	0.788	1	4
Pronunciation score	2.07	0.728	1	4
Intonation score	2.18	0.823	1	4
Number of pauses	34.95	6.916	18	50
Words per 5 seconds	10.726	1.277	8.050	14.300
Pitch range	3809.358	27.148	3730.00	3875.00
Spectral clarity	0.034	0.002	0.031	0.041

4.2 Preliminary performance evaluation of the mobile application through score correlations with acoustic indicators

To evaluate the preliminary performance of the mobile app in assessing children's speech fluency, we calculated Spearman correlation coefficients between the app-generated fluency score and the acoustic indicator of detected pauses. One hundred and sixty synthetic audio samples were analyzed, providing sufficient variability and sample size to justify the use of statistical parameters. First, a strong negative correlation ($\rho = -0.784$, $p < 0.01$) was found between the fluency score and the number of pauses. As the number of pauses increased, the fluency score assigned by the AI decreased. This result confirms that the application reliability penalizes fragmented or discontinuous speech patterns. Table 2 presents the results of the correlation analysis between the fluency score and the number of pauses.

Table 2. Correlation analysis between the fluency score and number of pauses

			Fluency Score	Number of Pauses
Spearman's Rho	Fluency score	Correlation coefficient	1.000	–0.790**
		Sig. (bilateral)	–	0.000
		N	160	160
	Number of pauses	Correlation coefficient	–0.790**	1.000
		Sig. (bilateral)	0.000	–
		N	160	160

Note: **p-Value < 0.01.

To assess the pronunciation capabilities of the mobile applications, we compared the pronunciation scores generated by the AI with the pitch range derived from analyzing 160 synthetic children's voice samples. We chose pitch range as the key metric, as research highlights its role in spotting articulatory challenges and gauging phonetic accuracy in children's voices. The analysis revealed a strong positive correlation ($\rho = 0.737$, p-value < 0.01) between the mobile application effectively picking up on voices with better articulatory clarity and tonal variation, aligning with samples that show wider pitch ranges. Overall, these findings confirm the application's reliability in identifying phonetic difficulties. Table 3 summarizes the correlation between the pronunciation scores and pitch range.

Table 3. Correlation between pronunciation score and pitch range

			Pronunciation Score	Pitch Range
Spearman's Rho	Pronunciation score	Correlation coefficient	1.000	0.737**
		Sig. (bilateral)	–	0.000
		N	160	160
	Pitch range	Correlation coefficient	0.737**	1.000
		Sig. (bilateral)	0.000	–
		N	160	160

Note: **p-Value < 0.01.

Regarding the intonation dimension, this was performed by correlating the intonation score provided by the generative AI and the acoustic indicator of spectral clarity. The results show a very high positive correlation ($\rho = 0.858$, $p < 0.01$) between the intonation score provided by the mobile application and the spectral clarity values obtained from the acoustic analysis of the 160 synthetic audios. Table 4 shows the results of the correlation analysis between the intonation score and the acoustic indicator of spectral clarity.

Table 4. Correlation between intonation score and spectral clarity

			Intonation Score	Spectral Clarity
Spearman's Rho	Intonation score	Correlation coefficient	1.000	0.858**
		Sig. (bilateral)	–	0.000
		N	160	160
	Spectral clarity	Correlation coefficient	0.858**	1.000
		Sig. (bilateral)	0.000	–
		N	160	160

Note: **p-Value < 0.01.

In addition to the quantitative correlations obtained, a visual spectral analysis was performed on three synthetic audio signals corresponding to the same participant to demonstrate the variability of the acoustic indicators and complement the interpretation of the results.

