Blue Voice：LLM-Powered Speech Rehabilitation System for Children with Hearing Loss

Yibo Meng, Qiaoren Wang, Bingyi Liu, Huidi Lu

Submitted to: JMIR Serious Games
on: September 02, 2025

Table of Contents

Blue Voice?LLM-Powered Speech Rehabilitation System for Children with Hearing Loss

Yibo Meng[1*]; Qiaoren Wang[2*]; Bingyi Liu[3*]; Huidi Lu[4]

[1]Tsinghua University Tsinghua University Beijing CN
[2] National University of Singapore Singapore SG
[3] University of Michigan, Ann Arbor Ann Arbor US
[4]Tsinghua University University of Oxford Oxford OX1 1HP, UK GB
[*]these authors contributed equally

Corresponding Author:
Huidi Lu
Tsinghua University
University of Oxford
Oxford OX1 1HP, UK
Oxford OX1 1HP, UK
GB

Abstract

Background: Children with hearing loss face significant barriers in speech and language acquisition, especially during critical developmental years. While computer-assisted instruction and gamified training have shown promise, many systems lack adaptability, personalization, and sustained engagement. Serious games, which merge educational goals with immersive game mechanics, have emerged as effective tools for rehabilitation and learning. Advances in large language models (LLMs) now provide opportunities to deliver semantic-level feedback and adaptive game-based content. However, empirical validation of LLM-driven serious games for speech rehabilitation remains limited.

Objective: This study aimed to design and evaluate Blue Voice, a home-based serious game for speech rehabilitation that integrates LLM-enabled real-time feedback, interactive training modules, and wearable sensing, with the goal of improving rehabilitation outcomes for children with hearing loss.

Methods: We adopted a mixed-methods approach. (1) Stakeholder interviews and surveys were conducted with 26 participants, including children with hearing loss, their parents, teachers, and speech therapists, to identify needs and barriers. (2) Two experimental studies were carried out. The pilot study involved 22 children between six and twelve years old to examine basic usability and interaction. A larger controlled experiment involved 22 children of the same age group. In this experiment, one group used the Blue Voice system with LLM-enabled real-time feedback, while the other group used the same system without LLM feedback. Usability metrics, engagement levels, and speech performance outcomes were collected. Quantitative data were analyzed using descriptive and inferential statistics, and qualitative feedback was thematically analyzed. Ethical approval was obtained from the institutional review board, and written informed consent was provided by the guardians of all participants.

Results: Stakeholder feedback indicated strong acceptance of home-based serious games for rehabilitation. In the pilot study with 22 children, usability ratings were high, with “ease of learning” scored at 5.0, “task completion without assistance” at 4.6, and “interface comprehension” at 4.6, while “feedback sufficiency” received a lower score of 3.8, suggesting areas for improvement. In the controlled experiment with 22 children, the experimental group using LLaMA-2–enabled real-time feedback achieved significantly higher speech accuracy in teacher evaluations (week 2: 4.7 ± 0.4 vs 4.1 ± 0.6, $t=2.73$, $P<.05$) and system assessments (week 2: 4.6 ± 0.3 vs 4.1 ± 0.5, $t=2.56$, $P<.05$) compared with the control group. Engagement ratings were also higher in the experimental group (4.8 ± 0.3) than in the control group (4.2 ± 0.5). Qualitative feedback highlighted that real-time semantic feedback improved children’s confidence, motivation, and expressive ability, while control group participants described the training as repetitive and less engaging.

Conclusions: Blue Voice demonstrates the potential of serious games in speech rehabilitation for children with hearing loss. By combining LLaMA-2 semantic feedback with game-based mechanics, the system not only improved speech accuracy but also fostered confidence for sustained engagement. The findings suggest that serious games enhanced with large language models can effectively complement traditional therapy and provide a scalable, child-centered digital rehabilitation solution.

(JMIR Preprints 02/09/2025:83431)
DOI: https://doi.org/10.2196/preprints.83431

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Blue Voice □ LLM-Powered Speech Rehabilitation System for Children with Hearing Loss

Abstract

Background:

Children with hearing loss face significant barriers in speech and language acquisition, especially during critical developmental years. While computer-assisted instruction and gamified training have shown promise, many systems lack adaptability, personalization, and sustained engagement. Serious games, which merge educational goals with immersive game mechanics, have emerged as effective tools for rehabilitation and learning. Advances in large language models (LLMs) now provide opportunities to deliver semantic-level feedback and adaptive game-based content. However, empirical validation of LLM-driven serious games for speech rehabilitation remains limited.

Objective:

This study aimed to design and evaluate Blue Voice, a home-based serious game for speech rehabilitation that integrates LLM-enabled real-time feedback, interactive training modules, and wearable sensing, with the goal of improving rehabilitation outcomes for children with hearing loss.

Methods:

We adopted a mixed-methods approach. (1) Stakeholder interviews and surveys were conducted with 26 participants, including children with hearing loss, their parents, teachers, and speech therapists, to identify needs and barriers. (2) Two experimental studies were carried out. The pilot study involved 22 children between six and twelve years old to examine basic usability and interaction. A larger controlled experiment involved 22 children of the same age group. In this experiment, one group used the Blue Voice system with LLM-enabled real-time feedback, while the other group used the same system without LLM feedback. Usability metrics, engagement levels, and speech performance outcomes were collected. Quantitative data were analyzed using descriptive and inferential statistics, and qualitative feedback was thematically analyzed. Ethical approval was obtained from the institutional review board, and written informed consent was provided by the guardians of all participants.

Results:

Stakeholder feedback indicated strong acceptance of home-based serious games for rehabilitation. In the pilot study with 22 children, usability ratings were high, with “ease of learning” scored at 5.0, “task completion without assistance” at 4.6, and “interface comprehension” at 4.6, while “feedback sufficiency” received a lower score of 3.8, suggesting areas for improvement. In the controlled experiment with 22 children, the experimental group using LLaMA-2–enabled real-time feedback achieved significantly higher speech accuracy in teacher evaluations (week 2: 4.7 ± 0.4 vs 4.1 ± 0.6, $t=2.73$, $P<.05$) and system assessments (week 2: 4.6 ± 0.3 vs 4.1 ± 0.5, $t=2.56$, $P<.05$) compared with the control group. Engagement ratings were also higher in the experimental group (4.8 ± 0.3) than in the control group (4.2 ± 0.5). Qualitative feedback highlighted that real-time semantic feedback improved children’s confidence, motivation, and expressive ability, while control group participants described the training as repetitive and less engaging.

