

Systematic Review

Interaction, Artificial Intelligence, and Motivation in Children's Speech Learning and Rehabilitation Through Digital Games: A Systematic Literature Review

Chra Abdoulqadir ^{*,†}  and Fernando Loizides [†] 

Department of Computer Science and Informatics, Cardiff University, Cardiff CF24 4AG, UK;
loizidesf@cardiff.ac.uk

* Correspondence: abdoulqadir1@cardiff.ac.uk

† These authors contributed equally to this work.

Abstract

The integration of digital serious games into speech learning (rehabilitation) has demonstrated significant potential in enhancing accessibility and inclusivity for children with speech disabilities. This review of the state of the art examines the role of serious games, Artificial Intelligence (AI), and Natural Language Processing (NLP) in speech rehabilitation, with a particular focus on interaction modalities, engagement autonomy, and motivation. We have reviewed 45 selected studies. Our key findings show how intelligent tutoring systems, adaptive voice-based interfaces, and gamified speech interventions can empower children to engage in self-directed speech learning, reducing dependence on therapists and caregivers. The diversity of interaction modalities, including speech recognition, phoneme-based exercises, and multimodal feedback, demonstrates how AI and Assistive Technology (AT) can personalise learning experiences to accommodate diverse needs. Furthermore, the incorporation of gamification strategies, such as reward systems and adaptive difficulty levels, has been shown to enhance children's motivation and long-term participation in speech rehabilitation. The gaps identified show that despite advancements, challenges remain in achieving universal accessibility, particularly regarding speech recognition accuracy, multilingual support, and accessibility for users with multiple disabilities. This review advocates for interdisciplinary collaboration across educational technology, special education, cognitive science, and human-computer interaction (HCI). Our work contributes to the ongoing discourse on lifelong inclusive education, reinforcing the potential of AI-driven serious games as transformative tools for bridging learning gaps and promoting speech rehabilitation beyond clinical environments.

Keywords: human-centered computing; accessibility theory; accessibility technologies; artificial intelligence in education; digital accessibility; educational technology advancements; concepts and paradigms; HCI theory; concepts and models



Academic Editors: Gabriele Gianini, Silvio Pagliara and Katerina Mavrou

Received: 31 May 2025

Revised: 1 July 2025

Accepted: 9 July 2025

Published: 12 July 2025

Citation: Abdoulqadir, C.; Loizides, F. Interaction, Artificial Intelligence, and Motivation in Children's Speech Learning and Rehabilitation Through Digital Games: A Systematic Literature Review. *Information* **2025**, *16*, 599. <https://doi.org/10.3390/info16070599>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As of 2022, 1.2 million children aged 0 to 12 were diagnosed with a speech disorder [1]. Research has shown that “[e]arly intervention for people who suffer from speech disorders would prevent many problems in the future” [2]. Therefore, this academic paper seeks to conduct a systematic analysis of digital games designed for the rehabilitation of speech disabilities in children. By addressing key questions, this review aims to show the current

research, identify gaps in existing knowledge, and provide insights for future research and development.

A serious game can be defined as “a game designed for a primary purpose beyond that of pure entertainment. The influence of games on the cognitive, emotional and social domains of players increases motivation and engagement of learners” [3]. In the context of this paper, we refer to “speech rehabilitation games” and “speech learning games” interchangeably. These are serious games designed for speech rehabilitation for people with speech disabilities.

Exploring digital games for children’s speech therapy is an area of active research. For example, there is a recent systematic review of children’s speech therapy games with 27 included papers [2]. The study shows that while the games show positive effects on children’s motivation, engagement, and satisfaction, some issues were identified [2]. These included frustration from failure, low self-esteem, environmental noise interference, mismatch between game difficulty and user needs, and limitations in speech recognition technologies. The study does not explore speech recognition technologies or other Artificial Intelligence potential. Another recent literature review examines the Automatic Speech Recognition (ASR) systems in 37 studies [4]. The review explores the accuracy and usability of ASR systems, both for commercial and non-commercial users with dysarthria. However, the study does not explore gamification solutions and how ASR can be integrated within serious games.

Understanding the level of independence exhibited by children while engaging with digital speech rehabilitation games is crucial for evaluating the practicality and accessibility of these tools. Examining the degree of autonomy can shed light on the usability and effectiveness of such games within diverse settings, including homes, schools, and clinical environments. Therefore, our first research question focuses on the levels of independence, whether the user can play the game independently or with the intervention of speech therapists.

The selection of interaction modalities plays a pivotal role in shaping the effectiveness of digital games for speech rehabilitation. This question explores the diverse ways in which children interact with these games and aims to identify patterns associated with improved speech outcomes. An in-depth analysis will contribute to the optimisation of game design to enhance therapeutic impact. As a result, our second research question will explore the user input methods that researchers are attempting to implement in digital speech rehabilitation games.

Our third research question explores the role of Natural Language Processing (NLP) in such speech rehabilitation games. It is crucial to examine how these new AI technologies are integrated into digital games for speech rehabilitation. This inquiry seeks to elucidate the specific contributions of AI, addressing its potential to adapt to individual needs, provide real-time feedback, and enhance the overall efficacy of rehabilitation interventions.

The final research question explores the effects of digital games beyond clinical efficacy; the motivation and engagement of children in speech rehabilitation are essential factors influencing the overall success of interventions. The final research question delves into the emotional and psychological dimensions of gaming, exploring how these platforms can be tailored to sustain children’s interest, boost motivation, and foster a positive attitude towards speech learning.

Exploring these critical questions, this paper aims to provide a holistic understanding of the current state of digital games in speech rehabilitation for children. The findings aid in informing practitioners, researchers, and developers but also contribute to the ongoing evolution of innovative and effective interventions in pediatric speech therapy, including our work in digital games for rehabilitation [5].

2. Research Methods

We adopted the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) [6] for identifying and presenting the literature. To achieve the aim of this review, we employed the use of the PESICO (Person (and problem), Environments, Stakeholders, Intervention, Comparison, Outcome) [7] framework. The PESICO framework allows defining stakeholders beyond the main users of the games, such as therapists. Its structure supports a Human–Computer Interaction (HCI)-oriented review where accessibility, independence, and engagement are key outcomes.

2.1. Person (and Problem)

We are focusing on children because early intervention can help prevent issues from developing later in life [2]. Upon reviewing published research in this area, the age researchers target is around two to twelve years old [2]. Therefore, this review paper aims to include research targeting children who are 2–12 years old. Some of the resources might not specify their target age range. We will include them if they target children with speech disabilities who can speak but have difficulty learning to speak and have the potential to improve their speech. Children falling into the category of speech or language impairment as defined by The Individuals with Disabilities Education Act (IDEA) [8].

2.2. Environments

We have considered digital games used in therapy clinics, at home, or in classroom settings. We have also considered PC and smaller screens, such as tablets or mobile phones. Voice input devices could be the device's internal microphone or an external device.

2.3. Stakeholders

- Children with speech disabilities: These are one of the target users in the solutions and published research papers.
- Caregivers or parents: Most projects in speech rehabilitation involve supervision by parents and caregivers. Thus, they are major stakeholders.
- Therapists: They are end users in many speech rehabilitation games for multiple purposes, such as setting up the game or reviewing feedback.
- Developers and Computer Scientists: This paper aims to conclude the common recommendations suggested by the research conducted in the digital speech rehabilitation game area.
- Researchers (Sociologists, Psychologists, Medical Professionals): This paper addresses some of the areas speech rehabilitation games target, as well as the areas for future research recommended by the researchers and this review of the state of the art. These could benefit researchers in related areas.

2.4. Intervention

We are looking at how digital games can be used in speech therapy with the help of AI. These games provide speech rehabilitation exercises. AI plays a large role in user interaction, affecting the speech rehabilitation experience. The intervention items reflect different elements used in digital games targeting children's speech therapy, which include the exercises implemented in speech rehabilitation games with the help of AI and NLP to enhance the overall experience.

2.5. Comparison

- Degree of children playing rehabilitation games independently.
- Speech recognition libraries in digital games.
- interaction and feedback implemented in different games.

2.6. Outcome

- Engagement and motivation.
- Greater independence and less intervention from carers or therapists.
- Enhanced pronunciation.

2.7. Research Questions

The following research questions were then defined:

1. **RQ1:** What is the degree of independence of children playing current speech rehabilitation or learning games?
2. **RQ2:** What interaction has been found to be effective in rehabilitating children with speech disabilities?
3. **RQ3:** What is the role of Artificial Intelligence, such as Natural Language Processing, within digital games for children's speech rehabilitation?
4. **RQ4:** What is the impact of using games for rehabilitation exercises on the motivation and engagement of children with speech disabilities?

2.8. Related Databases

This review focuses on speech rehabilitation through digital games, particularly from a technological and interaction design perspective. ACM Digital Library and IEEE Xplore are domain-specific databases rich in HCI, accessibility, speech recognition, AI, and serious games literature.

1. ACM Digital Library: Provides research articles related to technological and computer science aspects, such as software and hardware.
2. IEEE Xplore: Provides the technological aspects of the search.

2.9. Search Terms

An exhaustive search was conducted on 2 May 2025 on the databases mentioned in Section 2.8. The syntax for these databases slightly differs. Therefore, we had to adjust the syntax to be compatible with the respective search engine and be specific. The target papers on the different databases remain the same. Below are the search terms used in the databases.

1. ACM Digital Library:
(Title:(speak* OR speech OR voice) AND (Title: (rehab* OR therapy OR serious AND NOT autis* AND NOT dyslex*)) AND Title:(gam*) AND Title:(child*)) OR (Keyword:((speech OR speak* OR voice) AND (rehab OR therapy) AND (gam*)))
2. IEEE Xplore:
(("Document Title":speech OR speak* OR voice) AND ("Document Title":rehab* OR therapy OR serious) AND ("Document Title":gam*) AND ("Document Title":child*)) OR (("Author Keywords":speech OR speak* OR voice) AND ("Author Keywords":rehab* OR therapy) AND ("Author Keywords":gam*)) NOT ("Document Title":autis* OR "Document Title":dyslex*)

2.10. Scope

We are looking at work that has been produced both as a training aid to carers/nurses and given to children as prototypes to try. This paper is from the Human-Computer Interaction and user experience perspectives. It does not provide medical perspectives or efficiency. We are looking at research used for speech rehabilitation, not used to teach educational content or meet curriculum objectives. Thus, the papers we were able to find did not include traditional teachers in schools. The papers testing their solutions in classrooms have objectives from the rehabilitation and therapy perspective, not pedagogical.