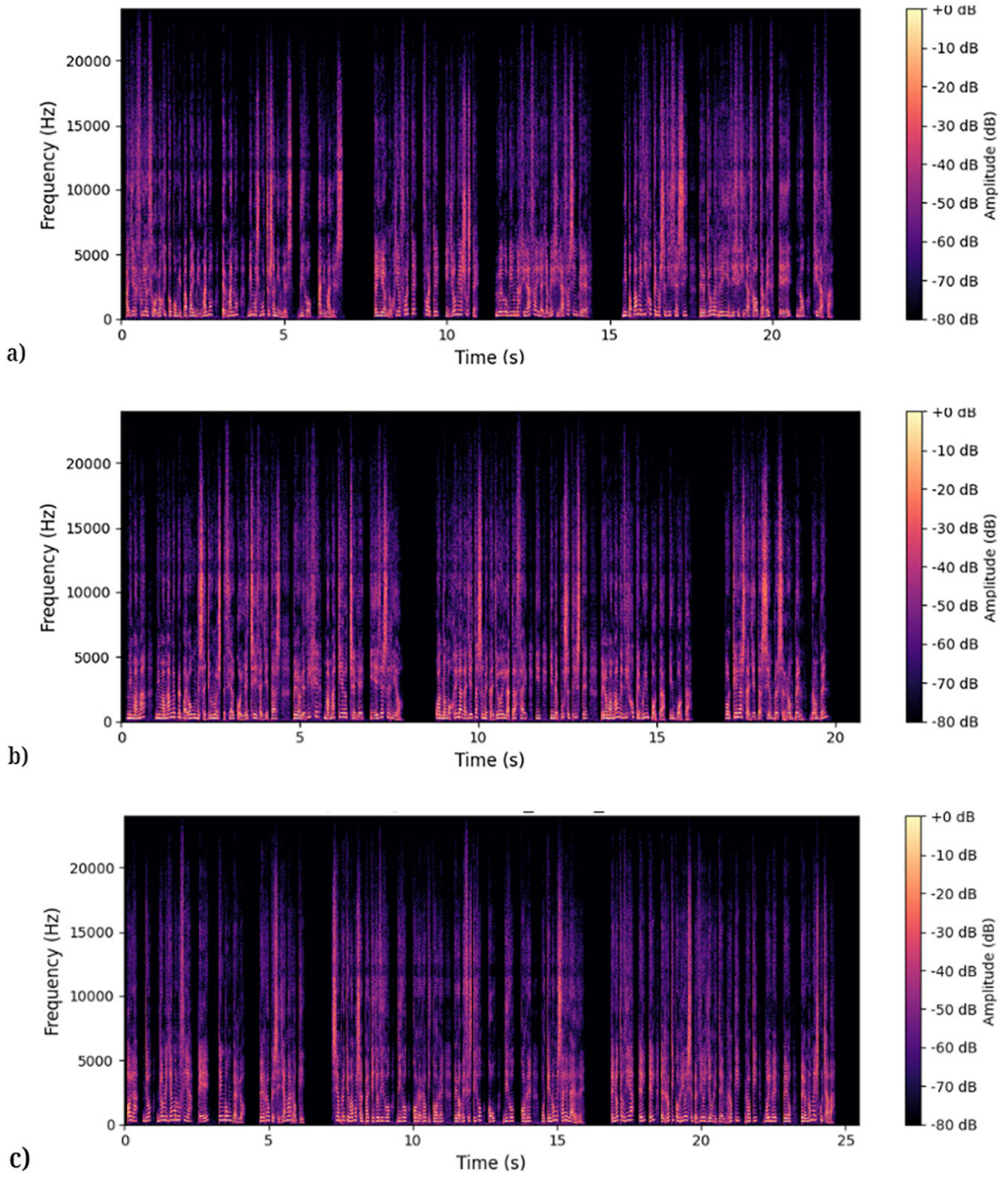

Fig. 7. Spectrograms of three synthetic audio signals analyzed for the same participant, (a) with 41 numbers of pauses, (b) with 26 numbers of pauses, and (c) with 35 numbers of pauses

Figure 7 shows the three spectrograms, which show marked variability in the density and distribution of frequencies, as well as the decibel levels over time. These spectrograms are consistent with the numerical results obtained in the correlation analysis. In Figure 7a, a higher number of pauses (41) and noticeably less dense areas, where low amplitude (decibel) intervals are frequent, can be seen. This reflects reduced fluency, as well as lower intonation and voice strength, also evidenced by the word rate (11.98/5 sec.), the pitch range (3,843.0), and the spectral clarity (0.033). In Figure 7b, the number of pauses is shown to decrease (26), showing greater continuity and areas of higher decibel intensity, which represent a more sustained voice with greater intonation. This pattern is accompanied by a higher word rate (12.34/5 sec.) and greater spectral clarity (0.036). Figure 7c, on the other hand, presents an intermediate number of pauses (35) and regions of moderate acoustic energy, with variations in decibel levels that indicate changes in voice strength

and modulation. The word rate was 13.30/5 sec, the pitch range was 3,843.2, and the spectral clarity was 0.034, representing a somewhat more fluent and intonated performance, although still with irregularities. These observations allow us to visualize how the differences in quantitative indicators—number of pauses, word rate, pitch range, spectral clarity—and amplitude patterns (decibels) are reflected in the structure of the spectrograms. Thus, the greater or lesser presence of high-amplitude areas and frequency continuity is related to the voice strength and intonation of the signal. Therefore, the correlation findings and the spectrogram representations allow us to demonstrate, in a first phase, the functional validity of the mobile application to detect variations in speech parameters in synthetic children's voice samples.