Conclusions:

Blue Voice demonstrates the potential of serious games in speech rehabilitation for children with hearing loss. By combining LLaMA-2 semantic feedback with game-based mechanics, the system not only improved speech accuracy but also fostered confidence for sustained engagement. The findings suggest that serious games enhanced with large language models can effectively complement traditional therapy and provide a scalable, child-centered digital rehabilitation solution.

Keywords: Serious Game; Children with Hearing Loss; Speech Rehabilitation; LLM; CAI; Human-Computer Interaction

Introduction

China has the world's largest population of individuals with hearing disabilities, totaling approximately 20.54 million—accounting for over 30% of the country's disabled population—with more than 4.6 million being children. Hearing impairments in children often create significant barriers to speech and language development, especially during the critical period between ages 3 to 12, which is essential for cognitive and social growth. Research indicates that early identification and intervention, including amplification and structured therapy, are closely linked to improved linguistic outcomes for this population 1.

CAI has been widely studied as an effective tool to support children with special educational needs, including those with hearing loss. Notable projects have confirmed that CAI can significantly enhance academic skills such as language arts and mathematics while maintaining cost-effectiveness. Studies have shown that CAI-based strategies improve learning achievement and retention in students with hearing impairments when compared to traditional methods such as sign language-only instruction 2. Moreover, intensive speech–language therapy combined with technological support has been associated with marked improvements in children's syntactic and lexical development 3.

In parallel, serious games have emerged as a promising approach in both education and rehabilitation, offering structured learning experiences embedded in engaging, game-like environments. Unlike traditional digital tools, serious games incorporate motivational elements such as challenges, rewards, and interactive storytelling, which sustain children's attention and encourage repeated practice. Evidence suggests that serious games not only improve knowledge retention and transfer but also enhance adherence and motivation in long-term therapeutic contexts [5,6]. For children with hearing loss, serious games provide an accessible and enjoyable way to engage in repetitive training, which is often necessary for effective speech rehabilitation.

While CAI demonstrates clear educational benefits, recent advances in artificial intelligence—particularly LLM—offer powerful new possibilities. LLM provide natural language understanding, real-time feedback, and semantic interaction capabilities that can significantly enhance CAI environments. When integrated into speech rehabilitation systems, LLM allow for more personalized, adaptive, and engaging learning experiences for children with hearing loss 4.

This study proposes the design and evaluation of an LLM-powered speech rehabilitation system tailored for children with hearing loss. The system integrates LLM-driven modules within a CAI framework to deliver individualized instruction, error correction, and interactive speech-language content. Preliminary evaluations indicate the system's potential to improve engagement and speech outcomes. This research aims to contribute an innovative, AI-enhanced approach to pediatric speech therapy and to inform future development of intelligent assistive technologies in the field 16.

Related Work

Profile of children with hearing impairment

Hearing disability is defined as a condition in which an individual experiences bilateral hearing loss or auditory dysfunction, leading to difficulty or inability in perceiving surrounding sounds clearly. The severity of hearing impairment is typically categorized according to the average threshold of hearing loss, as outlined in the Table 1. Below 5.

Table 1. Classification of degree of hearing loss

Degree of hearing disability	Hearing disability level	Degree of hearing loss	Manifestations of hearing impairment without the aid of hearing aids
Severity	Hearing disability level 1	≥91 dB HL	Inability to rely on the sense of hearing for verbal communication, extreme limitations in activities such as comprehension and communication, and very serious obstacles to participation in social life.
↑	Hearing disability level 2	81- 90 dB HL	Severe limitations in activities such as understanding and communication, and serious barriers to participation in social life.
	Hearing disability level 3	61- 80 dB HL	Moderate limitations in activities such as understanding and communication, and moderate barriers to participation in social life.
Mild	Hearing disability level 4	41- 60dB HL	Mild limitations in activities such as understanding and communication, and mild impairment in participation in social life.

Auditory impairments significantly hinder children's language development, especially since language acquisition closely depends on early hearing ability. Although hearing aids and cochlear implants can improve auditory function, many hearing-impaired children still face delays in comprehension and expression due to reduced auditory input in early childhood. These challenges often arise not from speech production deficits, but from a lack of stimulation during critical developmental periods. To address this, experts like Liu Yong advocate for personalized and regularly adjusted speech rehabilitation programs tailored to each child's abilities to better support their cognitive and linguistic growth 21.

Analysis of design related to children with hearing impairment

In recent years, both domestic and international designers and researchers have made significant progress in the development of language training tools for children with hearing impairments. Many of these products integrate advanced technologies with user-centered design to enhance rehabilitation outcomes while making the learning process more engaging.

“Little Ear” is a language rehabilitation training tool designed specifically for children aged 3–6 with hearing impairments. It combines image recognition and speech recognition technologies to support early-stage language learning in a visual and interactive way. Similarly, COMMU, a pronunciation correction device developed by Kyuseok Lee and Hyunjin Kim, captures a user’s speech, guides them through the pronunciation process, and connects to a mobile phone that displays speech waveforms and transcribed content. The device uses AI algorithms to analyze speech clarity and provide real-time feedback. By allowing children to practice independently, COMMU reduces social anxiety often associated with speech training and reflects a thoughtful approach to user dignity and comfort.

These innovative products demonstrate how technology can be used not only to improve the efficiency of speech rehabilitation, but also to make the experience enjoyable and motivating. By emphasizing interactivity, emotional comfort, and professional feedback, these tools represent a promising direction for the future of language training for children with hearing loss.

Preliminary User Research

To understand the perspectives of stakeholders regarding speech rehabilitation training for children with hearing loss, identify practical barriers in training implementation, and inform the design of a home-based digital intervention system, we conducted a two-stage user research study involving surveys and interviews.

Participant Recruitment

We recruited 26 participants through a combination of school-based and institutional outreach, including 15 hearing-impaired children (aged 6–12) and their parents, 5 teachers from special education schools for the deaf, and 6 professional speech therapists from language rehabilitation centers. Participants were invited via direct contact with institutions and participated voluntarily. The study protocol was approved by the Institutional Review Board of Shandong University and adhered to the principles of the Declaration of Helsinki (2008 revision). Written informed consent was obtained from all participants or their legal guardians.