2.11. Selection Process

A backward and forward reference search was also performed: the backward search identified the references from certain research, while the forward search identified the research that cited a certain study. This revealed a total of 609 potentially relevant articles (see Figure 1). The titles and abstracts were exported to a spreadsheet for them to be reviewed. To identify studies, we used the following criteria:

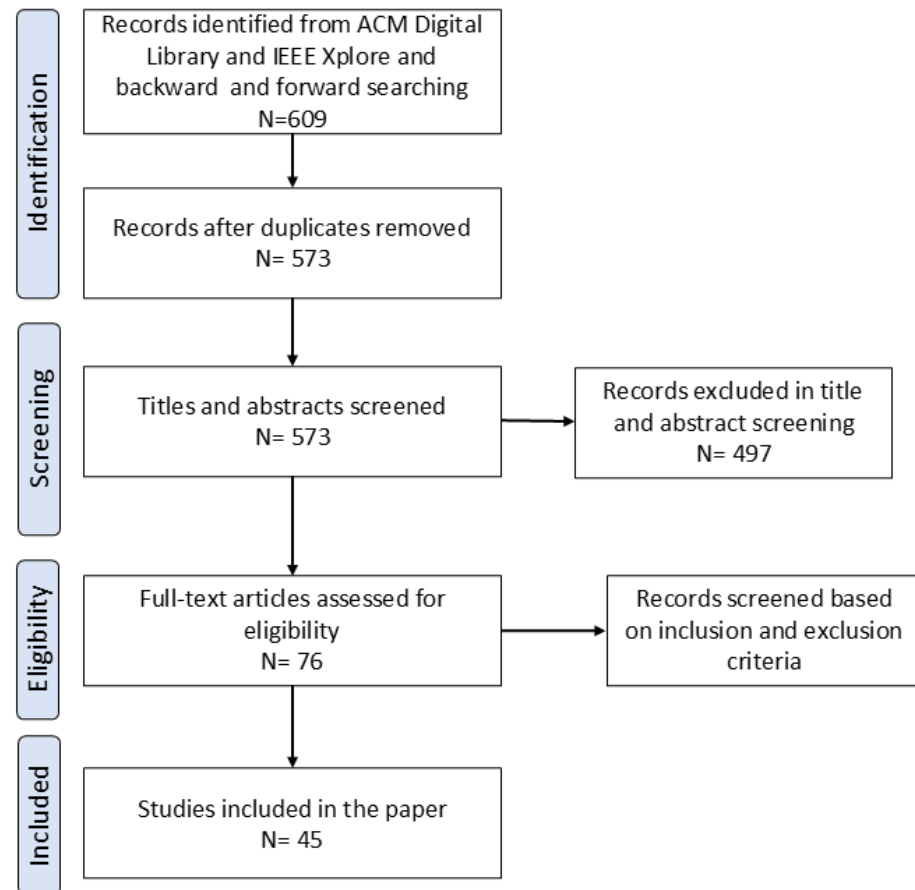


Figure 1. Study selection process flowchart. (PRISMA Checklist has been shown in Supplementary File).

2.11.1. Inclusion Criteria

- Studies published between 2011 and 2023 inclusive (commercially available ubiquitous interaction with speech recognition influences people and technology enhancement, with Siri being launched in 2011).
- Studies need to be identified as targeting children up to secondary school. It ensures the majority target audience is within the age range we include (2–12).
- Studies shall be identified that contain children with no, or almost no, speech ability in their current state, but with the potential to improve their speech.
- Studies shall target rehabilitation or serious games in order to enhance the users' speech.
- We only consider papers written in the English language
- Studies shall be published rather than in-press.
- Studies considered in the backward and forward searching can be published anywhere as long as they are relevant and specific to the topic.

2.11.2. Exclusion Criteria

The results included related articles of different types of rehabilitation, reports, narratives, and papers in other languages. Below are our exclusion criteria applied to the results

and backwards and forwards searches. These criteria were applied in the title and abstract screening and in the full-text eligibility stages shown in Figure 1.

- Dyslexia: There is a fine line between speech disabilities and dyslexia, and we want to ensure that the papers target children with speech disabilities rather than reading, writing, and spelling.
- Case reports, narrative reviews, and opinion pieces shall be excluded.
- Studies shall not be in the peer-reviewing stage.
- Studies focusing on languages other than English shall be excluded.

Selecting studies that use the same metrics will allow the authors to compare one thing that is common across studies. Our focus is on Human–Computer Interactions; therefore, we excluded clinical and therapeutic research that does not include the intervention of digital games. Figure 1 presents the comprehensive numerical details. The manuscripts were exported to a spreadsheet and assessed manually. AI automation tools were not used. Finally, we tried finding answers and recording them in a spreadsheet for each of the research questions. Table A4 shows the different articles used to answer each research question.

3. Results

3.1. Conducting the Review

The second author determined the scope of the paper, while the first author focused on the databases and search terms. The research scope is concluded based on previous research conducted in this area [4,5]. The second author supervised and carried out the final review of the screening results. Both authors reviewed the papers independently, compared them, and discussed conclusions accordingly. The papers were examined based on the research questions and were compared against the inclusion and exclusion criteria. The relevant pieces of information were categorised into separate tables to oversee the body of knowledge and draw conclusions. We used visual inspection to go through each paper based on the inclusion and exclusion criteria. We began with the exclusion criteria and filtered down the papers. We have not used any software or Large Language Models (LLMs).

3.2. Summary of Results

A total of 45 papers were included and analysed in this study. The papers were diverse and from different countries. Figure 2 shows the different countries the selected papers were written. The papers also provided speech rehabilitation and exercises in different languages other than English, such as Turkish, Portuguese, and Greek. These different languages affected their speech rehabilitation focus. For example, ref. [9] focuses on isolated sibilant exercises in the Portuguese language because they are commonly used by speech–language pathologists. Moreover, Figure 3 shows the number of included studies per year, with the highest number of relevant papers published in 2021.

Moreover, we have identified and summarised the papers according to the types of tests conducted on their proposed solutions. Some of the articles provide theoretical models and frameworks, in which case the researchers have not tested them on end-users, while some papers propose solutions, but they intend to have tests in their future work. Other articles mention testing on different stakeholders of their prototypes. Figure 4 shows the percentage of each category. The number of participants in the articles that included tests ranged from one child [10], tested on a 6-year-old, up to 90 participants [9], who collected data from three schools in the Lisbon area.

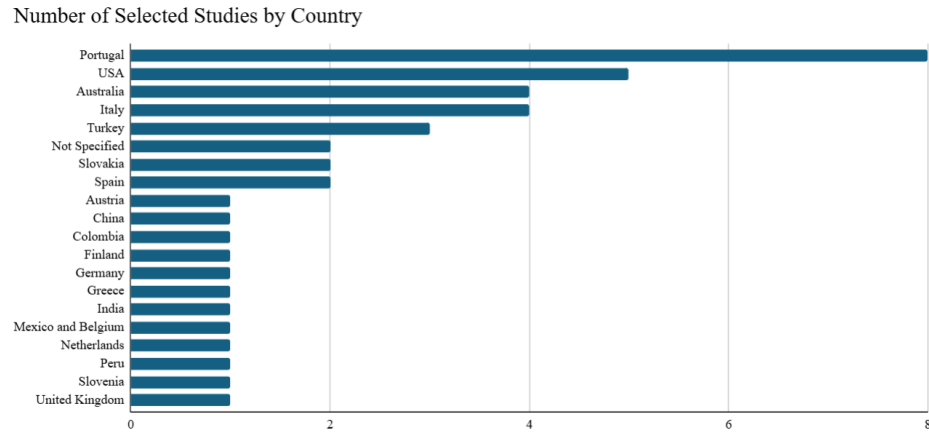


Figure 2. Number of selected papers by country.

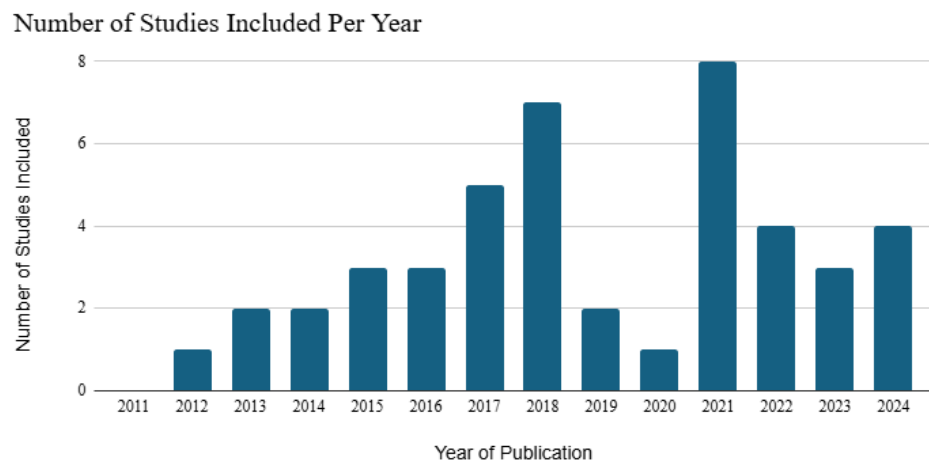


Figure 3. Number of studies included per year.

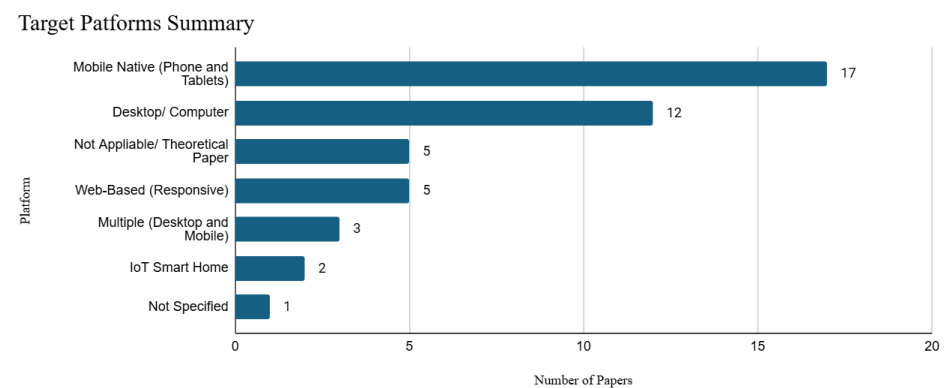


Figure 4. Percentage of test categories in the selected studies.

The age ranges targeted in the different studies vary. Most of the papers specified their target age range group eligible to use their solutions. Some of the articles did not specify their age range nor the age of the participants if they had tested their solutions. Some did not mention their target age range but specified the participant ages. Determining the age range for some others was not applicable as the papers were theoretical, such as focusing on the accuracy of AI models used in speech therapy. The minimum age targeted by the articles was 2 years old for both recruited participants and target groups. Similarly, the maximum age for both groups was 12 years old. Figure 5 shows the percentages of age ranges specified in the papers.

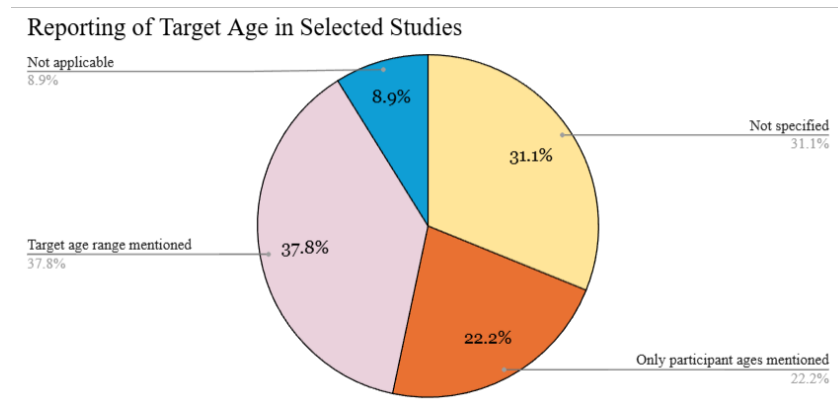


Figure 5. Percentage of age ranges and participant ages specified.

The summary of the platforms the papers targeted shows that mobile platforms are the most commonly chosen for speech rehabilitation games. However, speech rehabilitation games span over a variety of platforms, including smart home, desktop, and web. A summary of the platforms mentioned in the papers is shown in Figure 6.

A major result we obtained was the fact that 34 papers did not target any other disabilities other than speech. There were only five papers involving multiple disabilities. The most common disability combination was hearing loss, with three papers mentioning it [9–11].