4.3 Expert judgment on application functionality and usability

A panel of 24 experts, composed of linguistic instructors and specialists in educational technology, evaluated the mobile applications' functionality and usability. The assessment instrument included five dimensions: clarity, sufficiency, organization, pertinence, and coherence. Each dimension was rated on a 5-point Likert scale (1 =Very poor, 2 = Poor, 3 = Fair, 4 = Good, 5 = Excellent). The results showed consistently high mean scores across all dimensions: clarity (4.7), sufficiency (4.6), organization (4.8), pertinence (4.7), and coherence (4.6). However, experts identified areas for improvement, particularly suggesting more explicit user guidance and further refinement of certain interface instructions. These findings confirm the applications' overall strengths while highlighting opportunities to enhance user experience in future iterations. Table 5 shows the results of the expert judgment evaluation by indicator.

Table 5. Result of the expert judgment evaluation

Criterion	Mean	Standard Deviation	Minimum	Maximum
Clarity	4.7	0.4	4	5
Sufficiency	4.6	0.5	4	5
Organization	4.8	0.4	4	5
Pertinence	4.7	0.4	4	5
Coherence	4.6	0.5	4	5

5 DISCUSSION

The findings of this study confirm the reliability and internal consistency of the mobile application, showing significant correlations between GAI-generated scores and classical acoustic indicators (pauses, pitch range, and spectral clarity) commonly used in pediatric speech evaluation. Although this Phase 1 limits ecological generalizability, the progressive validation framework provides a roadmap for future multilingual and cross-cultural studies, including the adaptation of the system to non-Spanish languages. These results indicate that the tool can become a practical alternative for automated assessment in clinical and educational settings, providing objective and reproducible analyses. Consistent with this, the study in [40] reported that mobile applications can detect alterations in children's voices through acoustic

parameters, although their work focused on binary classification. Our approach goes further by integrating multidimensional scoring and personalized feedback, offering a more comprehensive assessment of speech difficulties. Likewise, the research in [41] demonstrated that mobile technologies enable remote monitoring and continuous feedback, but their study centered on therapeutic support rather than on the automated correlation of acoustic parameters, an aspect specifically addressed in our study. Similarly, the work in [42] highlighted that automatic scoring systems and interactive spectrogram visualization not only improve objectivity but also enhance user engagement and progress tracking, findings aligned with the principles of our proposal. In the same vein, the investigation in [43] confirmed the usefulness of mobile applications for remote and daily measurement of acoustic indicators, showing the feasibility of evaluating parameters such as jitter using smartphones. This reinforces the role of mobile technologies and automated algorithms in the functional assessment of children's voices. Additionally, the analysis in [44] found that acoustic voice analysis with mobile applications can achieve reliability levels comparable to traditional systems, even under uncontrolled conditions. Our study expands this scope by validating not only fluency but also intonation and pronunciation, while integrating personalized recommendations enabled by generative artificial intelligence.

The proposed four-phase framework ensures a methodical progression from synthetic data validation to real-world deployment and outlines a scalable and ethical AI protocol. Specifically, Phase 1's baseline correlations between GAI scores and acoustic indicators—demonstrated here with Spanish synthetic samples—serve as a foundation for Phases 2 and 3, where real children's voices from diverse linguistic backgrounds (e.g., English or Portuguese in bilingual classrooms) can be tested in simulated and clinical settings.

This progression enables fine-tuning of the DELE-based rubric—embedded in the GAI prompt—for phonetic adaptations in Romance languages, while incorporating alternative rubrics (e.g., CEFR [Common European Framework of Reference for Languages] or language-specific prosodic frameworks) for tonal or non-Indo-European languages such as Mandarin or Hindi, thereby facilitating Phase 4's longitudinal evaluation in multicultural educational programs. In low-resource educational settings and multilingual clinical environments, pilot implementations can assess usability and compare AI outputs with clinician judgments, ultimately informing global standards for AI-assisted speech assessment and reducing diagnostic disparities across languages.

Other works have explored complementary dimensions. For example, the research in [45] showed that mobile applications can improve vocal self-perception and self-assessment, consistent with our findings on the motivational value of automated feedback for speech practice. In terms of acoustic robustness, the research in [46] indicated that combining multiple parameters allows the differentiation of pathological voice in children, noting that greater dispersion in values reflects reduced phonatory stability-a pattern also observed in our application, particularly in synthetic samples simulating severe dysphonia. Moreover, the work in [47] demonstrated that including real-time error correction and individualized pronunciation analysis modules strengthens user self-regulation and facilitates more effective remediation. This supports the importance of multidimensional feedback for continuous improvement, as proposed in our system. Similarly, the investigation in [48] emphasized that, although mobile applications achieve results comparable to standard clinical techniques for some acoustic parameters, further work is required in methodological standardization, device diversity, and usability factors

that represent opportunities for future research and the development of more robust tools. Finally, the work in [49] confirmed that mobile applications can measure acoustic levels with high precision in clinical environments when using calibrated microphones, although accuracy decreases with very low sound levels or uncalibrated equipment. This finding underlines the need to consider technical constraints in the clinical validation phase of our application. Complementing this, the findings in [50] highlighted the benefits of combining computer-assisted analysis with real-time feedback and automatic error correction, reinforcing the importance of robust algorithms and user-centered designs for maximizing educational impact.