Questionnaire-Based Survey

In the first phase, participants completed a structured questionnaire designed to explore current practices and perceptions surrounding speech rehabilitation, with a focus on home-based training using mobile devices. The questionnaire comprised both closed and semi-open-ended questions, organized into four modules: (1) Basic demographic information (age, role, caregiving responsibilities) to establish user profiles; (2) Training behaviors (locations, frequency, involvement of professionals) to assess current routines; (3) Perceptions and attitudes (convenience, effectiveness, perceived obstacles) to reveal cognitive and motivational factors; and (4) Expectations and willingness (adoption of remote support, ideal support formats) to inform digital solution design. Questions included practical scenarios such as “Do you believe speech training at home with mobile devices is convenient?” and “What challenges do you foresee in home-based rehabilitation?”

The questionnaires were distributed and completed in a special education school environment, under the guidance of two trained researchers. Responses were collected on-site and subsequently coded, organized, and analyzed using Microsoft Excel. The average completion time was approximately 25 minutes per participant.

Interviews and Focus Group

To supplement the survey findings and explore individual experiences in greater depth, we conducted four semi-structured interviews with stakeholders representing diverse perspectives: a special education teacher, a speech-language therapist, and two parents of hearing-impaired children. All the interviews were conducted at the Special Education School of Zhong District, Jinan City. Interview questions addressed both demographic background and open-ended themes related to daily language learning, support needs, and technology acceptance. The qualitative data provided insights into the emotional, logistical, and pedagogical challenges encountered in language rehabilitation, particularly in non-institutional settings. All participants were informed of the study’s purpose and procedures, and their responses were anonymized to ensure confidentiality.

The user research revealed a strong interest in home-based training models, especially when supported by professional remote guidance and digital tools. These insights directly informed the definition of target user personas and the identification of system requirements for the proposed LLM-powered speech rehabilitation platform.

Research Findings

High Acceptance of Home-Based Rehabilitation

Survey results indicate a generally positive attitude toward home-based speech rehabilitation using mobile devices among all three participant groups. Among the children surveyed, 86.7% expressed a preference for practicing at home, and 80.0% believed that engaging in pronunciation exercises within a gamified environment was both easier and more enjoyable. Additionally, 73.3% reported that they would rather use a mobile phone or tablet for practice than attend in-person sessions, citing a more relaxed and comfortable home atmosphere that encourages active participation. Parents echoed this sentiment, with 86.7% agreeing that home-based training was convenient or highly convenient, emphasizing benefits such as flexible scheduling and time saved on transportation. Notably, 80.0% of parents were willing to support home-based training if remote professional guidance was available, and 66.7% pointed out the added benefit of enhanced parent-child interaction, which allows them to better understand their child’s language development. However, 33.3% of parents raised concerns, primarily regarding their ability to facilitate training effectively and the potential for distractions in the home environment. Among special education teachers, 80.0% acknowledged the feasibility of home-based training, viewing it as an effective supplement to classroom instruction, particularly beneficial for students with shorter attention spans or emotional variability. Teachers emphasized that a familiar environment could reduce stress and improve expressive frequency. However, they also cautioned that without standardized and professional guidance, there is a risk of reinforcing incorrect pronunciation patterns. All participating speech therapists supported the use of mobile-based home training, citing its potential to expand access to speech rehabilitation services and improve consistency in training, especially in the context of limited professional resources. Across all stakeholder groups, 92.3% considered mobile-assisted home rehabilitation convenient or highly convenient; 88.5% believed it would enhance therapeutic outcomes; and 84.6% expressed willingness to adopt such methods if professional remote support was provided. Open-ended responses highlighted key advantages including flexible timing, a familiar environment, and increased engagement, while major concerns included lack of parental expertise, child distraction, and device usability challenges.

Core Challenges in Language Acquisition

Interview findings identified a central challenge in the speech development of hearing-impaired children: the lack of effective and age-appropriate language training methods. Participants emphasized that language therapy goes beyond correcting phonetic or grammatical errors—it plays a vital role in fostering children's confidence and expressive abilities. Educators and therapists reported recurring classroom challenges such as poor phonemic awareness, limited attention spans, delayed cognitive development, and difficulty with emotional comprehension. Due to their auditory limitations, many children face significant difficulties in both articulation and expressive communication. As a result, some exhibit a reluctance to speak at all, which further exacerbates their communicative challenges. Interviewees also noted that cognitive delays are common among hearing-impaired children, who often require additional time and scaffolding to interpret their surroundings and articulate thoughts effectively. Experts highlighted the potential of AI to assist in this context by enabling more responsive, semantically accurate, and context-aware language interaction.

Opportunities and Constraints in Gamified Learning

Both parents and professionals expressed enthusiasm about gamified approaches to speech rehabilitation. Games were seen as highly engaging, developmentally appropriate, and well-aligned with children's digital habits. Respondents highlighted features such as visual interactivity, reward mechanisms, and the ability to present information in playful, intuitive ways. These aspects were believed to support sustained participation and improve language comprehension. However, stakeholders also raised critical design concerns: games must match users' developmental stages and be realistic and goal-oriented, but not overly complex. Excessive design sophistication or non-linear pathways may confuse children or increase the risk of overuse and distraction. Respondents advised that game structure should provide clear learning paths without relying on typical gaming conventions that encourage addiction or aimlessness.

Design Principles

The system should adopt gamification strategies tailored to the cognitive and emotional characteristics of hearing-impaired children, providing an engaging and intuitive training experience that encourages sustained participation; integrate LLM to generate age-appropriate pronunciation prompts, deliver semantically rich and context-sensitive feedback, and personalize training based on individual language development trajectories; ensure standardized and real-time speech correction feedback using speech recognition technologies aligned with professional norms to prevent reinforcement of incorrect pronunciation patterns; support remote guidance from certified speech-language therapists, including asynchronous evaluations and synchronous consultations, to ensure professional involvement while reducing dependency on institutional visits; offer multimodal interaction formats—such as voice, image, and animated instructions—to accommodate diverse learning styles and sensory needs; and optimize usability for the home environment through flexible training schedules, distraction-aware interaction design, and minimal setup complexity for parents and caregivers.

System Design

Functional Module

To address the limitations in continuous and personalized speech rehabilitation resources for children with hearing impairments, we developed a gamified home-based speech training system named Blue Voice. The system comprises two main components: a mobile application that delivers interactive speech training tasks and feedback, and a wearable smart wristband capable of real-time emotional state monitoring. The backend is powered by LLM and speech recognition technology. This system is specifically designed for children aged 6–12 with hearing impairments, aiming to support articulation and language expression in a low-stress home environment while enabling parents and teachers to monitor progress and provide timely intervention.