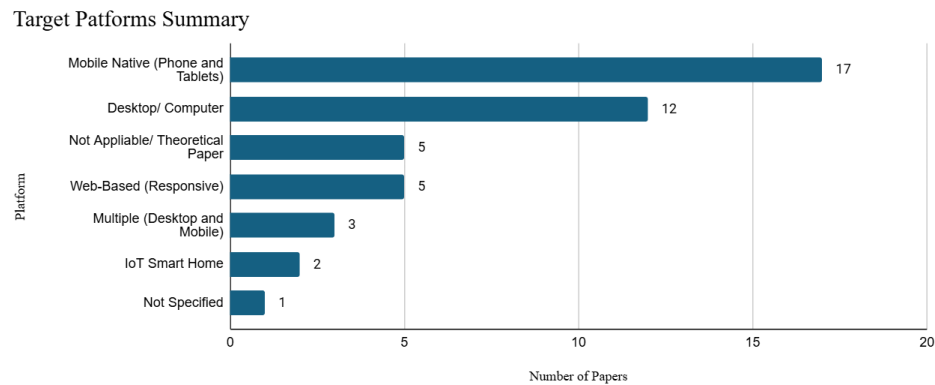


Figure 6. Number of papers per platform mentioned.

3.3. RQ1: What Is the Degree of Independence of Children Playing Current Speech Rehabilitation or Learning Games?

Speech rehabilitation games offer a spectrum of approaches. We have reviewed the games based on the number of end-user categories needed to prepare the game, enabling the child to play it, and whether supervision is necessary during the gameplay. Some of the children’s rehabilitation games are supervised by therapists or instructors, while some allow autonomous play after the instructor prepares the necessary exercises. Other games function independently without direct supervision.

Some games are supervised to provide more personalised feedback and monitor progress. These games target therapists or adults to create new exercises or reinforce pronunciation. For example, in the Fanima game, children do not progress autonomously [12]. Progression depends on the therapist’s real-time classification of the child’s speech via a web platform. The game only moves forward when the therapist validates the spoken input. Some of these games allow the patient to play autonomously once the exercises are prepared [13–16]. Others require supervision and reinforcement [10,17,18]. Researchers

have given the option to download the game from Google Play, but they suggest the involvement of a therapist or supervisor to prepare and personalise the exercises [15]. The feedback of these games targets therapists to oversee the progress and implement further training [10,13,15,18,19].

Researchers have developed a model they call SEGA-ARM for auditory rehabilitation serious games [10]. In their case study, the therapist has an essential role in choosing the series of phonemes and exercises necessary for the patient. Their model tracks the advancement of the player to give feedback to the therapist. This model supports the fact that therapists shall be involved in serious games for speech rehabilitation as a main end-user group, starting the application cycle.

Researchers also explore the potential of smart home technologies in assisting therapists through voice assistants [13,20]. Some research investigation revolves around the execution of therapeutic exercises outside conventional clinical settings by harnessing the capabilities of existing smart home technologies [13]. The process initiates with therapists preparing and assessing the exercises for effectiveness. Parents or caregivers facilitate the setup of the smart home environment, enabling the child to participate in these prescribed exercises. The child's engagement and performance are recorded for subsequent evaluation by the therapist, forming a comprehensive cycle that integrates technology and therapeutic practice, expanding the potential for remote and technologically supported rehabilitation approaches.

Rubin and Kurniawan worked on a mobile game with a similar workflow, starting with the therapist recording the child's speech to upload it and use it for the speech recognition system, "[t]he therapist will record a few sentences of the child speaking with both proper speech and cleft speech. The system will upload the files to the server, which will perform the adaptation on the base model and send the resulting model back to the device" [21].

Speech therapy is also considered within classroom settings [22–24]. Nanavati, Bernardine, and Steinfeld have worked on deploying speech therapy games in India [24]. Their results show that the students could play the games autonomously without the help of the teachers but with the help of the "alpha" students. The "alpha students" are described as those who understand the game more quickly than others. They have noticed that these students tend to help others, depending on how much control the teachers maintained over computer usage [24]. Even though it increased the flexibility of the teachers and the efficiency of the teaching process, the system was designed for classroom environments with the teachers as one of their main end-user groups.

The game "Into the Forest" operates autonomously, designed to facilitate a child's progress without the continual guidance of a therapist or teacher [11]. Through the guidance of an arrow, the game allows independent navigation, reducing the reliance on external assistance. It focuses on teaching and reinforcing vocabulary expected of a child under the age of 7. The primary goal is to encourage prolonged engagement without the necessity of supervision or a clinical environment. Similarly, ref. [9] has created a game for extensive training that can be used at home without the supervision of therapists or parents. Researchers present Logopedic's Escape, a high-autonomy voice-controlled game where children interact via spoken commands [25]. Children are able to control the gameplay via voice input without therapist intervention during play [25].

In the exploration of speech rehabilitation games, it is evident that various approaches exist, each tailored to address the unique needs and circumstances of the users. Research targets different end-user independence levels depending on the specific condition requirements. Figure 7 shows that most of the papers provide solutions with therapists or parents as their main stakeholders. While supervised games offer effective feedback and targeted progress, the evolution of technology has presented opportunities for games that

empower children to engage independently. The gamification of smart home technologies and therapeutic practices shows the potential for remote, technology-driven rehabilitation and learning approaches, expanding the horizons of therapeutic interventions beyond traditional classroom settings. Table 1 shows the references and their relevant levels of child independence.

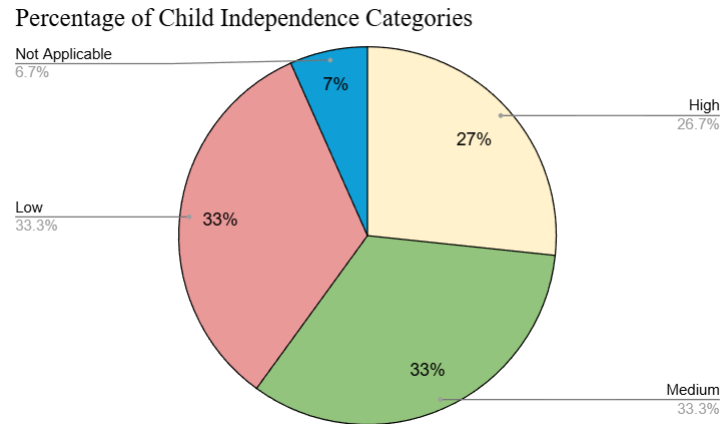


Figure 7. Percentage of papers falling into different child independence categories.

Table 1. Classification of selected papers based on the level of child independence in their proposed solutions.

References	Level of Independence	Notes
[9,11,14,23,26–33]	High	Child plays independently without therapist supervision (therapist supervision is optional if available).
[14–16,21,28,34–43]	Medium	Therapist prepares the game, but the child plays autonomously.
[10,12,13,17,18,20,22–24,34,44–48]	Low	Requires supervision and reinforcement by therapists.

While the level of independence varies across speech rehabilitation games, this independence is closely linked to how children interact with the games. The design of input methods, such as voice commands, phoneme recognition, and feedback mechanisms, plays a critical role in enabling or limiting autonomous engagement. Therefore, the next section explores the types of interaction modalities used in these games and evaluates which have been most effective in supporting speech rehabilitation outcomes for children.

3.4. RQ2: What Interaction Has Been Found to Be Effective in Rehabilitating Children with Speech Disabilities?

Three main designs were identified in our review process of game development for speech rehabilitation. Most researchers developed their games in 2D format meeting the requirements for speech rehabilitation [5,9,12,16,19,21,25,28]. The main user interaction for these games is speech recognition with different words or phrases depending on the target speech disability. On the other hand, some researchers used 3D environments leveraging over-the-shoulder perspectives [11,18,31]. They utilised similar speech recognition interactions to navigate through the game’s map. They concluded that the 3D environment is motivational, leading to a positive user experience. Moreover, research has been conducted around Augmented Reality (AR) to enhance motivation and user interaction [47]. The diverse design choices in speech rehabilitation games, whether leveraging 2D, 3D, or Augmented Reality, demonstrate formats to enhance user interaction and engagement,

emphasising the critical role of design in shaping the efficacy and user experience within these therapeutic tools.

Visual feedback was a key interaction method across several papers. *Fanima* used image prompts and sound replay to support recognition [12], while *Logopedic's Escape* included animated avatars and mouth gestures to guide pronunciation [25]. In the Slovak web game, confidence scores were visualised through token movement on a board [43]. These implementations suggest that visual cues enhance focus, reinforce correct pronunciation, and help reduce errors, especially in children with limited reading or attention skills.

Our next consideration for user interaction involves examining the voice control modes utilised in the therapy process. Research has been conducted around continuous speech and voice availability to control the game [19]. For example, in “*Flappy Voice*”, which is an inspiration from *Flappy Bird*, the player needs to speak so that the bird does not fall. It is controlled by vocal loudness continuously until the level is complete. In another paper, the interaction is via whistle sounds (blowing/sucking), which trigger in-game events [43]. The games use pitch detection and volume analysis. In addition, some of the games are designed around specific word pronunciation in specific parts of the game [11,21]. Ref. [11] has a set of predefined words relative to the age of the child. This is similar to [21], whose game asks the user to alter certain words expected for a child to know at a certain age. Other games use phoneme recognition [22,27,37,38,47]. These are mainly the vowel sounds aimed at phonological performance. Finally, some games' focus is on sibilant consonants, developed to practice consonants [9,26].

The final aspect we consider in our user interaction review is how feedback and evaluation are communicated to users in these games. Two user experience approaches have been the most common in speech rehabilitation games. Real-time feedback is recommended by eight papers. Customisation and adaptation of the difficulty levels or preferences were also mutual in six articles. This provides insight and recommendations to game developers and researchers in this field, as it affects the user experience and game efficiency. Table 2 shows the different references recommending these elements in game design. Moreover, considering the socio-cultural context of the target audience while developing games is highlighted by Nanavati, Bernardine, and Steinfeld [24]. Their games are designed to help children who are deaf or hard of hearing (DHOH) explore and understand their voices in Bengaluru, India. They highlight the fact that social and cultural stigmas and preferred languages shall be considered when designing speech rehabilitation games.

Table 2. Most common design recommendations for speech rehabilitation games.

Design Recommendation	Supporting Studies
Real-time feedback	[12,16,17,25–28,31,32,36,37,41,42,45,47]
Customisation and adaptation	[12,16,20,21,25,26,35–37,42,45]

The majority of the papers provide an overall score after the child's voice input [5,11,27,31,42]. This score is calculated by comparing the child's voice input to the expected input. Duval et al. provide feedback using animated facial expressions [28]. They want to create “natural and realistic experiences within the game” showing “primary emotions”. Some researchers use the main character's behaviour to provide feedback. For instance, if the child mispronounces the proposed word, either the character stops and waits for another attempt or the exercise has to be repeated from the beginning [9]. In this instance, researchers believe that this form of feedback is “highly intuitive”.

The exercise methods are diverse and not limited to a single approach. Researchers have explored various techniques, including exercises for increasing loudness, sustaining vowel sounds, improving consonant pronunciation, as well as working on pitch, volume

control, and emotional expression and engagement. Among these, vowel and pitch exercises have received the most attention. Table 3 provides an overview of the studies that focus on different rehabilitation techniques, with some studies incorporating multiple methods to enhance overall effectiveness. It is important to note that these decisions are affected by the target audience and the language preference. For example, [9] have created a serious game for training sibilant consonants because the distortion of sibilant sounds is common in Portuguese-speaking children.

Table 3. Most common types of rehabilitation exercises in the chosen papers.