From an educational technology viewpoint, the mobile application presents viable pathways for incorporation into everyday teaching routines and early intervention activities. In classroom settings, instructors could integrate it for systematic audio evaluations during expressive language exercises, using the participant oversight module to document progress and generate summaries aligned with individualized instructional goals. Such deployment would allow prompt, interface-based guidance on articulation precision and rhythmic variation, facilitating rapid adaptation in collaborative tasks and creating inclusive spaces that provide targeted support while maintaining natural group dynamics. To further promote equality, the tool could incorporate adaptive prompts that account for sociolinguistic variability, such as regional accents or code–switching in multilingual classrooms, thereby supporting diverse learner profiles and reducing biases in AI-driven assessments by modeling specific language varieties through sociolinguistically-informed corpora [51]. Such adaptations may also leverage bilingual advantages in executive functions, as bilingual children consistently outperform monolinguals on attention and inhibition tasks, enhancing cognitive flexibility in diverse learning environments [52].

These implementations align with the subsequent phases of the validation framework, extending the app's use from controlled synthetic trials to authentic classroom and therapeutic contexts. This approach echoes evolving instructional paradigms powered by AI that promote early detection of needs and adaptive learning across diverse groups [53]. Likewise, within early therapeutic programs, practitioners might embed the application's rhythmic appraisal into engaging, activity-based sessions, combining data-informed feedback with expert review to accelerate verbal development and sustain continuous monitoring [54]. By merging automated insights with educator input, these mechanisms not only extend speech evaluation to underserved areas but also empower teachers to integrate health-related observations into the core curriculum, fostering inclusive language growth and reducing disparities in access to speech support.

Beyond its local validation scope, the mobile tool has strong potential for cross-linguistic scalability. The DELE-based rubric, integrated into the mobile app, can be reprogrammed to assess speech parameters in other languages, such as English, Portuguese, or French, by adjusting phonetic rules and prosodic parameters within the AI layer. This flexibility allows for the creation of culturally and linguistically responsive speech assessment modules adaptable to diverse contexts. Furthermore, the cloud infrastructure and mobile interface ensure transferability to low-resource settings, where access to specialized speech therapy remains limited. Lightweight AI inference and Firebase synchronization enable offline use and integration into school and community programs. From a policy perspective, this approach aligns with international initiatives such as the UNESCO Digital Learning Framework and the WHO guidelines for children's communication development, offering a scalable pathway for equitable access to AI-supported speech assessment and early intervention. To address sociolinguistic variability, future adaptations could integrate models

trained on diverse dialects and translinguistic practices, mitigating potential biases in prosodic detection for non-standard varieties and enhancing equity in multilingual classrooms, particularly for low-SES learners in global south contexts via context-focused NLP (natural language processing) to bridge vocabulary gaps and foster cultural awareness [55]. Recent advances in AI-edtech underscore this potential, with studies showing that inclusive language models improve speech perception flexibility in linguistically diverse children, positioning our tool within the emerging landscape of equitable AI for learning as greater input diversity promotes gradient categorization over rigid boundaries [56]. Automated tools such as those reviewed here further support this by enabling scalable analysis of daylong recordings in underrepresented cultural and clinical contexts, addressing gaps in linguistic diversity [57].

6 CONCLUSION

The results of this study indicate that the GAI-based mobile application demonstrates functional validity for the automated assessment of fluency, pronunciation, and intonation in children's speech. Significant correlations were identified between the generated scores and acoustic indicators, including a strong inverse relationship between fluency score and number of pauses ($\rho = -0.790$), direct correlations between pronunciation score and pitch range ($\rho = 0.737$), and between intonation score and spectral clarity ($\rho = 0.858$). Descriptive and spectrographic analyses further revealed mean decibel levels ranging from 15 to 33 dB, speech rates of 11.98 to 13.30 words per five seconds, pitch ranges around 3,843 Hz, and spectral clarity values between 0.033 and 0.036. The visual inspection of spectrograms supported these quantitative findings, illustrating consistent trends between acoustic energy distribution and the assigned scores. Additionally, expert judgment confirmed the applications' high levels of clarity, sufficiency, organization, pertinence, and coherence, supporting their usability and practical relevance in diverse settings. Therefore, it is concluded that this tool may be useful for the objective, multidimensional evaluation and monitoring of speech difficulties in clinical, educational, and family contexts, given its alignment with established acoustic parameters.