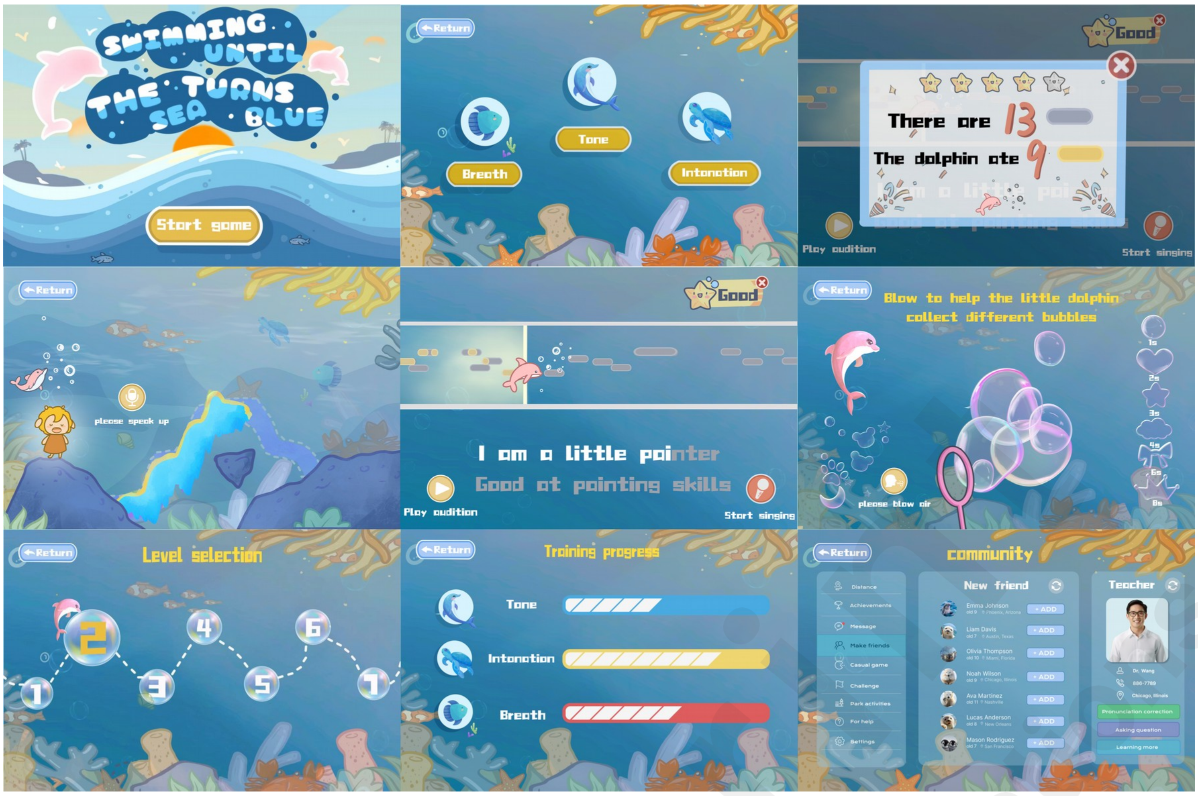
Speech Training Platform

The mobile application serves as the primary interface through which children engage in interactive speech rehabilitation. It is compatible with smartphones, tablets, and VR devices. The speech training modules are divided into three functional categories: *breath control training*, *pitch-rhythm training*, and *intonation modulation*, as summarized in Table 2. All training tasks are guided by a virtual character, "Dolly," modeled after the Chinese white dolphin. Users can control the game environment through voice commands, facilitating language learning in a playful, engaging format. These training modules are structured and interactive, aiming to meet the specific rehabilitation needs of children with hearing loss and enhance their speech capabilities through repeated, enjoyable practice 9.

Table 2. Training Modes and Function Descriptions

Mode Name	Description
Breath Control Training	The user blows into the microphone to control the size of on-screen bubbles. When sufficient breath is detected, the bubble bursts and points are awarded. This mode helps improve children's breath control for more stable and accurate articulation.
Pitch-Rhythm Training	The system selects a song of appropriate difficulty based on the child's current ability. Users sing to control a fish character that collects bubbles; better pitch and rhythm yield higher scores. Encouraging text prompts are displayed throughout the game to maintain motivation and engagement.
Intonation Modulation	The user adjusts vocal pitch to guide "Dolly" across a series of bridges. When the correct pitch is achieved, the bridge turns yellow and allows progression. Clearing all bridges results in level completion and experience rewards. This mode enhances pitch perception and modulation.

Fig.1 Blue voice: interface display of the speech rehabilitation training system for hearing-impaired children (the three modules from left to right in the middle are Intonation Modulation, Pitch-Rhythm Training and Breath Control Training respectively)



Smart Wristband

The smart wristband is worn on the child's wrist and integrates basic motion tracking and emotional recognition functions. It continuously captures behavioral data during training sessions, including heart rate variability, activity frequency, and concentration levels. These data are uploaded in real-time to the backend of the mobile application and used to dynamically adjust training intensity and pacing. For instance, if the system detects signs of emotional distress or reduced attention, it can automatically slow down task progression or switch to less demanding activities. Teachers can also remotely assess the child's training status through wristband feedback, enabling timely intervention and support *7*.

Intelligent Feedback and Content Generation Module

The system leverages the open-source large language model LLaMA-2 to evaluate children's pronunciation accuracy, prosody, and overall speech fluency in real time, providing semantic-level feedback and expert-like correction. The child's voice input is processed through speech recognition algorithms to assess both acoustic features and semantic alignment. The LLaMA-2 module analyzes the speech in real time and delivers targeted feedback and instructional guidance, thereby enhancing language expression and improving the clarity and fluency of verbal communication. All feedback is presented using child-friendly language and is accompanied by personalized practice prompts to ensure both professional rigor and motivational encouragement *10*.

System Implementation

Software Development The Blue Voice application adopts a cross-platform architecture, with its core training interface developed using the Unity engine. Unity's real-time rendering capabilities and flexible interaction design allow for the integration of rich visual elements and responsive voice-controlled components, making the system highly suitable for child-focused speech rehabilitation scenarios. To further support adaptive and intelligent feedback, the backend integrates the open-source large language model LLaMA-2, which enables real-time semantic analysis of children's speech and the generation of personalized instructional responses. The system consists of three main training modules, each implemented with distinct technical approaches.

(1) Breath Control Module

This module utilizes Unity's microphone input API to capture the user's continuous airflow signals. A custom algorithm analyzes the amplitude and duration of the airflow sound to determine whether it meets preset thresholds, which in turn drives the dynamic scaling and bursting of visual bubbles. The module is optimized for graphical frame rate and processing latency, ensuring immediate visual feedback during training sessions.