Type of Rehabilitation Exercise	Supporting Studies
Loudness	[17,38]
Vowel Exercises	[9,25,27,32,37,38,47]
Consonant Words	[9,12,16,25,26,41,42,47]
Pitch	[17,24,38,45]
Volume	[24,28,45]
Emotional Expression and Engagement	[20,23,43]
Customisation and adaptation	[12,16,20,21,35–37,41,45]

Moreover, Table A2 presents details of the AI techniques, interaction methods, and intended age groups across the reviewed studies. A range of AI-driven approaches was identified, including template matching, dynamic scoring, and voice assistant integration. Interaction modalities varied from simple touch and voice commands to multimodal feedback and smart home automation. While most studies targeted children between the ages of 2 and 12, some did not specify their intended age range. This summary enhances comparability and highlights design trends in AI-supported speech rehabilitation games.

The effectiveness of speech rehabilitation games relies on diverse design elements and feedback mechanisms. From the prevalence of 2D environments to the emergence of 3D and Augmented Reality formats, design choices significantly impact user engagement. Varied voice control modes, phoneme recognition, and consonant practice further enhance interaction. Feedback methods, such as overall scores, facial expressions, or character behaviour, are important in user engagement.

The interaction methods employed in speech learning games, ranging from speech recognition and pitch control to visual and real-time feedback, demonstrate the importance of design. However, the success of these interactions often relies on the technologies used. For example, the role of Artificial Intelligence, and more specifically Natural Language Processing (NLP), is central to enabling accurate speech recognition, adaptive feedback, and personalised learning. The following section investigates how AI technologies have been integrated into these games and assesses their impact on speech rehabilitation for children.

3.5. RQ3: What Is the Role of Artificial Intelligence (AI), Such as Natural Language Processing (NLP), Within Digital Games for Children’s Speech Rehabilitation?

When employing AI and machine learning for speech rehabilitation, the choice of a speech recognition algorithm is one of the initial decisions in the implementation stage. The reviewed studies reveal a diverse trend in the application of ASR technologies for children’s speech rehabilitation, shown in Table A1. Some studies emphasise data quality and collection, such as therapist-validated or child-specific corpora, which are necessary for training ASR models that can accommodate the variability in children’s speech. Others present fully implemented voice-controlled games, relying on real-time pitch, phonation, or sound intensity rather than full speech recognition to engage users. While systems

like “Apraxia World” [36] and “Into the Forest” [11] integrate speaker-dependent models and template matching to improve detection accuracy, several newer studies employ lightweight audio features, such as pitch and volume, instead of conventional ASR. They favour responsiveness and simplicity over linguistic precision. These approaches often enable greater customisation and therapist control, making them accessible in both clinical and home settings. A few systems report formal accuracy benchmarks, but many highlight successful engagement and motivational outcomes, particularly through visual and audio feedback.

Moreover, latency or response time is highly dependent on the algorithm. Choosing the incorrect algorithm creates end-user frustrations, as evident in research [34]. For example, research shows that Whisper-Local struggles with phoneme detection, but it performs adequately at the word level [25]. To maximise accuracy, some of the researchers use speech recognition based on Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) [16]. They claim that the model recognises correct pronunciation of syllables and specific sounds, necessary for the games. The careful selection of a speech recognition algorithm is critical in ensuring effective functionality and user satisfaction, underscoring the impact of this decision on the overall success of speech rehabilitation AI systems. Meanwhile, background noise cancellation is another key element.

Twenty-eight papers use NLP and speech recognition in their solutions. Speech recognition has its challenges affecting its efficacy. As a result, researchers try to minimise them by exploring different approaches. Background noise and the quality of the voice highly affect the efficiency and accuracy of the NLP algorithm and speech recognition system. Therefore, the microphone and the overall hardware used are considerably important in a developed game. Interestingly, research suggests that Kinect microphone is better than most built-in or conventional microphones for voice recognition [40]. One of the primary reasons is its capability to automatically suppress noise, allowing the game to be used in natural speech environments [40]. With this, one can minimise the processing necessary to remove background noise. Additionally, the hardware storage and processing capacity affect the accuracy of the speech recognition system. For example, mobile devices became one of the significant barriers for processing and response time in the “Speech Adventure” game [21]. Challenges around hardware are mentioned again when researchers had to re-record their audio in a soundproof studio to avoid background noise being recorded and used in their system [29]. This is mainly because tablet computers’ audio quality was limited compared to the requirements of the machine learning system. In short, hardware quality significantly impacts speech recognition accuracy and user experience in speech rehabilitation gaming environments. The ultimate goal is to minimise the hardware necessary to run the applications while maximising efficiency.

Another step in utilising AI and NLP in speech rehabilitation research is adjusting different voices, microphone status, and noise levels. This includes normalisation [19] or calibration [32]. It is necessary because environmental noise is often a challenge in speech recognition systems [2]. Implementing AI techniques to help therapists in the Automatic Speech Recognition (ASR) tuning to the “patient’s speech according to disease severity” is helpful to avoid frustration and inaccurate feedback [45]. Therefore, the dataset used in the AI model directly affects the efficiency and accuracy of the voice recognition and results in the game. Barletta, Cassano, Pagano, and Piccinno have emphasised this and proposed a model where people participate in the “basic creation of phoneme records” [44]. They propose that the same method could be used by the therapist so that the AI is trained with realistic data and recognises phonemes correctly. Thus, fine-tuning voice, microphone settings, and noise levels, coupled with AI-driven ASR tuning to match disease severity,

significantly impacts the dataset quality, influencing the accuracy and efficiency of voice recognition in speech rehabilitation games.

While the majority of the papers use ASR and automatic scoring, some have used human evaluators to avoid ASR errors. Hair, Monroe, Ahmed, Ballard, and GutierrezOsuna claim that using human evaluators in their system helps them avoid “frustration from ASR errors” [36]. Likewise, issues related to speech recognition accuracy have occurred from acoustic mismatches between training and testing data [48]. Indeed, speech recognition accuracy is a challenge [39]. However, research to enhance the accuracy of the models is active. Diogo, Eskenazi, J. Magalhães, and Cavaco targeted accuracy issues [27]. They implemented a “robust scoring model” to provide real-time feedback. Some used Google voice recognition API and received a score of 73.2% for accuracy, which they consider high [30]. Similarly, ref. [42] used ASR to include the Slovak language. All in all, addressing challenges in speech recognition accuracy, from ASR–human discrepancies to acoustic mismatches or included languages, remains an active area of research, striving to enhance models for robust, real-time feedback.

Whisper AI is mentioned in only one of the studies [25]. Most studies focus on lightweight and offline models, such as Template Matching (TM), Goodness of Pronunciation (GOP)-based models, or embedded ASR. These are computationally lightweight, more predictable, and easier to integrate into constrained platforms, such as tablets or mobile applications. Even though Whisper-Local shows promise in the solution, it struggles with phoneme-level recognition in speech learning games [25]. This study was conducted in February 2025, showing that Whisper AI is not yet considered efficient for speech learning games.

While AI technologies, such as speech recognition and NLP, enhance the technical accuracy and adaptability of speech learning games, their effectiveness also depends on how engaging and motivating they are for children. Positive user experience, sustained attention, and emotional investment are crucial for long-term therapeutic impact. The next section explores how game design strategies support motivation and engagement and how these emotional factors contribute to the success of speech rehabilitation exercises.

3.6. RQ4: What Is the Impact of Using Games for Rehabilitation Exercises on the Motivation and Engagement of Children with Speech Disabilities?

Research has shown that games used in speech rehabilitation keep the children motivated and more engaged throughout the session [22,23,25,32]. This is accomplished using various methods. For example, controlling the game’s avatar through speech motivates the player to use speech to move around and discover the environment [11]. Other projects provide autonomous modes in their games, resulting in higher degrees of engagement [23]. Speech recognition techniques also enhance motivation in conducting repetitive speech exercises [32]. Additionally, level progression helps players “maintain their motivation through continuous feedback on their performance” [16,33]. For instance, one of the games uses an ice-cream progress bar to keep track of the child’s performance and keep them motivated to finish successfully [47]. The use of selective, personalised digital games in speech rehabilitation enhances and affects children’s motivation considerably [22]. Gameplay elements, such as levels and avatars, are also mentioned in [25] to keep the child motivated.

Concepts of reward and punishment are necessary to keep the child competitive and eager to finish [10,16,34,35,47]. The SEGA-ARM model suggests including forms of reward and punishment systems, such as “points, badges, levels, and leaderboards”, and provide ways to earn and lose points when designing speech rehabilitation games [10]. Another article emphasises motivation through goal setting and unlockable achievements [43]. It mentions that plans can be personalised and children can see progress through rankings and rewards. Research results show that children and speech–language pathologists (SLPs)

prefer “games with rewards, challenges, and multiple difficulty levels” [34]. Other forms of rewards utilise time intervals. Apraxia World’s reward and punishment system is through giving different time intervals to utter the word, 10 s for correct and 5 s for incorrect pronunciations [35]. This gave them positive results on the child’s motivation throughout the game. Overall, digital games specifically aimed at rehabilitation have been observed to produce more acceptance rates and positive effects on children [49]. Gamification sustains children’s motivation through diverse engagement strategies effectively. While limited assessments directly link games to speech performance, positive outcomes reinforce the potential of digital games in enhancing children’s speech rehabilitation.

Figure 8 shows the frequency of different gamification elements identified across the reviewed studies. Feedback and difficulty scaling are used the most. This shows the importance of real-time responses and adaptive challenges in maintaining engagement. Rewards and progression and narrative elements were also common to motivate continued play. Fewer studies incorporated avatars, session design, or control-based mechanics, suggesting these areas are less explored. This distribution reflects current priorities in speech rehabilitation game design. There are opportunities to expand on strategies less commonly used, such as autonomy and personalisation for children with speech impairments. As discussed in research question 1, most games involve therapists or caregivers for configuration or supervision. A detailed table, Table A3, is presented in the appendix section to show the details of the gamification elements and their results for engagement over time.

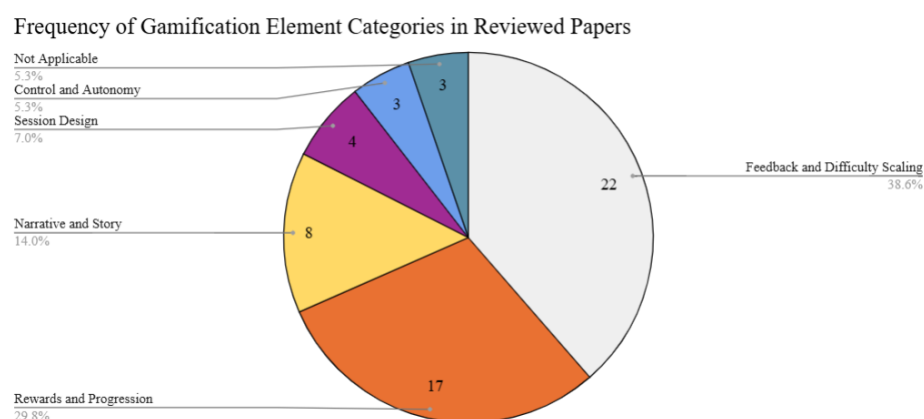


Figure 8. Frequency of gamification element categories in reviewed papers.

Speech rehabilitation games have demonstrated success in maintaining children’s motivation and engagement during therapy sessions through varied strategies, encompassing speech-controlled avatars, autonomous play modes, and speech recognition techniques.