7 LIMITATIONS OF THE STUDY

A main limitation of this study is the exclusive use of synthetic samples and the validation in a controlled experimental setting, which restricts the generalization of the results to real-life scenarios with greater variability. Future research should evaluate the application with natural child voice recordings in a clinical environment, assess its performance across diverse devices and conditions, and explore user acceptance by examining perceived ease of use and usefulness. Efforts to standardize evaluation metrics and interfaces are also recommended to further enhance its clinical and educational relevance.

8 DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors did not use generative AI or AI-assisted technologies for writing, editing, or content creation. All text, figures,

and tables were produced manually by the authors through collaborative academic work. Limited use of Grammarly was employed solely for basic grammar and spelling checks, without altering the original content or generating new material. The authors take full responsibility for the integrity and accuracy of the content.

9 REFERENCES

[1] N. T. Vidal and E. P. Espinoza, "The prevention of voice disorders in the fifth-grade students of the primary education," *Opuntia Brava*, vol. 11, no. 1, pp. 1–15, 2023. https://opuntiabrava.ult.edu.cu/index.php/opuntiabrava/article/view/693/653

[2] N. C. Ruiz, S. J. Jiménez, M. C. Q. Martínez, and Y. Solovieva, "The correction of psycho-pedagogical reading difficulties in Spanish," *Avances en Psicología Latinoamericana,* vol. 37, no. 2, pp. 361–374, 2019. https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.6628

[3] S. P. Suggate, "A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions," *Journal of Learning Disabilities*, vol. 49, no. 1, pp. 77–96, 2016. https://doi.org/10.1177/0022219414528540

[4] I. Vásquez-Venegas, H. León-Valdés, and K. Sáez-Carrillo, "Prosodic performance in reading aloud in children with autism spectrum disorder," *Estudios filológicos*, vol. 72, pp. 91–110, 2023. https://doi.org/10.4067/S0071-17132023000200091

[5] N. Jordán, F. Cuetos, and P. Suárez-Coalla, "Prosody in the reading of children with specific language impairment," *Infancia y Aprendizaje*, vol. 42, no. 1, pp. 87–127, 2019. https://doi.org/10.1080/02103702.2018.1550161

[6] L. Nercelles-Carvajal, N. Pizarro-Silva, and P. Sepúlveda-Torres, "Aggravating factors of dysphonia in preschool children: Differences between children with and without dysphonia," *Rev. Salud. Pública*, vol. 22, no. 5, pp. 486–490, 2020. https://doi.org/10.15446/rsap.v22n5.78180

[7] D. Centenod and M. Penna, "Characterization of patients with dysphonia evaluated in the pediatric voice unit of the Dr. Luis Calvo Mackenna Hospital," *Rev. Otorrinolaringol. Cir. Cabeza Cuello*, vol. 79, no. 1, pp. 18–24, 2019. https://revistaotorrino-sochiorl.cl/index.php/orl/article/view/633

[8] D. Centeno, L. Nercelles, J. Valenzuela, and C. Catalán, "Descriptive analysis of the pediatric voice handicap index in children with benign vocal fold pathology," *Rev. Otorrinolaringol. Cir. Cabeza Cuello*, vol. 82, no. 1, pp. 16–20, 2022. https://revistaotorrino-sochiorl.cl/index.php/orl/article/view/5

[9] F. Betances and H. Vallés, "Prevalence of infantile dysphonia in the school of early childhood and primary education 'Ensanche' of teruel," *Rev. Otorrinolaringol. Cir. Cabeza Cuello*, vol. 79, no. 1, pp. 159–166, 2019. https://revistaotorrino-sochiorl.cl/index.php/orl/article/view/606

[10] O. O. Abayomi-Alli, R. Damaševicius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: A systematic review," *Electronics*, vol. 11, no. 3795, pp. 1–32, 2022. https://doi.org/10.3390/electronics11223795

[11] M. K. Reddy, P. Alku, and K. S. Rao, "Detection of specific language impairment in children using glottal source features," *IEEE Access*, vol. 8, no. 1, pp. 15274–15279, 2020. https://doi.org/10.1109/ACCESS.2020.2967224

[12] I. Sindhu and M. S. Sainin, "Automatic speech and voice disorder detection using deep learning—A systematic literature review," *IEEE Access*, vol. 2, pp. 49667–49698, 2024. https://doi.org/10.1109/ACCESS.2024.3371713