(2) Pitch-Rhythm Module

This module integrates third-party pitch detection libraries to extract frequency and rhythm features from the user's singing input. Through real-time mapping logic, the system translates pitch accuracy and rhythm alignment into in-game scoring mechanisms. The audio input is timestamped and used to trigger animation feedback and reward logic, maintaining tight synchronization between user performance and the background music.

(3) Intonation Control Module

Based on Fast Fourier Transform (FFT) algorithms, this module estimates the user's vocal pitch in real time and adjusts the vertical movement of the virtual character "Dolly" (a Chinese white dolphin) accordingly. A state machine manages the transition and color change of bridges in the game, ensuring compatibility with various speech speeds and reducing false positives during pitch detection.

The front-end includes a monitoring dashboard developed using native Android and iOS toolkits. This interface allows teachers and parents to view training logs, module-level performance, and participation trends. Speech recognition is implemented by invoking mainstream cloud service APIs, which return phonetic and semantic data for subsequent feedback generation 8.

Hardware Integration (1) Hardware Design

The Blue Voice system is equipped with a smart wristband that integrates a heart rate sensor and a six-axis motion sensor. These components collect basic physiological and behavioral data in real time during training. Data is transmitted via Bluetooth Low Energy (BLE) to the mobile device, where it is parsed and analyzed to calculate attention and stress indicators based on predefined thresholds and sliding window averages. For example, emotional arousal is assessed by computing heart rate variability over 30-second intervals, while attention level is inferred by analyzing motion frequency and periods of stillness. These indicators are used to dynamically adjust the pacing of training modules. All sensor data is structured using timestamped JSON packets and synchronized with the Unity-based training modules to enable real-time adaptive feedback based on the user's state *11*.

(2) Industrial Design

The wristband is designed using miniaturized sensor modules and medical-grade silicone housing. Its overall structure is lightweight and comfortable for children to wear and can remain securely attached to the wrist without interfering with hand movement or device operation. The design ensures both ergonomic fit and functional stability to support uninterrupted data collection during training.

System prototype testing and evaluation

The purpose of this assessment was to validate the usability and child-friendliness of the Blue Voice language training application, identify areas for improvement in interface design and interaction flow, and collect user feedback to inform subsequent optimization. The evaluation focused on ease of use, learning efficiency, and the appropriateness of the app's interface and gamification elements for children aged 6–12.

Experimental Design and Methodology

Participants We recruited 22 participants (female: 11; male: 11; average age: 8.2 ± 1.64) through both online and offline channels. Recruitment was conducted via direct outreach to special education schools and rehabilitation centers, as well as parent groups on WeChat. We first distributed digital flyers with study details, followed by phone contact or in-person communication with interested parents to confirm eligibility and obtain consent. Each participant received a compensation of 130 RMB upon completion of the evaluation session as a token of appreciation.

Experimental Flow Each child participated in an individual evaluation session. Researchers guided them through the use of the Blue Voice application without direct intervention during interaction. Children were allowed to freely explore and complete tasks within the app, including each of the three training modules. Researchers observed their interaction behavior and recorded performance across a set of predefined evaluation criteria. After the interaction, each child completed a brief structured usability questionnaire. The key focus areas of observation included interface clarity, navigation fluency, task completion ability, engagement with game elements, and responsiveness to feedback *17*.

Evaluation Metrics This study was approved by the Institutional Review Board of Shandong University. All procedures followed the ethical principles outlined in the Declaration of Helsinki (2008 revision). Written informed consent was obtained from the parents or legal guardians of all participating children before enrollment in the study. Children were informed in age-appropriate language about the purpose of the research and were free to withdraw at any time without consequences.

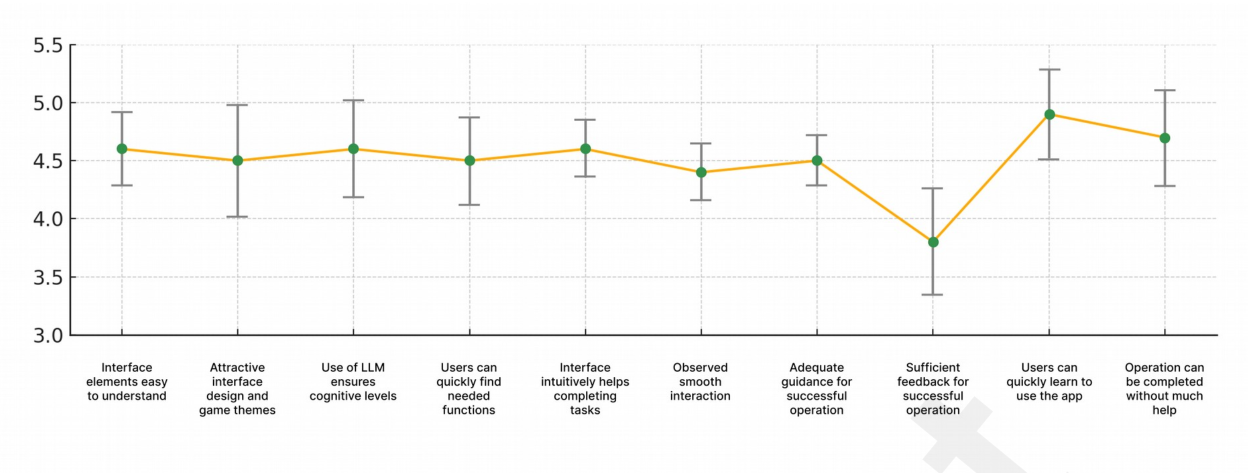
To ensure confidentiality and data protection, all collected speech and interaction data were anonymized at the point of collection. Personally identifiable information was removed and replaced with unique identification codes. Audio and video data were encrypted and stored on secure, password-protected servers accessible only to the research team. In accordance with institutional and national regulations, all raw data were permanently deleted after statistical analysis was completed.

Prototype test results

Usability Performance The evaluation results, summarized in Fig.1, indicate that the Blue Voice app achieved generally positive usability performance. Children reported high scores in areas such as “ease of learning” was 5.0, “task completion without assistance” was mean: 4.6, and “intuitive interface” was 4.6. These findings suggest that the app's core interaction model is well-suited for its target age group. The interface was largely seen as understandable and navigable, and users were able to engage with the app with minimal external help.

Several limitations were identified. Some interface elements were reported to be too complex for younger users, particularly within the training calendar and query modules. The average score for “clear and intuitive interface guidance” was 4.2, and “feedback sufficiency” scored the lowest at 3.8, indicating a need for more explicit visual or auditory feedback during task execution. In addition, the gamification elements, while present, were considered insufficiently dynamic to sustain engagement during longer interactions.