The results show increasing support for child autonomy, diverse and effective interaction methods, growing AI integration, and strong use of motivational design. Speech recognition accuracy and hardware limitations remain key challenges. Together, these findings show a shift towards more adaptive, engaging, and accessible speech rehabilitation games. The next section discusses the gaps, opportunities for AI enhancement, and the need for more inclusive, scalable game design for children with speech disabilities.

4. Discussions and Future Work

This systematic review has given us a more coherent and bespoke literature of an important and evolving area of digital game rehabilitation for children with speech disabilities. Using four targeted questions, we are able to highlight areas of ecological and external validity in existing evidence, identify research gaps, establish the strength of evidence, and provide a foundation for meta-analysis.

We have identified clear evidence and guidance on the creation of such games. The crucial role of therapists in designing exercises and monitoring progress cannot be overstated, but the necessity for readily accessible, previously set exercises, particularly in underserved areas or for children without regular access to therapists, is apparent. The varying game designs, some reliant on therapist involvement and others promoting independent engagement, showcase the evolving area of speech rehabilitation games. The results of our analysis indicate a wide spectrum of approaches in terms of the degree of independence children exhibit while playing speech rehabilitation games. These games cater to therapists as a primary end-user, emphasising the role of expert guidance in optimising rehabilitation outcomes. This suggests a shift toward more inclusive and self-directed rehabilitation approaches, acknowledging the evolving role of technology in empowering children with speech disabilities. This emphasises the need for a balanced consideration of therapist guidance, peer support, and independent engagement, underscoring the importance of tailoring interventions to the unique needs and preferences of children with speech disabilities.

The need to minimise therapist or teacher involvement has been identified as a growing priority, especially in contexts with limited access to professionals. Studies, such as refs. [12,41], propose the integration of AI-based classifiers as a viable path forward. By automating the classification of phonemes or utterances, these tools would enable children to engage with therapy content more independently while reducing the need for real-time supervision and validation. This points toward a future where technology plays a more central role in both delivery and assessment.

The diversity in voice control methods caters to the individualised needs of children with varying speech disabilities, offering a range of exercises to enhance phonological performance. Current work underscores the importance of tailoring feedback approaches to suit the preferences and needs of individual users, acknowledging the multifaceted nature of speech rehabilitation. In essence, the various design choices in interaction modalities, voice control, and feedback mechanisms highlight the adaptability of speech rehabilitation games. These choices are pivotal in shaping user engagement, emphasising the need for a thoughtful and inclusive design approach to cater to the diverse needs of children with speech disabilities. The use of reward and punishment systems, as observed in the SEGA-ARM model [10], introduces a competitive element, fostering enthusiasm among players. Level progression serves as a motivational tool, providing continuous feedback on performance. Concepts of selective and personalised game design contribute to increased motivation and positive effects on children. The positive impact of reward systems on the child's motivation throughout the game emphasises the potential of gamification in speech rehabilitation.

The incorporation of AI, particularly NLP, in speech rehabilitation games involves critical decisions, starting with the selection of speech recognition algorithms. Background noise cancellation emerges as a key element, with the choice of microphone and hardware significantly influencing the efficiency and accuracy of NLP algorithms. The preference for Kinect microphones [40], for example, due to their noise suppression capabilities, shows the importance of optimising hardware to create natural speech environments and minimise processing requirements. The role of AI and NLP in speech rehabilitation games is pivotal, influencing algorithm selection, hardware optimisation, dataset quality, and the ongoing pursuit of accuracy improvements. The intersection of human expertise and AI capabilities holds promise for creating more effective and personalised interventions in the field of speech rehabilitation.

Our findings also suggest that gamified approaches not only engage children more effectively but also lead to improved speech accuracy. In [41], children with dysarthria demonstrated fewer mispronunciations and a clear preference for game-based collection

methods over traditional adult imitation. This preference was consistent across both subjective engagement surveys and objective speech error rates. Such evidence supports the case for continued development of child-centred, interactive methods that embed speech tasks within rewarding, playful experiences.

There are areas that our review has shown to need addressing. The concept of what is considered “fun” differs from one person to another, as mentioned in [36]. Data collection and analysis from these games is an area for improvement because most games rely on therapists to evaluate the data, give feedback, and adjust accordingly. Improving speech recognition accuracy is still necessary and response time and latency is an active frustration element in most games. We are hypothesising that with NLP technology becoming more commonplace, this technical factor will be overcome naturally.

With the minimal multiple disabilities targeted in the solutions, mentioned in Section 3.2, future work should be accessible to people with more than one disability. While Abdoulqadir, Loizides, and Hoyos focus on game applications targeting dual disabilities [50], more research is necessary for serious games targeting more than one disability. Similarly, most solutions are not accessible to minority languages despite the fact that preferred languages are recommended to be considered by researchers and developers [24]. Speech rehabilitation games accessible for minority languages should be addressed more in future research.

In Section 3.6, we showed the different elements used in games to support speech therapy. Gamified speech therapy applications implement motivational elements to keep and enhance children’s motivation. This has been applied and studied in schools specific for deaf and hard-of-hearing students [24]. Speech therapy games support inclusive education by providing opportunities for children with speech learning disabilities to practice in class [24]. Moreover, most studies showed autonomy and enhanced children’s engagement, as detailed in Table A3. Helping students practice independently shifts classroom focus from teacher-focused to student-focused. The games are accessible for students with speech impairments, aligning with inclusive education goals. However, there are pedagogical limitations within the speech learning solutions. Most solutions are designed for individual use, with limited support for peer-to-peer interaction. They are often culturally and linguistically specific. To better support diverse classroom environments, future research should prioritise multilingual and customisable designs. This is important for educational settings with diverse students from different backgrounds.

In summary, the results indicate that speech rehabilitation games effectively maintain motivation and engagement through diverse gamification strategies. While challenges in directly linking games to speech performance persist, the positive outcomes reinforce the potential of digital games in enhancing children’s speech rehabilitation. Notably, the results for a sustained vowel game show that “all children improved their achieved MPT (maximum production time) progressively during the user test” [38]. This shows the positive effects of digital games on children’s speech performance. Few papers mention the effectiveness of speech rehabilitation games on the child’s actual speech performance. AI and NLP are evolving and have large potential, which shows the need for continued research and innovation in gamified approaches to address the unique needs of children with speech disabilities. Through this work, we also aim to address the need for more domain-specific and flexible design guidelines that are now specific to certain disabilities and presented at a higher level.

Limitations of the Study

This review has multiple limitations that should be acknowledged. Firstly, it focuses on the technological, interactional, and motivational aspects of speech rehabilitation games, rather than evaluating their clinical efficacy. It does not assess whether these games result in measurable therapeutic improvements as judged by speech–language professionals.

Secondly, while the study includes a diverse set of papers, most target children with isolated speech disabilities, with limited inclusion of dual or complex disability scenarios. The results cannot be generalised to broader populations. Thirdly, the review is based on English-language publications and may under-represent work in minority languages. Most of the included studies report short-term evaluations or prototypes, and few provide longitudinal data on sustained engagement or long-term therapeutic outcomes. Finally, this review falls under the Human–Computer Interaction (HCI) field. It does not consider psychological or non-digital interventions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/info16070599/s1>, PRIMASA checklist. Reference [51] are cited in the Supplementary Materials.

Author Contributions: C.A.: Formal analysis, investigation, resources, visualisation, writing—original draft preparation, writing—review and editing. F.L.: Conceptualisation, methodology, validation, funding acquisition, resources, supervision, writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by GOOGLE.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to express our sincere gratitude to Google for their invaluable support and feedback throughout this project. In particular, we are deeply thankful to Abhipreeti Chauhan, Karo Caran, and Victor Tsaran for their consistent guidance, insightful feedback, and inspiration that shaped this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. ASR Technologies Used in Children’s Speech Rehabilitation Games

Table A1. Comparison of ASR technologies in game-based systems.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[35]	Template-Based	Template Matching (TM): 72% F1, Goodness of Pronunciation: 69% F1	Template Matching outperformed GOP; TM was better at identifying mispronunciations. Children improved with both.	Template Matching (TM), GOP algorithm using Kaldi acoustic models trained on Librispeech corpus (960 h of adult speech)

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[52]	Rule-Based/ Hybrid	Not mentioned	Speech scoring system validated in small studies; storybook format found most motivating. Offline ASR effective.	Offline ASR, custom speech scoring
[34]	Rule-Based/ Hybrid	Not mentioned	No direct ASR accuracy reported; focuses on integration with Alexa and making therapy more engaging through automation.	Alexa voice assistant, AI scoring
[11]	Other/ Unspecified	70% threshold for good pronunciation	Children could control the game via voice; 70% score considered a good result. Emphasises motivation and independence.	Word Detection Package, Windows UDP Voice Recognition Server
[44]	Data Collection/ Preprocessing	Not mentioned (conceptual framework)	No ASR tested; proposes a crowd-sourced dataset creation and validation model to improve ASR performance in therapy.	Collaborative dataset with therapist validation
[9]	Rule-Based/ Hybrid	Not mentioned	Game offers visual feedback for /s/, /z/, etc.; real-time processing used; accuracy and scalability not evaluated.	Custom audio analysis tool for sibilant sounds

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[45]	Other/ Unspecified	Not mentioned	Supports parent-guided home use with simple detection logic; emphasises usability more than recognition precision.	Sound detection and interactive speech interface
[17]	Other/ Unspecified	Therapist-rated as accurate enough for dysphonic voices	System filters ambient noise and tracks pitch; FFT preferred over auto-correlation due to sensitivity needs.	Pitch and loudness detection via Fast Fourier Transform (FFT), Kinect mic array
[18]	Other/ Unspecified	Not mentioned	Focus is on structured language interaction and logging; limited use of automated speech recognition.	Therapist-guided interaction; voice and scenario logging
[26]	Machine Learning/ AI-Based	86.1% (cross-validation)	Promising accuracy for phoneme-level classification, not real-time but useful for therapeutic monitoring.	Deep Neural Network (DNN)-based classifier for sibilant consonant detection
[28]	Other/ Unspecified	Not quantified	Prototype shows motivational benefit despite ASR limitations, aims for future precision improvements.	Conversational speech recognition (prototype)

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[2]	Other/ Unspecified	Not consistent across tools	ASR adoption remains fragmented; real-time, child-friendly ASR is still underdeveloped; environmental noise challenges.	Varied across reviewed studies
[36]	Machine Learning/ AI-Based	Not evaluated (manual marking used)	ASR integration planned; manual system ensures evaluation consistency but limits scalability.	Wizard of Oz (manual) with plans for Pocket-Sphinx
[23]	Other/ Unspecified	Not mentioned	Focuses on autonomous engagement; does not evaluate ASR.	Not specified
[30]	Other/ Unspecified	Not quantified	ASR used for independent vocabulary evaluation; supports scalable remote therapy.	Integrated speech recognition in mobile app (ASR library unspecified)
[22]	Other/ Unspecified	Not applicable	Focus on ICTs and their general benefit; no use of ASR.	Not implemented (focuses on ICT tools)
[20]	Other/ Unspecified	Not applicable	Describes future potential of ASR and IoT for home-based therapy; not yet implemented.	Planned conversational interfaces
[31]	Rule-Based/ Hybrid	Not quantified	Indicates speech quality improvement tracking but lacks model-specific detail.	Custom scoring algorithm based on speech input and pronunciation comparison