[13] J. Su and W. Yang, "Artificial intelligence in early childhood education: A scoping review," *Computers and Education: Artificial Intelligence*, vol. 3, no. 1, pp. 1–13, 2022. https://doi.org/10.1016/j.caeai.2022.100049

[14] A. Bhardwaj, M. Sharma, S. Kumar, S. Sharma, and S. P. Chander, "Transforming pediatric speech and language disorder diagnosis and therapy: The evolving role of artificial intelligence," *Health Sciences Review*, vol. 12, no. 1, pp. 1–7, 2024. https://doi.org/10.1016/j.hsr.2024.100188

[15] S. Papadakis, S. H. Lytvynova, S. M. Ivanova, and I. A. Selyshcheva, "Advancing lifelong learning with AI-enhanced ICT: A review of 3L-Person 2024," in *CEUR Workshop Proceedings (9th International Workshop on Professional Retraining and Life-Long Learning using ICT: Person-Oriented Approach—3L-Person 2024)*, Lviv, Ukraine, 2023. https://ceur-ws.org/Vol-3781/paper00.pdf

[16] J. McKechnie, B. Ahmed, R. Gutierrez-Osuna, P. Monroe, P. McCabe, and K. J. Ballard, "Automated speech analysis tools for children's speech production: A systematic literature review," *International Journal of Speech-Language Pathology*, vol. 1, no. 1, pp. 1–17, 2018. https://doi.org/10.1080/17549507.2018.1477991

[17] M. Sarosa, D. Febiyanti, and H. Darmono, "Design and implementation of voice time, time indicator application for diabetic retinopathy patients," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, no. 2, pp. 144–159, 2020. https://doi.org/10.3991/ijim.v14i02.11436

[18] Z. Shi, D. Chung, Y. Du, J. Zhang, S. Raina, and M. Mataric, "Is AI ready to support speech therapy for children? A systematic review of AI-enabled mobile Apps for pediatric speech therapy," in *Interaction Design and Children (IDC '25)*, Reykjavik, Iceland. ACM, New York, NY, USA, 2025, pp. 479–493. https://dl.acm.org/doi/10.1145/3713043.3728841

[19] L. Bilic, M. Ebner, and M. Ebner, "A voice-enabled game based learning application using Amazon's echo with Alexa voice service: A game regarding geographic facts about Austria and Europe," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, no. 3, pp. 226–232, 2020. https://doi.org/10.3991/ijim.v14i03.12311

[20] D. Glista, R. O'Hagan, M. Servais, and N. Jalilian, "Adolescent-centered mHealth applications in a collaborative care model: A virtual focus group study with audiologists," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 8, pp. 2794–2810, 2024. https://doi.org/10.1044/2024_JSLHR-23-00679

[21] A. Zainuddin, N. A. Zubir, N. A. Aminuddin, N. D. Khirul Ashar, and M. E. Mahadan, "Appliance control with IOT-Arduino of voice command detection for mobility impaired people," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 23, pp. 164–177, 2021. https://doi.org/10.3991/ijim.v15i23.22147

[22] R. D. B. Contreras-Manrique, L. Contreras-Manrique, and A. M. Figueroa-Hernández, "Inclusion of students with hearing disabilities through the ListenApp mobile application," *Ingeniería y Competitividad*, vol. 24, no. 1, pp. 1–18. https://revistaingenieria.univalle.edu.co/index.php/ingenieria_y_competitividad/article/view/11070

[23] S. N. Awan, R. Bahr, S. Watts, M. Boyer, R. Budinsky, Bridge2AI Voice Consortium, and Y. Bensoussan, "Validity of acoustic measures obtained using various recording methods, including smartphones with and without headset microphones," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 1, pp. 1712–1730, 2024. https://doi.org/10.1044/2024_JSLHR-23-00759

[24] V. W. Zhang, A. Sebastian, and J. J. M. Monaghan, "Automated speech intelligibility assessment using AI-based transcription in children with cochlear implants, hearing aids, and normal hearing," *Journal Clinical Medicine*, vol. 14, no. 15, p. 5280, 2025. https://doi.org/10.3390/jcm14155280

[25] F. Albedah, "Artificial intelligence in language education: A systematic review of multilingual applications, large language models, and emerging challenges," *Language Teaching Research Quarterly*, vol. 49, pp. 247–268, 2025. https://doi.org/10.32038/ltrq.2025.49.13

[26] T. Zhang, X. Liu, G. Liu, and Y. Shao, "PVR-AFM: A pathological voice repair system based on non-linear structure," *Journal of Voice*, vol. 37, no. 5, pp. 648–662, 2021. https://doi.org/10.1016/j.jvoice.2021.05.010