Fig. 2: Average User Evaluation Score



Controlled Experiment on LLM-Enabled Feedback

Methods

Participants We recruited 22 children with hearing loss (female: 10; male: 12; average age: 8.4 ± 1.52) who had previously participated in the prototype usability evaluation. All participants were between six and twelve years old and were randomly assigned into two groups of equal size. The experimental group (n=11) used the *Blue Voice* system with LLaMA-2–enabled real-time feedback, while the control group (n=11) used the same system without integration of the large language model, receiving only simple correctness prompts. Written informed consent was obtained from parents or legal guardians prior to participation, and the study received approval from the Institutional Review Board of Shandong University.

Fig 3. Experimental flowchart

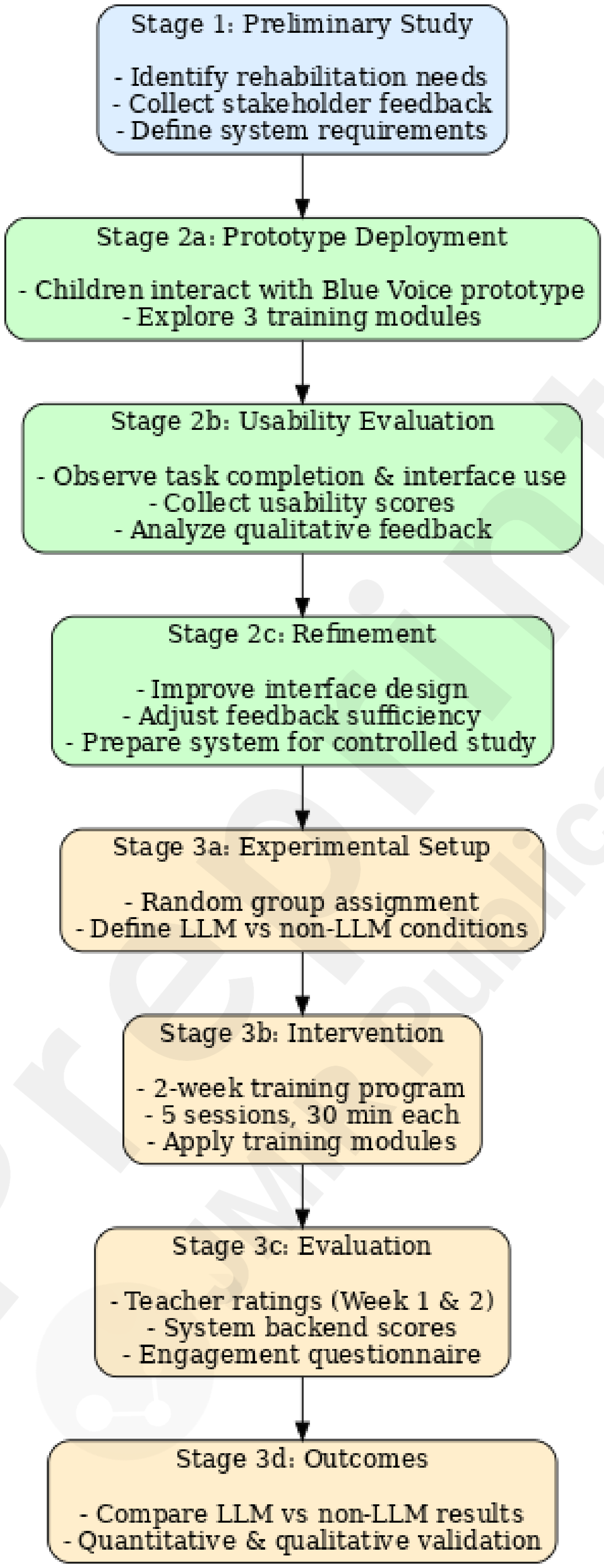


Fig 4. The control group participants used the three training modules of the Blue Voice system in a two-week usability testing process.

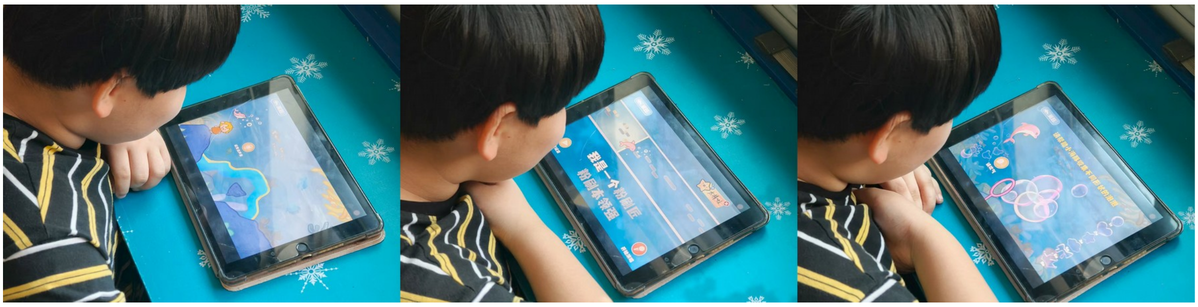


Table 3. Participant characteristics.

Variable	Experimental group (N=11)	Control group (N=11)	Statistic	P value
Age (years), mean (SD)	8.5 (1.6)	8.3 (1.5)	t=0.32	0.75
Sex (female), n (%)	5 (45.5)	5 (45.5)	χ^2=0.00	1.00
Hearing loss severity (moderate–severe), n (%)	7 (63.6)	6 (54.5)	χ^2=0.18	0.67
Previous speech training, n (%)	11 (100)	11 (100)	—	—
Native language (Mandarin), n (%)	11 (100)	11 (100)	—	—

Experimental Flow Each child participated in a structured training program conducted in a home-based setting over a two-week period. During this stage, participants engaged in five training sessions, each lasting approximately 30 minutes. Both groups completed the same three training modules of the *Blue Voice* application, including breath control, pitch-rhythm training, and intonation modulation. In the experimental group, the system processed children's speech through automatic recognition algorithms, followed by semantic-level analysis from LLaMA-2, which generated personalized corrective feedback and guidance. The control group, in contrast, interacted with the identical system interface but received only fixed, non-adaptive feedback indicating whether a response was correct or incorrect. Researchers remotely monitored the sessions to ensure smooth progress without direct intervention. At the end of the intervention, all participants completed a standardized post-test evaluation of speech performance, along with brief usability and engagement questionnaires to capture both objective outcomes and subjective perceptions.