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[32]	Other/ Unspecified	Not quantified; adjusted with calibration and thresholding	Highlights the value of lightweight real-time vowel recognition over full ASR for engagement and responsiveness.	Formant-based vowel detection using LPC and FFT; implemented in C++ with OpenFrameworks
[47]	Other/ Unspecified	High robustness claimed, no specific number	Strong emphasis on robustness to background noise for practical settings.	Robust HMM-based phoneme recogniser with noise-robust features
[27]	Other/ Unspecified	73.98% to 85.93%	Demonstrates reliable classification for different voice exercises with low false negatives.	SVM-based classifier for sustained vowels and pitch variation using MFCCs and F0 features
[10]	Other/ Unspecified	Not applicable	Conceptual metamodel; no empirical results but supports integration of speech recognition as part of user modeling.	Proposed framework includes phoneme recognition; no specific implementation
[24]	Machine Learning/ AI-Based	Not quantified, but describes 'robust detection' and 'usable feedback'	Feasible and acceptable in resource-constrained settings over the long term.	Specific engine not stated
[40]	Other/ Unspecified	Moderate; poor recognition for children's accents observed	Popular gameplay can boost motivation, but recognition issues must be addressed.	Microsoft Speech SDK with Kinect microphone array
[21]	Machine Learning/ AI-Based	High (95% overall, 97.5% phoneme accuracy in pilot)	Highly promising; customisable for cleft-specific errors and works on mobile	PocketSphinx via OpenEars speech recognition API for iOS

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[53]	Machine Learning/ AI-Based	Not quantified in this paper.	Users reported positive engagement; ASR viable with offline use cases.	Not stated
[14]	Machine Learning/ AI-Based	Perceived as neutral-to-good by users (Likert average 3.2–4.2)	Effective and offline-friendly; supports critical listening and user-specific adaptation.	PocketSphinx via OpenEars and RapidEars (offline)
[38]	Other/ Unspecified	Not mentioned	Pitch-based voice input is viable for therapy reinforcement but lacks full ASR functionality.	Pitch detection (not specific)
[46]	Other/ Unspecified	Not mentioned	Highlights motivational value of games but lacks discussion on ASR tools or outcomes.	Not specified
[49]	Other/ Unspecified	Not quantified	Focus is on UX methodology and assessment; ASR is minimal and used for basic voice response.	Microphone input with pitch/timbre evaluation
[15]	Other/ Unspecified	Not mentioned	Game supports phonological therapy via simplified input; ASR engine details not specified.	Concept-matching and speech stimulus response
[13]	Machine Learning/ AI-Based	Not measured; assumed platform-native	Smart assistants enhance motivation and scheduling; effectiveness tied to ecosystem, not custom ASR.	Voice assistant platforms (Amazon Alexa, Google Home)

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[41]	Other/ Unspecified	Not mentioned (focus on feasibility and motivation)	Promising for generating child speech datasets but lacks immediate feedback or therapeutic use.	Gamified speech data collection interface with ASR analysis post hoc
[12]	Other/ Unspecified	Not evaluated directly	ASR integrated indirectly; system focuses on therapist-aided assessment rather than full automation.	Embedded phonetic-phonological processing via therapist dashboard
[48]	Other/ Unspecified	Not reported; proposal-focused paper	Highlights potential of adaptive speech interfaces; lacks empirical ASR results.	Exploratory use of speech recognition and voice features
[29]	Other/ Unspecified	Not applicable	Focus is on structured audio input and feedback rather than ASR analysis.	Tablet-based phoneme training app; audio playback, no ASR
[39]	Machine Learning/ AI-Based	Acknowledges ASR accuracy challenges for disordered speech	Commercial VAs show potential but current ASR not robust enough for speech disorder needs without retraining.	Voice assistants, ASR
[43]	Other/ Unspecified	Not mentioned	No full ASR used; pitch-based sound input replaces speech recognition for engagement and tracking.	Pitch detection via Audio Input Handler using whistles
[16]	Other/ Unspecified	Not specified	game focuses on practice rather than advanced speech analysis.	Basic voice analysis tools

Table A1. Cont.

Reference	Technology Category	Accuracy	ASR Technology Conclusion	Details
[42]	Rule-Based/ Hybrid	Word Error Rate (WER) reduced to 24.8% using augmented data	Custom-trained models significantly improved child speech recognition and user engagement.	Custom ASR model trained on Slovak children’s speech using wav2vec2
[25]	Machine Learning/ AI-Based	Not quantified; challenges with phoneme detection noted	Whisper-Local shows promise but struggles with precise phonetic level recognition for therapy.	Whisper-Local used for speech-to-text
[5]	Rule-Based/ Hybrid	Confidence-based matching (specific rates not disclosed)	Custom ASR models support word articulation scoring; confidence levels drive game feedback and progression.	Pre-trained and custom-trained speech recognition models using Raspberry Pi
[54]	Machine Learning/ AI-Based	Not quantified, but personalised to individual phonetic inventories	Voiceitt shows strong potential for severe impairments by enabling real-time, individualised Augmentative and Alternative Communication (AAC) support using AI-driven speech recognition.	Voiceitt®-AI-based non-standard speech recognition system using deep learning and pattern classification

Appendix B. Interaction Modalities

Table A2. Interaction modalities used in selected studies.

Reference	AI Techniques	Interaction Modalities	Target Age Group/ Participant Age Group
[35]	Template Matching, GOP, Automatic Mispronunciation Detection	Touch (tablet joystick/buttons), voice input, audio feedback	Children aged 5–12

Table A2. Cont.

Reference	AI Techniques	Interaction Modalities	Target Age Group/ Participant Age Group
[52]	Offline Speech Recognition, Dynamic Curriculum	Voice input, audio feedback, storybook-style navigation	Not explicitly stated
[34]	AI-based correction, Alexa Skill integration	Voice input (via Alexa), smart home automation	Children aged 4–8
[11]	Voice Command Detection, Speech Scoring via UDP Server	voice input, game avatar control, visual and audio guidance	Children aged 2–6
[44]	Collaborative dataset generation, Gamified data collection	Voice recording, therapist validation	Not mentioned
[9]	Custom audio analysis (sibilant energy extraction)	Visual waveform display, voice input	Children aged 6–10
[45]	Sound detection, Therapist-defined interactive feedback	Voice input, parental and therapist interface	Not explicitly stated; intended for home-based use by children
[37]	Phonation time analysis, Intensity threshold detection	Voice input only, audio feedback	Children with articulation challenges (not specifically aged)
[17]	FFT-based pitch and loudness estimation, Kinect microphone array	Voice input (pitch-based control), visual feedback	Adolescents and adults with dysphonia (generalisable elements for older children)
[18]	Therapist-guided recording and object interaction logging	Voice input, 3D object interaction, therapist-controlled environment	Children aged 5 (tested), supports various therapy scenarios
[26]	Deep Neural Network (sibilant classification)	Voice input, visual feedback	Not specified; children with sibilant errors
[28]	Conversational agent prototype, visual/audio reinforcement	Voice input, visual prompts, adaptive feedback	Not explicitly stated
[2]	Analysis of AI-based games (ASR, NLP, Feedback)	Varies; includes voice, touch, gesture	Primarily 4–12 years (based on included studies)
[36]	Planned PocketSphinx ASR, manual SLP scoring	Voice input, touch control, custom prompts	Children aged 2–14
[23]	Autonomous task triggering, Interaction tracking	Touch input, visual/audio prompts	Children and young adults with Down Syndrome

Table A2. Cont.

Reference	AI Techniques	Interaction Modalities	Target Age Group/ Participant Age Group
[30]	ASR for vocabulary recognition, Manual review for other stages	Voice input, Video/audio uploads, virtual pet	Children post-cleft lip surgery (unspecified age)
[22]	Overview of ICT tools (not specific to AI)	Software-based, therapist-controlled tools	Not explicitly mentioned
[20]	Smart home orchestration via EUD, Proposed ASR integration	Voice input, tablet control, IoT devices (lights, TV)	Children in home therapy (not explicitly stated)
[31]	Speech quality analysis and task-specific scoring	Voice input, visual feedback through game UI	Not explicitly stated
[32]	Formant tracking via Linear Predictive Coding (LPC)/ FFT for vowel detection	Voice input, real-time retro game interface	Not explicitly stated
[47]	the Naive Bayes (NB), Support Vector Machines (SVM) and Kernel Density Estimation (KDE) were compared; The best results were obtained with the flat KDE with Silverman's bandwidth using MFCCs	Voice input, speech playback, visual feedback	Not explicitly stated
[27]	SVM classifiers for sustained vowel and pitch variation	Voice input, visual feedback via screen interface	Not explicitly stated
[10]	Model-driven design incorporating phoneme recognition and user profiling	Voice input, audio-visual interaction modules (conceptual)	Framework for various user types, including children with hearing loss
[24]	specific technology not stated, multilingual model support	Voice input, visual feedback via game characters	Not explicitly stated
[40]	Speech analysis using Microsoft SDK, loudness-based input	Voice input, Kinect gestures, visual rewards	7 and 10
[21]	Speech pattern matching using PocketSphinx with cleft-specific adaptation	Voice input, touch screen interaction, storybook format	Children aged 2–3
[53]	PocketSphinx, phoneme scoring	Voice input only, Visual/audio game feedback	Not explicitly stated

Table A2. Cont.

Reference	AI Techniques	Interaction Modalities	Target Age Group/ Participant Age Group
[14]	Custom phoneme scoring, offline ASR via OpenEars	Voice input, audio prompts, tablet interaction	Target children, tested on adults 24–31
[38]	Pitch analysis (signal processing only, no full ASR)	Voice input (sustained vowel)	Not explicitly stated
[46]	Not specified (exploratory discussion)	Potentially touch and voice (not evaluated)	Not explicitly stated
[49]	UX framework development with voice-based input consideration	Voice input, therapist-led observation, touch	Children with cochlear implants (early to mid-childhood)
[15]	Speech stimulus and concept-response (simplified speech processing)	Voice input, game-based touch interface	Children aged 3–6 with phonological disorders
[13]	Voice assistant orchestration (Alexa, Google Home)	Voice input, smart device interaction (lights)	Children in home-based therapy (general use case)
[41]	Speech data collection with planned ASR analysis	Voice input, touch interface, animated characters	Children aged 5–8 with dysarthria
[12]	Therapist-controlled phonological data processing	Voice input, tablet game interface, therapist dashboard	Not explicitly stated
[48]	Proposal of adaptive ASR-based interfaces for therapy	Voice input (planned), adaptive feedback	Children with speech disabilities (ages unspecified)
[29]	Structured phoneme training, No ASR	Touch input, audio playback, visual rewards	Study 1: mean age of 6 years and 6 months; Study 2: mean age of 7 years 9 months
[39]	Commercial ASR and NLP	Voice input, smart assistant responses, screen prompts	Children with speech impairments (general home use)
[43]	Pitch detection and audio monitoring (no ASR)	Voice input (whistle/pitch), visual mobile interface	Children aged 4–10 with orofacial myofunctional disorders
[16]	Simple voice analysis tools (unspecified)	Voice input, basic visual/audio prompts	Children in early speech therapy (ages not specified)
[42]	Wav2Vec2 model fine-tuned on child speech	Voice input, movement-based game interface	Children speaking Slovak (age not specified)

Table A2. Cont.

Reference	AI Techniques	Interaction Modalities	Target Age Group/ Participant Age Group
[25]	Whisper-Local model for real-time speech recognition	Voice input, visual/audio feedback, directional movement	Children with speech impairments (ages not specified)
[5]	Custom-trained language models for speech articulation scoring	Voice input, mouse (in shooter), keyboard (optional), visual/audio feedback	Children aged 6–10
[54]	Deep learning, pattern clustering, voice donor output, speaker-dependent ASR	Voice input, real-time AI interpretation, voice output	Children and adults with speech disabilities, such as cerebral palsy, dysarthria, and autism

Appendix C. Details of Gamification Tactics Mentioned in the Selected Papers

Table A3. details of gamification elements used in speech therapy games and their reported impact on user engagement over time.