[27] S. Papadakis *et al.*, "Embracing digital innovation and cloud technologies for transformative learning experiences," in *Proceedings of the 11th Workshop on Cloud Technologies in Education (CTE 2023)*, Kryvyi Rih, Ukraine, 2023. https://ceur-ws.org/Vol-3679/paper00.pdf

[28] S. Lagan, P. Aquino, M. R. Emerson, K. Fortuna, R. Walker, and J. Torous, "Actionable health app evaluation: Translating expert frameworks into objective metrics," *npj Digital Medicine*, vol. 3, no. 100, pp. 1–8, 2020. https://doi.org/10.1038/s41746-020-00312-4

[29] A. Deniz-Garcia *et al.*, "Quality, usability, and effectiveness of mhealth apps and the role of artificial intelligence: Current scenario and challenges," *Journal of Medical Internet Research*, vol. 4, no. 25, pp. 1–26, 2023. https://doi.org/10.2196/44030

[30] Q. Grundy, "A review of the quality and impact of mobile health apps," *Annual Review of Public Health*, vol. 43, no. 1, pp. 117–134, 2021. https://doi.org/10.1146/annurev-publhealth-052020-103738

[31] World Health Organization, "Global strategy on digital health 2020–2025," 2021. https://www.who.int/publications/i/item/9789240020924

[32] A. E. Roberts *et al.*, "Evaluating the quality and safety of health-related apps and e-tools: Adapting the mobile app rating scale and developing a quality assurance protocol," *Internet Interventions*, vol. 24, no. 1, pp. 1–13, 2021. https://doi.org/10.1016/j.invent.2021.100379

[33] B. Vasey *et al.*, "Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI," *Nature Medicine*, vol. 28, pp. 924–933, 2022. https://doi.org/10.1038/s41591-022-01772-9

[34] S. Agarwal *et al.*, "Guidelines for reporting of health interventions using mobile phones: Mobile health (mHealth) evidence reporting and assessment (mERA) checklist," *BMJ*, vol. 352, no. 1, pp. 1–10, 2016. https://doi.org/10.1136/bmj.i1174

[35] X. Liu *et al.*, "Guidelines for reporting clinical trials of artificial intelligence interventions: CONSORT-AI extension," *Pam American Journal of Public Health*, vol. 48, no. 1, pp. 1–15, 2024. https://doi.org/10.26633/RPSP.2024.13

[36] S. C. Rivera *et al.*, "Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension," *Nature Medicine*, vol. 26, no. 1, pp. 1351–1363, 2020. https://doi.org/10.1038/s41591-020-1037-7

[37] J. McKechnie, B. Ahmed, R. Gutierrez-Osuna, P. Monroe, P. McCabe, and K. J. Ballard, "Automated speech analysis tools for children's speech production: A systematic literature review," *International Journal of Speech-Language Pathology*, vol. 20, no. 6, pp. 583–598, 2018. https://doi.org/10.1080/17549507.2018.1477991

[38] Y. Win and T. Masada, "Myanmar text-to-speech system based on Tacotron-2," in *International Conference on Information and Communication Technology Convergence (ICTC),* 2020, pp. 578–583. https://doi.org/10.1109/ICTC49870.2020.9289599

[39] H. Suh, A. Dangol, H. Meadan-Kaplansky, C. A. Miller, and J. A Kientz, "Opportunities and challenges for AI-based support for speech-language pathologists," in *CHIWORK '24,* Newcastle upon Tyne, United Kingdom, 2024, pp. 1–14. https://doi.org/10.1145/3663384.3663387

[40] G. Desolda, R. Lanzilotti, A. Piccinno, and V. Rossano. "A system to support children in speech therapies at home," in *CHItaly '21,* Bolzano, Italy, 2021, pp. 1–5. https://doi.org/10.1145/3464385.3464745

[41] M. L. Farooq, D. Bigioi, R. Jain, W. Yao, M. Yiwere, and P. Corcoran, "Synthetic speaking children – Why we need them and how to make them," *Computer Science,* vol. 1, no. 11, pp. 1–6, 2023. https://arxiv.org/abs/2311.06307

[42] A. Fazel *et al.*, "SynthASR: Unlocking synthetic data for speech recognition," in *Interspeech 2021,* Brno, Czechia, 2021, pp. 896–900. https://doi.org/10.21437/Interspeech.2021-1882

[43] C. Terblanche, T. T. Schnoor, M. Harty, and B. V. Tucker, "The development of synthetic child speech in three South African languages," *Augmentative and Alternative Communication*, vol. 1, no. 1, pp. 1–13, 2024. https://doi.org/10.1080/07434618.2024.2374312