Statistical Analysis

Descriptive statistics (mean, SD) were calculated for all continuous variables. The Kolmogorov-Smirnov test was performed to assess normality. Independent-sample *t* tests were conducted to compare postintervention outcomes between the experimental group and the control group. Paired-sample *t* tests were used to evaluate pre- and postintervention changes within each group. For subjective questionnaire data, thematic qualitative analysis was applied to open-ended responses to extract key themes regarding usability, motivation, and feedback clarity. A significance threshold of $P<.05$ was used for all quantitative analyses. Statistical analyses were conducted using SPSS software (version 26.0; IBM Corp).

Result

Average Value and Variance

Professional teacher evaluation During the first week, teachers rated the speech accuracy of the experimental group at 4.3 ± 0.5, compared with 3.8 ± 0.6 in the control group. By the second week, accuracy further improved to 4.7 ± 0.4 in the experimental group versus 4.1 ± 0.6 in the control group. Over the two-week intervention, teachers consistently reported that the experimental group

showed clearer articulation, better rhythm control, and greater improvement in confidence.
System-based evaluation Automatic analysis from the *Blue Voice* system confirmed these trends. In the first week, the experimental group achieved 4.2 ± 0.4 in system accuracy scores, while the control group scored 3.9 ± 0.5. By the second week, the experimental group reached 4.6 ± 0.3, compared with 4.1 ± 0.5 in the control group. Engagement levels, measured by task completion and interaction frequency, were also higher in the experimental group (4.8 ± 0.3) than in the control group (4.2 ± 0.5).

Table 4. Weekly and overall speech ability cores

Assessment method	Group	Week 1	Week 2	Overall
Teacher evaluation	Experimental (LLM)	4.3 ± 0.5	4.7 ± 0.4	4.5 ± 0.4
	Control (non-LLM)	3.8 ± 0.6	4.1 ± 0.6	3.9 ± 0.6
System evaluation	Experimental (LLM)	4.2 ± 0.4	4.6 ± 0.3	4.4 ± 0.3
	Control (non-LLM)	3.9 ± 0.5	4.1 ± 0.5	4.0 ± 0.5

The Result of The One-Sample T-Test

For the experimental group, teacher ratings were significantly higher than the test value in both week 1 (*t*=3.40, *P*=.007) and week 2 (*t*=4.22, *P*=.002). System-based evaluations yielded similar results, showing significance in week 1 (*t*=3.18, *P*=.010) and week 2 (*t*=3.86, *P*=.003). In the control group, teacher ratings were also above the test value in week 1 (*t*=2.28, *P*=.045) and week 2 (*t*=2.79, *P*=.019). Likewise, system evaluations indicated significant differences in week 1 (*t*=2.31, *P*=.043) and week 2 (*t*=2.60, *P*=.026). These results confirm that both teacher and system evaluations reflected effective improvements beyond the baseline level across the two groups.

Table 5. One-sample t test results

Group	Evaluation method	Week 1	Week 2
Experimental	Teacher	*t*=3.40, *P*=.007*	*t*=4.22, *P*=.002*
Experimental	System	*t*=3.18, *P*=.010	*t*=3.86, *P*=.003*
Control	Teacher	*t*=2.28, *P*=.045	*t*=2.79, *P*=.019
Control	System	*t*=2.31, *P*=.043	*t*=2.60, *P*=.026

Paired-sample *t* tests

Teacher evaluations showed that in week 1 the difference between the experimental and control groups approached significance (*t*=2.05, *P*=.052), while in week 2 the difference was statistically significant (*t*=2.73, *P*=.0046(*P*<.05). System-based evaluations revealed a similar trend, with week 1 results showing a marginal difference (*t*=1.98, *P*=.061) and week 2 results reaching significance (*t*=2.56, *P*=.0032(*P*<.05).

Table 6. Independent-sample t test results

Evaluation method	Week 1	Week 2
Teacher evaluation	*t*=2.05, *P*=.052	*t*=2.73, *P*=.0046*
System evaluation	*t*=1.98, *P*=.061	*t*=2.56, *P*=.0032*

Analysis of questionnaire data of subjective scales

Children in the experimental group frequently described the system as "fun," "encouraging," and "like playing a game." They highlighted that real-time feedback allowed them to "know mistakes immediately" and "try again with confidence," with several reporting a sense of pride after correcting their pronunciation. By contrast, children in the control group characterized the training as "clear but repetitive," noting that binary right-or-wrong feedback reduced their interest over time.

Parents' reports suggested that children in the experimental group showed stronger motivation and more consistent participation than in conventional therapy. Parents observed that interactive and game-like features sustained their children's attention, and some noticed improvements in everyday communication at home, such as clearer pronunciation during reading. In the control group, however, parents noted that although the system was easy to use, the lack of personalized feedback led to occasional distraction and declining motivation.

Across both groups, participants recommended further enhancements, such as incorporating more gamified elements, providing multiple voice options, and enabling adaptive difficulty levels.

Discussion

This study demonstrates the feasibility and effectiveness of the Blue Voice system, which integrates LLaMA-2–enabled real-time feedback to support speech rehabilitation for children with hearing loss. Quantitative results showed that the experimental group significantly outperformed the control group in both teacher and system evaluations. Specifically, teacher ratings revealed significant group differences in week 2 (t=2.73, P<.05), and system evaluations confirmed this pattern (t=2.56, P<.05). One-sample t tests further validated the reliability of the data, with experimental group scores consistently higher than the baseline criterion (t=4.22, P<.01 for teacher evaluations in week 2).

Qualitative findings reinforced these results: children in the experimental group reported greater enjoyment and motivation, parents observed stronger participation and improvements in everyday communication, and teachers highlighted that semantic-level guidance not only corrected pronunciation but also enhanced rhythm, prosody, and expressive ability. By contrast, children in the control group reported that feedback was repetitive, and parents noted a gradual decline in motivation.

These outcomes align with the concept of serious games, which integrate therapeutic or educational goals with game-like elements to sustain engagement and motivation. Serious games have been shown to transform repetitive training tasks into enjoyable activities, thereby improving adherence and long-term learning outcomes. In this study, experimental group participants described the training as fun and game-like, reflecting the motivational advantages widely associated with serious games. By embedding rehabilitation tasks into playful interactions and reinforcing them with semantic-level feedback, Blue Voice can be regarded not only as a rehabilitation system but also as a serious game that merges clinical rigor with engaging gameplay.