Reference	Gamification Element Category	Gamification Element Details	Engagement Over Time
[35]	Avatar and Customisation, Rewards and Progression	Avatars, coins, in-game store, power-ups, progression through levels	Daily session cap, personalisation, story-based progression, high user enjoyment reported
[52]	Feedback and Difficulty Scaling, Narrative and Story	Narrative storybook format, characters, adaptive difficulty	User-centred design, sustained motivation reported across studies
[34]	Feedback and Difficulty Scaling	Smart reminders, voice assistant interaction	Not emphasised; focus on automation and convenience
[11]	Avatar and Customisation, Control and Autonomy	Avatar control, voice-controlled game actions, object collection	Use of guidance arrow, repetitive play encouraged, no negative feedback
[44]	Rewards and Progression	Gamified data collection (star ratings, score-based validation)	Conceptual only; no long-term play evaluation conducted
[9]	Feedback and Difficulty Scaling, Rewards and Progression	Visual waveform feedback, score thresholds, progress bar	Session-based progression, configurable thresholds by therapist
[45]	Rewards and Progression	Task completion tracking, rewards	Designed for routine home use; long-term engagement monitored by caregivers

Table A3. Cont.

Reference	Gamification Element Category	Gamification Element Details	Engagement Over Time
[37]	Avatar and Customisation, Feedback and Difficulty Scaling	Avatar animation (flying bird), real-time voice control, feedback via game success/failure	Adaptive difficulty, multiple intensity levels; plans for scenario expansion
[17]	Feedback and Difficulty Scaling	Pitch-controlled visual objects, real-time feedback, points system	Designed for continuous repetition; therapist adjustable goals
[18]	Narrative and Story	Explorable 3D environment, object manipulation, scenario-based storytelling	Customisable scenarios, therapist-led exploration, voice logs for follow-up
[26]	Feedback and Difficulty Scaling	Feedback animations, task scoring	Emphasised pronunciation monitoring; designed for iterative use
[28]	Narrative and Story	Narrative, empathetic characters, procedural content generation, visual/audio rewards	Replayability via PCG; immersion driven by plot and character empathy; volume-based input challenges
[2]	Avatar and Customisation, Feedback and Difficulty Scaling	Variable; includes points, character-based feedback, customisable environments	Highlights sustained use challenges; low self-esteem after several failures; reviews importance of meaningful rewards and personalisation
[36]	Avatar and Customisation, Rewards and Progression	Avatars, level progression, coin and star collection, rewards	Therapist-driven content adjustment; motivational elements include store, upgrades, and interactive feedback
[23]	Feedback and Difficulty Scaling, Rewards and Progression	Autonomous play mode, feedback animations, score keeping	Increased play duration and independence; tested on users with Down Syndrome to assess motivation
[30]	Session Design	Virtual pet feeding, candy collection, session rewards	Motivation sustained through pet care dynamics; repeated sessions encouraged by daily decay of pet health

Table A3. Cont.

Reference	Gamification Element Category	Gamification Element Details	Engagement Over Time
[22]	Not Applicable	Not applicable (review of technologies)	Mentions importance of user motivation but does not analyse tactics directly
[20]	Narrative and Story, Control and Autonomy	Smart home fantasy scenarios, character response via devices	Focus on emotional reinforcement and parental configurability; plans for immersive future development
[31]	Feedback and Difficulty Scaling	Game-like scoring, immediate feedback, audio rewards	Tracks improvement over sessions; encourages continued effort with task-based incentives
[32]	Feedback and Difficulty Scaling, Control and Autonomy	Retro game mechanics, vowel-triggered character control, visual reaction	Sustained engagement through nostalgia-style play and real-time voice response
[47]	Rewards and Progression	Progress bar (ice cream) and reward (virtual button)	Motivates accurate phoneme production using score and repetition logic
[27]	Rewards and Progression	Progress indicators, visual rewards, real-time correctness display	Encourages voice control improvement by minimising false negatives; increases confidence
[10]	Feedback and Difficulty Scaling	Metamodel supports points, feedback loops, challenge levels	Design-driven personalisation aims to retain users by adapting difficulty and reward schemes
[24]	Feedback and Difficulty Scaling, Rewards and Progression	Character feedback, level-based rewards, visual progress tracking	Long-term classroom deployment; increased confidence and repeat play observed
[40]	Avatar and Customisation	Pac-Man style gameplay, voice-triggered avatar, power-up mechanics	Popular game structure increases motivation, though ASR accuracy limits usability
[21]	Feedback and Difficulty Scaling, Narrative and Story, Rewards and Progression	Story-driven progression, pop-the-balloon mini-games, auditory feedback	Motivates cleft speech repetition with scenario advancement after repeated attempts

Table A3. Cont.

Reference	Gamification Element Category	Gamification Element Details	Engagement Over Time
[53]	Rewards and Progression	Score system, character animation, word repetition tasks	Children enjoyed progress tracking and character response; found voice interaction intuitive
[14]	Feedback and Difficulty Scaling, Rewards and Progression, Session Design	Point scoring, session progression, patient-specific challenges	Critical listening emphasised over pure game reward; motivates goal achievement with therapist-set plans
[38]	Rewards and Progression	Visual progression	Visual reinforcement encourages vocal control; suitable for short, repetitive sessions
[46]	Not Applicable	Games as a motivational wrapper (discussion only)	Proposes using game elements to overcome training anxiety and boost participation
[49]	Feedback and Difficulty Scaling, Narrative and Story	Narrative elements, multi-sensory feedback	Focus on personalisation and sensory accessibility for engagement measurement
[15]	Feedback and Difficulty Scaling	Colourful animation, audio response to correct/incorrect input	Designed to provide entertaining structure to speech sound exercises, aligned with therapy goals
[13]	Session Design	Daily challenges, verbal praise, smart home cues (lights)	Integrates therapy into household routines; supports emotional motivation via device responses
[41]	Feedback and Difficulty Scaling	Character animations	Children showed high participation and motivation during speech recording sessions
[12]	Feedback and Difficulty Scaling	Game-like assessment interface, animated feedback	Game structure improves cooperation and enjoyment during assessment; therapists report better child focus

Table A3. Cont.

Reference	Gamification Element Category	Gamification Element Details	Engagement Over Time
[48]	Narrative and Story	Proposed use of audio-visual storytelling, point rewards	Planned use of emotionally engaging interfaces to motivate therapy adherence
[29]	Feedback and Difficulty Scaling, Rewards and Progression	Progress charts, colorful feedback, sound playback	Focus on therapist-defined targets; game elements used to maintain attention in early learners
[39]	Feedback and Difficulty Scaling, Session Design	Conversational prompts, daily therapy reminders, praise	Encourages routine formation through smart assistant dialogue and friendly voice interactions
[43]	Rewards and Progression	Animation, Whistle-driven progress, visual rewards	Reinforces participation through sound-controlled progression; motivates daily practice with app interaction
[16]	Feedback and Difficulty Scaling	Colorful prompts, word repetition scoring	Encourages pronunciation through repetition and animated feedback; suitable for early intervention
[42]	Rewards and Progression	Game character movement tied to ASR output, score display	Speech quality influences in-game control, increasing repetition and goal-oriented speaking
[25]	Rewards and Progression	Level-based progression, visual/audio cues, score system	Voice triggers movement and progression; integrates usability testing for continued motivation
[5]	Feedback and Difficulty Scaling	Vertical shooter and adventure platformer, word-triggered obstacles, adaptive difficulty, audio-visual rewards	Confidence-based gameplay promotes repetition and motivation; two distinct styles of gameplay suit varied interaction abilities
[54]	Not Applicable	None (not game-based)	Motivation derived from restored communication ability; app promotes inclusion, autonomy, and real-world interactions

Appendix D. Resources Used to Answer the Research Questions

Table A4. Research papers used to answer each research question.

Research Question 1	Research Question 2	Research Question 3	Research Question 4
[10]	[9]	[34]	[34]
[9]	[26]	[30]	[49]
[13]	[27]	[27]	[10]
[10]	[28]	[45]	[47]
[14]	[47]	[19]	[22]
[19]	[18]	[48]	[23]
[17]	[31]	[21]	[35]
[15]	[11]	[2]	[47]
[18]	[21]	[40]	[35]
[24]	[19]	[39]	[23]
[11]	[21]	[25]	[32]
[21]	[24]	[16]	[33]
[12]	[12]	[42]	[11]
[25]	[25]		[16]
[16]	[42]		[43]
[43]	[43]		
	[5]		

References

- Fitch, J. Pediatric Speech Disorder Diagnoses More than Doubled Amid COVID-19 Pandemic. 2023. Available online: <https://www.contemporarypediatrics.com/view/pediatric-speech-disorder-diagnoses-more-than-doubled-amid-covid-19-pandemic> (accessed on 21 January 2025).
- Saeedi, S.; Bouraghi, H.; Seifpanahi, M.S.; Ghazisaeedi, M. Application of digital games for speech therapy in children: A systematic review of features and challenges. *J. Healthc. Eng.* **2022**, *2022*, 4814945. [CrossRef] [PubMed]
- Larson, K. Serious games and gamification in the corporate training environment: A literature review. *TechTrends* **2020**, *64*, 319–328. [CrossRef]
- Jaddoh, A.; Loizides, F.; Rana, O. Interaction between people with dysarthria and speech recognition systems: A review. *Assist. Technol.* **2023**, *35*, 330–338. [CrossRef] [PubMed]
- Nicolaou, K.; Carlton, R.; Jaddoh, A.; Syed, Y.; James, J.; Abdoulqadir, C.; Loizides, F. Game based learning rehabilitation for children with speech disabilities: Presenting two bespoke video games. In *BCS HCI '23, Proceedings of the 36th International BCS Human-Computer Interaction Conference, York, UK, 28–29 August 2023*; BCS Learning & Development Ltd.: Swindon, UK, 2023.
- Moher, D.; Shamseer, L.; Clarke, M.; Gherzi, D.; Liberati, A.; Petticrew, M.; Shekelle, P.; Stewart, L.A.; Group, P.P. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **2015**, *4*, 1–9. [CrossRef] [PubMed]
- Schlosser, R.W.; O'Neil-Pirozzi, T.M. Problem Formulation in Evidence-based Practice and Systemic Reviews. *Contemp. Issues Commun. Sci. Disord.* **2006**, *33*, 5–10. [CrossRef]
- Parentctrhub. Categories of Disability Under Part B of IDEA—Center for Parent Information and Resources. 2024. Available online: <https://www.parentcenterhub.org/713categories/#speech> (accessed on 3 January 2025).
- Anjos, I.; Grilo, M.; Ascensão, M.; Guimarães, I.; Magalhães, J.; Cavaco, S. A serious mobile game with visual feedback for training sibilant consonants. In *Proceedings of the International Conference on Advances in Computer Entertainment, London, UK, 14–16 December 2017*; pp. 430–450.
- Céspedes-Hernández, D.; Pérez-Medina, J.L.; González-Calleros, J.M.; Rodríguez, F.J.Á.; Muñoz-Arteaga, J. Sega-arm: A metamodel for the design of serious games to support auditory rehabilitation. In *Proceedings of the XVI International Conference on Human Computer Interaction, Vilanova i la Geltrú, Spain, 7–9 September 2015*; pp. 1–8.