[44] V. Giannouli and M. Banou, "The intelligibility and comprehension of synthetic versus natural speech in dyslexic students," *Disability and Rehabilitation: Assistive Technology*, vol. 1, no. 1, pp. 1–10, 2019. https://doi.org/10.1080/17483107.2019.1629111

[45] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 310, pp. 1–9, 2020. https://doi.org/10.1186/s12911-020-01332-6

[46] V. Muralidharan, A. Burgart, R. Daneshjou, and S. Rose, "Recommendations for the use of pediatric data in artificial intelligence and machine learning ACCEPT-AI," *npj Digital Medicine*, vol. 6, no. 166, pp. 1–6, 2023. https://doi.org/10.1038/s41746-023-00898-5

[47] S. Y. Chng *et al.*, "Ethical considerations in AI for child health and recommendations for child-centered medical AI," *npj Digital Medicine*, vol. 8, no. 152, pp. 1–10, 2025. https://doi.org/10.1038/s41746-025-01541-1

[48] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "Voice disorder detection via an m-health system: Design and results of a clinical study to evaluate Vox4Health," *BioMed Research International*, vol. 1, no. 1, pp. 1–18, 2018. https://doi.org/10.1155/2018/8193694

[49] C. R. Doarn *et al.*, "Design and implementation of an interactive website for pediatric voice therapy-the concept of in-between care: A telehealth model," *Telemed. J. E. Health.*, vol. 25, no. 5, pp. 415–422, 2019. https://doi.org/10.1089/tmj.2018.0108

[50] O. Chamorro-Atalaya *et al.*, "Voice analytics for the identification of university student satisfaction, from whatsapp audio messaging," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 21, pp. 219–227, 2023. https://doi.org/10.3991/ijet.v18i21.39073

[51] E. U. Grillo, "An online telepractice model for the prevention of voice disorders in vocally healthy student teachers evaluated by a smartphone application," *Perspect ASHA Spec Interest Groups*, vol. 2, no. 3, pp. 63–78, 2017. https://doi.org/10.1044/persp2.SIG3.63

[52] S. Fujimura *et al.*, "Real-time acoustic voice analysis using a handheld device running android operating system," *Journal of Voice*, vol. 34, no. 6, pp. 823–829, 2020. https://doi.org/10.1016/j.jvoice.2019.05.013

[53] E. V. Leer and N. Porcaro, "Feasibility of the fake phone call: An iOS app for covert, public practice of voice technique for generalization training," *Journal of Voice*, vol. 33, no. 5, pp. 659–668, 2019. https://doi.org/10.1016/j.jvoice.2018.02.014

[54] R. Martínez-Olalla *et al.*, "Analysis of voice quality in children with Smith-Magenis syndrome," *Journal of Voice*, vol. 40, no. 1, pp. 1–11, 2024. https://doi.org/10.1016/j.jvoice.2024.09.026

[55] Y. Jin, "Design of students' spoken English pronunciation training system based on computer VB platform," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 6, pp. 41–52, 2019. https://doi.org/10.3991/ijet.v14i06.10154

[56] J. Grieve *et al.*, "The sociolinguistic foundations of language modeling," *Front. Artif. Intell.*, vol. 7, no. 1, pp. 1–18, 2025. https://doi.org/10.3389/frai.2024.1472411

[57] A. Yurtsever, J. A. E. Anderson, and J. G. Grundy, "Bilingual children outperform monolingual children on executive function tasks far more often than chance: An updated quantitative analysis," *Developmental Review*, vol. 69, no. 1, pp. 1–20, 2023. https://doi.org/10.1016/j.dr.2023.101084

10 AUTHORS

Maritza Arones holds a Bachelor's degree in Educational Sciences with a specialization in Mathematics and Physics, and a PhD in Education from the National University "San Luis Gonzaga" in Ica, Peru (E-mail: marones@unica.edu.pe).

Irma Aybar-Bellido holds a PhD in Education and is a Professor at the National University of "San Luis Gonzaga" in Ica, Peru (E-mail: irma.aybar@unica.edu.pe).

Willy Adauto-Medina holds a Bachelor's degree in Language and Writing, a Master's degree in Communication Didactics. He is a Professor at the National Technological University of Lima South, Lima, Peru (E-mail: wadauto@untels.edu.pe).

Santiago Rubiños-Jimenez holds a PhD in Engineering and is an Electrical Engineer. He is a Professor at the National Technological University of Lima South, Lima, Peru (E-mail: srubinos@untels.edu.pe).

José Antonio Arévalo-Tuesta holds a PhD in Administration and Education and a Master's degree in Educational Management. He is a Professor at the National University Federico Villarreal, Lima, Peru (E-mail: jarevalotu@unfv.edu.pe).