System Usability and Child User Interaction

The evaluation results suggest that the Blue Voice system demonstrates good overall usability among child users. Most dimensions received high scores, with "ease of learning" rated at 5.0 and "independent task completion" and "interface comprehension" both at 4.6, indicating that the system is generally easy for children to operate and understand. However, certain areas showed room for improvement. Notably, "feedback sufficiency" scored the lowest at 3.8, and "interface guidance clarity" scored 4.2, suggesting that some users had difficulty recognizing system responses or navigating specific functions.

To address these issues, we propose simplifying the interface design, using larger icons, and setting clearer operational guidelines. Adding interactive tutorials will help children familiarize themselves with the app's functions more quickly 1819. Integrating stronger game elements will enhance the user experience, making it more enjoyable and educationally effective 14.

Feedback Accuracy and Adaptability

The system's performance in providing LLaMA-2–driven feedback demonstrated technical effectiveness, particularly in recognizing pronunciation accuracy and responding in real time. However, inconsistencies were noted. For example, slight tempo deviations in the pitch-rhythm module were sometimes overlooked, while subtle errors in the intonation module occasionally elicited ambiguous responses. These findings indicate that the system requires further calibration of threshold sensitivity and refinement of semantic feedback logic to better match expert-level responsiveness. Moreover, while adaptive content generation based on user history showed promise, it must be optimized to accommodate diverse speech patterns and cognitive profiles among children with varying degrees of hearing impairment.

Limitations

When interpreting the findings, it should be noted that participants were mainly recruited from rehabilitation centers and special education schools, which may not fully represent the broader population of children with hearing loss in community or home settings. Although teacher and system evaluations demonstrated strong consistency, certain discrepancies in automated feedback were observed, indicating that the calibration of threshold sensitivity and semantic feedback logic requires further refinement. Moreover, the scope of this study was limited to usability, engagement, and short-term speech performance, while other important outcomes such as language comprehension, social interaction, and academic development were not assessed.

Conclusions

Blue Voice demonstrates the potential of serious games powered by LLM in pediatric speech rehabilitation. By integrating LLM real-time feedback with interactive and gamified training modules, the system not only improved speech accuracy and prosody but also enhanced motivation, engagement, and overall learning experience. Both quantitative and qualitative results demonstrated that children using Blue Voice with adaptive feedback achieved greater improvements than those receiving only fixed correctness prompts. High usability scores further confirmed that the system is accessible and intuitive for child users, while expert and parental feedback emphasized its capacity to complement traditional therapy. Although challenges remain in refining automated feedback and expanding applicability across diverse user groups, Blue Voice provides a promising foundation for scalable, child-centered rehabilitation tools that bridge clinical expertise with advanced AI technology.

Acknowledgements

The authors would like to express their sincere gratitude to Shandong University for providing institutional support and ethical oversight throughout the study. Special thanks are extended to Jinan Zhiyu Speech Studio for their professional assistance in participant recruitment and data collection. The authors also gratefully acknowledge the support of the Shandong Disabled Persons' Federation, whose collaboration made this research possible.

Conflicts of Interest

None declared.

References

1. C. Yoshinaga-Itano, A. L. Sedey, D. K. Coulter, and A. L. Mehl. 1998. Language of early- and later-identified children with hearing loss. Pediatrics 102, 5 (1998), 1161-1171.

2. S. R. Easterbrooks and B. Stephenson. 2012. Computer-assisted instruction in reading for students who are deaf: Empirical evidence for effectiveness. Education and Treatment of Children 35, 4 (2012), 576-594.
3. H. Li and L. Wang. 2018. Computer-assisted instruction in special education: A review. Journal of Special Education Technology 33, 3 (2018), 158-167.
4. E. J. Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine 25, 1 (2019), 44-56.
5. M. P. Moeller. 2000. Early intervention and language development in children who are deaf and hard of hearing. Pediatrics 106, 3 (2000), e43.
6. J. L. DesJardin. 2006. Family empowerment: Supporting language development in young children who are deaf or hard of hearing. Volta Review 106, 3 (2006), 275.
7. S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers. 2012. A review of wearable sensors and systems with application in rehabilitation. Journal of NeuroEngineering and Rehabilitation 9, 1 (2012), 21.
8. J. C. Read and P. Markopoulos. 2013. Child–computer interaction. International Journal of Child-Computer Interaction 1, 2 (2013), 2-6.
9. J. Hamari, J. Koivisto, and H. Sarsa. 2014. Does gamification work?--a literature review of empirical studies. In Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14). IEEE, 3025-3034. DOI: https://doi.org/10.1109/HICSS.2014.377.
10. C. K. Yiu and C. K. Lee. 2016. Speech recognition technology in computer-aided language learning for children. International Journal of Speech Technology 19, 3 (2016), 623-634.
11. J. P. Hourcade. 2008. Interaction design and children. Foundations and Trends in Human–Computer Interaction 1, 4 (2008), 277-392.
12. S. Nittrouer. 2010. Early development of children with hearing loss. ASHA Leader 15, 10 (2010), 14-17.
13. D. J. Ertmer and J. A. Mellon. 2001. Beginning to talk: Speech development in the first year after cochlear implantation. Ear and Hearing 22, 4 (2001), 371-381.
14. Y. Wang and P. V. Paul. 2011. Integrating computer technology into literacy instruction for students who are deaf or hard of hearing. Journal of Deaf Studies and Deaf Education 16, 3 (2011), 397-415.
15. M. J. Guralnick. 2011. Why early intervention works: A systems perspective. Infants & Young Children 24, 1 (2011), 6-28.
16. Y. Xiong and D. Wang. 2021. Artificial intelligence in speech assessment: A review. Speech Communication 127 (2021), 1-13.
17. F. D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13, 3 (1989), 319-340.
18. H. Knoors and M. Marschark. 2012. Language planning for the 21st century: Revisiting bilingual language policy for deaf children. Journal of Deaf Studies and Deaf Education 17, 3 (2012), 291-305.
19. G. H. Saunders and K. V. Echt. 2007. Use of outcome measures to assess intervention efficacy in hearing rehabilitation. Trends in Amplification 11, 2 (2007), 67-81.
20. H. Song, S. Hao, Y. Chen, et al. 2024. The effect of different visual feedback interfaces of music training games on speech rehabilitation in hearing-impaired children: An fNIRS study. Neuroscience Letters 843 (2024), 138010.
21. Y. Liu. 2024. Proactive detection and early intervention to open up the world of sound for children with hearing impairment—Interview with Huang Lihui, researcher of the Institute of Otorhinolaryngology, Peking Tongren Hospital, Capital Medical University. Women's and Children's Health Guide (07) (2024), 1-3.