11. Nasiri, N.; Shirmohammadi, S.; Rashed, A. A serious game for children with speech disorders and hearing problems. In Proceedings of the 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), Perth, Australia, 2–4 April 2017; pp. 1–7.
12. Antunes, I.; Antunes, A.; Madeira, R.N. Designing Serious Games with Pervasive Therapist Interface for Phonetic-Phonological Assessment of Children. In Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia, Vienna, Austria, 3–6 December 2023; pp. 532–534.
13. Cassano, F.; Pagano, A.; Piccinno, A. Supporting speech therapies at (smart) home through voice assistance. In Proceedings of the International Symposium on Ambient Intelligence, Salamanca, Spain, 6–8 October 2021; pp. 105–113.
14. Duval, J.; Rubin, Z.; Segura, E.M.; Friedman, N.; Zlatanov, M.; Yang, L.; Kurniawan, S. SpokeIt: Building a mobile speech therapy experience. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Barcelona, Spain, 3–6 September 2018; pp. 1–12.
15. Madeira, R.N.; Macedo, P.; Reis, S.; Ferreira, J. Super-fon: Mobile entertainment to combat phonological disorders in children. In Proceedings of the 11th Conference on Advances in Computer Entertainment Technology, Madeira, Portugal, 11–14 November 2014; pp. 1–4.
16. Haluška, R.; Pleva, M.; Reňák, A. Development of Support Speech Therapy Game. In Proceedings of the 2024 International Conference on Emerging eLearning Technologies and Applications (ICETA), Stary Smokovec, Slovakia, 24–25 October 2024; pp. 174–179.
17. Lv, Z.; Esteve, C.; Chirivella, J.; Gagliardo, P. A game based assistive tool for rehabilitation of dysphonic patients. In Proceedings of the 2015 3rd IEEE VR International Workshop on Virtual and Augmented Assistive Technology (VAAT), Arles, France, 23 March 2015; pp. 9–14.
18. Cagatay, M.; Ege, P.; Tokdemir, G.; Cagiltay, N.E. A serious game for speech disorder children therapy. In Proceedings of the 2012 7th International Symposium on Health Informatics and Bioinformatics, Nevsehir, Turkey, 19–22 April 2012; pp. 18–23.
19. Lan, T.; Aryal, S.; Ahmed, B.; Ballard, K.; Gutierrez-Osuna, R. Flappy voice: An interactive game for childhood apraxia of speech therapy. In Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play, Toronto, ON, Canada, 19–21 October 2014; pp. 429–430.
20. Cassano, F.; Piccinno, A.; Regina, P. End-user development in speech therapies: A scenario in the smart home domain. In Proceedings of the End-User Development: 7th International Symposium, IS-EUD 2019, Hatfield, UK, 10–12 July 2019; Proceedings 7; pp. 158–165.
21. Rubin, Z.; Kurniawan, S. Speech adventure: Using speech recognition for cleft speech therapy. In Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments, Rhodes, Greece, 29–31 May 2013; pp. 1–4.
22. Drosos, K.; Voniati, L.; Christopoulou, M.; Kosma, E.I.; Chronopoulos, S.K.; Tafiadis, D.; Peppas, K.P.; Toki, E.I.; Ziavra, N. Information and Communication Technologies in Speech and Language Therapy towards enhancing phonological performance. In Proceedings of the 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 21–23 October 2021; pp. 187–192.
23. Escudero-Mancebo, D.; Corrales-Astorgano, M.; Cardeñoso-Payo, V.; González-Ferreras, C. Evaluating the impact of an autonomous playing mode in a learning game to train oral skills of users with down syndrome. *IEEE Access* **2021**, *9*, 93480–93496. [[CrossRef](#)]
24. Nanavati, A.; Dias, M.B.; Steinfeld, A. Speak up: A multi-year deployment of games to motivate speech therapy in India. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Canada, 21–26 April 2018; pp. 1–12.
25. Baranyi, R.; Weber, L.; Aigner, C.; Hohenegger, V.; Winkler, S.; Grechenig, T. Voice-Controlled Serious Game: Design Insights for a Speech Therapy Application. In Proceedings of the 2024 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 14–15 November 2024; pp. 1–4.
26. Costa, W.; Cavaco, S.; Marques, N. Deploying a Speech Therapy Game Using a Deep Neural Network Sibilant Consonants Classifier. In Proceedings of the Progress in Artificial Intelligence: 20th EPIA Conference on Artificial Intelligence, EPIA 2021, Virtual Event, 7–9 September 2021; Proceedings 20; pp. 596–608.
27. Diogo, M.; Eskenazi, M.; Magalhaes, J.; Cavaco, S. Robust scoring of voice exercises in computer-based speech therapy systems. In Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; pp. 393–397.
28. Duval, J.; Rubin, Z.; Goldman, E.; Antrilli, N.; Zhang, Y.; Wang, S.H.; Kurniawan, S. Designing towards maximum motivation and engagement in an interactive speech therapy game. In Proceedings of the 2017 Conference on Interaction Design and Children, Stanford, CA, USA, 27–31 June 2017; pp. 589–594.
29. Gačnik, M.; Starčič, A.I.; Zaletelj, J.; Zajc, M. User-centred app design for speech sound disorders interventions with tablet computers. *Univers. Access Inf. Soc.* **2018**, *17*, 821–832. [[CrossRef](#)]

30. Garay, A.P.A.; Benites, V.S.V.; Padilla, A.B.; Galvez, M.E.C. Implementation of a solution for the remote management of speech therapy in postoperative cleft lip patients using speech recognition and gamification. In Proceedings of the 2021 IEEE Engineering International Research Conference (EIRCON), Lima, Peru, 27–29 October 2021; pp. 1–4.
31. Nasiri, N.; Shirmohammadi, S. Measuring performance of children with speech and language disorders using a serious game. In Proceedings of the 2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, Italy, 11–13 June 2017; pp. 15–20.
32. Tan, C.T.; Johnston, A.; Bluff, A.; Ferguson, S.; Ballard, K.J. Retrogaming as visual feedback for speech therapy. In Proceedings of the SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications, Shenzhen, China, 3–6 December 2014; pp. 1–5.
33. Tori, A.A.; Tori, R.; dos Santos Nunes, F.D.L. Serious game design in health education: A systematic review. *IEEE Trans. Learn. Technol.* **2022**, *15*, 827–846. [[CrossRef](#)]
34. Barletta, V.; Calvano, M.; Curci, A.; Piccinno, A. A New Interactive Paradigm for Speech Therapy. In Proceedings of the IFIP Conference on Human-Computer Interaction, York, UK, 28 August–1 September 2023; pp. 380–385.
35. Hair, A.; Ballard, K.J.; Markoulli, C.; Monroe, P.; Mckechnie, J.; Ahmed, B.; Gutierrez-Osuna, R. A longitudinal evaluation of tablet-based child speech therapy with Apraxia World. *ACM Trans. Access. Comput. (TACCESS)* **2021**, *14*, 1–26. [[CrossRef](#)]
36. Hair, A.; Monroe, P.; Ahmed, B.; Ballard, K.J.; Gutierrez-Osuna, R. Apraxia world: A speech therapy game for children with speech sound disorders. In Proceedings of the 17th ACM Conference on Interaction Design and Children, Trondheim, Norway, 19–22 June 2018; pp. 119–131.
37. Lopes, M.; Magalhães, J.; Cavaco, S. A voice-controlled serious game for the sustained vowel exercise. In Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, Osaka, Japan, 9–12 November 2016; pp. 1–6.
38. Lopes, V.; Magalhães, J.; Cavaco, S. Sustained Vowel Game: A Computer Therapy Game for Children with Dysphonia. In Proceedings of the INTERSPEECH, Graz, Austria, 15–19 September 2019; pp. 26–30.
39. Qiu, L.; Abdullah, S. Voice assistants for speech therapy. In Proceedings of the Adjunct 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, Virtual Event, 21–26 September 2021; pp. 211–214.
40. Tan, C.T.; Johnston, A.; Ballard, K.; Ferguson, S.; Perera-Schulz, D. sPeAK-MAN: Towards popular gameplay for speech therapy. In Proceedings of the 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death, Melbourne, Australia, 30 September–1 October 2013; pp. 1–4.
41. Liu, N.; Barakova, E.; Han, T. A Novel Gamified Approach for Collecting Speech Data from Young Children with Dysarthria: Feasibility and Positive Engagement Evaluation. In Proceedings of the 2024 17th International Convention on Rehabilitation Engineering and Assistive Technology (i-CREATE), Shanghai, China, 23–26 August 2024; pp. 1–5.
42. Ondáš, S.; Staš, J.; Ševc, R. Speech recognition as a supportive tool in the speech therapy game. In Proceedings of the 2024 34th International Conference Radioelektronika (RADIOELEKTRONIKA), Zilina, Slovakia, 17–18 April 2024; pp. 1–4.
43. Karthan, M.; Hieber, D.; Pryss, R.; Schobel, J. Developing a gamification-based mhealth platform to support orofacial myofunctional therapy for children. In Proceedings of the 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), L'Aquila, Italy, 22–24 June 2023; pp. 169–172.
44. Barletta, V.; Cassano, F.; Pagano, A.; Piccinno, A. A collaborative ai dataset creation for speech therapies. In Proceedings of the CoPDA2022–Sixth International Workshop on Cultures of Participation in the Digital Age: AI for Humans or Humans for AI? Frascati, Italy, 7 June 2022; pp. 81–85.
45. Desolda, G.; Lanzilotti, R.; Piccinno, A.; Rossano, V. A system to support children in speech therapies at home. In Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter, Bolzano, Italy, 27–29 June 2021; pp. 1–5.
46. Elo, C.; Inkinen, M.; Autio, E.; Vihriälä, T.; Virkki, J. The role of games in overcoming the barriers to paediatric speech therapy training. In Proceedings of the International Conference on Games and Learning Alliance, Tampere, Finland, 30 November–2 December 2022; pp. 181–192.
47. Grossinho, A.; Guimaraes, I.; Magalhaes, J.; Cavaco, S. Robust phoneme recognition for a speech therapy environment. In Proceedings of the 2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH), Orlando, FL, USA, 11–13 May 2016; pp. 1–7.
48. Nayar, R. Towards designing speech technology based assistive interfaces for children's speech therapy. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 609–613.
49. Cano, S.; Collazos, C.A.; Aristizábal, L.F.; Gonzalez, C.S.; Moreira, F. Towards a methodology for user experience assessment of serious games with children with cochlear implants. *Telemat. Inform.* **2018**, *35*, 993–1004. [[CrossRef](#)]
50. Abdoulqadir, C.; Loizides, F.; Hoyos, S. Enhancing mobile game accessibility: Guidelines for users with visual and dexterity dual impairments. In Proceedings of the International Conference on Human-Centred Software Engineering, Reykjavik, Iceland, 8–10 July 2024; pp. 255–263.

51. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)] [[PubMed](#)]
52. Duval, J. A mobile game system for improving the speech therapy experience. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vancouver, BC, Canada, 9–14 July 2017; pp. 1–3.
53. Ahmed, B.; Monroe, P.; Hair, A.; Tan, C.T.; Gutierrez-Osuna, R.; Ballard, K.J. Speech-driven mobile games for speech therapy: User experiences and feasibility. *Int. J. Speech-Lang. Pathol.* **2018**, *20*, 644–658. [[CrossRef](#)] [[PubMed](#)]
54. Murero, M.; Vita, S.; Mennitto, A.; D'Ancona, G. Artificial intelligence for severe speech impairment: Innovative approaches to AAC and communication. In Proceedings of the PSYCHOBIT, Naples, Italy, 4–5 October 2020.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.