ArticuMotion: Towards Assessing Motor Speech Disorders via Gamification

Ghada Alsebayel
Northeastern University
Boston, Massachusetts, USA
alsebayel.g@northeastern.edu

Mahsa Nasri
Northeastern University
Boston, Massachusetts, USA
nasri.m@northeastern.edu

Caleb Myers
Northeastern University
Boston, Massachusetts, USA
myers.ca@northeastern.edu

Giovanni Troiano
Northeastern University
Boston, Massachusetts, USA
g.troiano@northeastern.edu

Elaheh Hatamimajoumerd
Northeastern University
Boston, Massachusetts, USA
e.hatamimajoumerd@northeastern.edu

Sarah Ostadabbas
Northeastern University
Boston, Massachusetts, USA
s.ostadabbas@northeastern.edu

Kristen Allison
Northeastern University
Boston, Massachusetts, USA
k.allison@northeastern.edu

Casper Harteveld
Northeastern University
Boston, Massachusetts, USA
c.harteveld@northeastern.edu

Figure 1: Two scenes from *ArticuMotion* featuring the Space Treasure Hunt (left) and Zoo Exploration Adventures (right).

ABSTRACT

Assessing speech disorders in early childhood is challenging, and Speech-Language Pathologists (SLPs) play a key role in addressing such a challenge. However, tools that support speech assessment are often not child-friendly, and SLPs struggle to keep young patients engaged. To compensate, we introduce *ArticuMotion*, a child-friendly app that supports the assessment of speech disorders while engaging children in gamified experiences. We use participatory design to co-create *ArticuMotion* with SLPs and test the resulting product in a user study with nine preschool children. *ArticuMotion* has promise as a gamified assessment for motor speech disorders and shows a potential avenue for designing clinical tools that are useful while being child-friendly.

IDC '24, June 17–20, 2024, Delft, Netherlands

ACM ISBN 979-8-4007-0442-0/24/06
https://doi.org/10.1145/3628516.3655815

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Consumer health**; *Computer games.*

KEYWORDS

Speech assessment, Children, Games for health, Gamification

ACM Reference Format:
Ghada Alsebayel, Mahsa Nasri, Caleb Myers, Giovanni Troiano, Elaheh Hatamimajoumerd, Sarah Ostadabbas, Kristen Allison, and Casper Harteveld. 2024. *ArticuMotion*: Towards Assessing Motor Speech Disorders via Gamification. In *Interaction Design and Children (IDC '24), June 17–20, 2024, Delft, Netherlands.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3628516.3655815

1 INTRODUCTION

Speech Disorders is an umbrella term for many conditions that affect the perception, production, articulation, or fluency of speech [2]. Studies estimate that speech disorders are among the most common disabilities in children in the U.S. and worldwide [50, 61]. According to the National Institute on Deafness and Other Communication Disorders (NIDCD) [56], nearly 1 in 12 (7.7%) U.S. children aged 3-17 have had a disorder related to voice, speech, language, or swallowing in the past 12 months. The prevalence of these disorders is highest among children aged 3-6 (11.0%). Further, the accuracy of screening tools for identifying speech and language disorders varies widely. The United States Preventive Services Task Force (USPSTF) [44] found inadequate evidence on the accuracy of screening instruments for speech disorders in primary care settings, and speech-language pathologists (SLPs), who specialize in evaluating and treating speech disorders, often find it hard to assess and diagnose these disorders accurately.

One reason why accurate assessment and diagnosis of speech disorders is challenging lies in the cognitive, linguistic, and motor functions required to master clear articulation of spoken language, as their complex interplay makes it difficult to isolate the etiology of speech disorders [63]. For example, motor speech disorders (i.e., dysarthria and childhood apraxia of speech) have a neurological basis [22], while other speech disorders may have unknown etiologies. Furthermore, speech disorders have varied presentations—children with speech disorders form a heterogeneous group, among which symptoms widely vary [76]. Currently, SLPs mainly use auditory-perceptual methods to assess and diagnose speech disorders [57], i.e., listening and observing the speech of an individual to identify patterns, errors, and characteristics that may indicate a speech disorder [10, 22]. However, as speech errors can overlap between different types of speech disorders, relying on auditory-perceptual judgment only makes accurate assessments and diagnosis challenging [19, 43], particularly in young children still acquiring speech skills, who may exhibit speech errors due to immaturity [19]. Besides, practical barriers arise due to restricted access to healthcare services [31, 52, 67] and increased caseload on clinicians who can effectively diagnose and treat these conditions [51].

Technology-mediated solutions can help support SLPs in their endeavors, improving the assessment, diagnosis, and treatment of speech disorders. Nowadays, SLPs can use computerized methods, such as acoustic analysis, to support speech assessment [65]. SLPs can also deliver remote therapy sessions using conferencing software [24, 28, 35], and integrate digital therapy programs to supplement in-clinic sessions with on-demand and at-home care [13, 29]. Among technology-mediated solutions for attending to speech disorders in children, games and gamification are emerging as promising avenues to deliver speech therapies that are both effective and engaging [25]. In that respect, Tommy et al. [73] highlighted the importance of considering fun and engagement when designing for children with speech disorders, showing how engaging game design boosts learning, performance, and retention. Furthermore, games can be made available on mobile devices for enhanced accessibility and ubiquity, enabling speech data to be collected and monitored to deliver tailored speech therapies.

These advantages have motivated the development of many speech therapy games, including *SpokeIt* [23], *Apraxia World* [32–34], *Tabby Talks* [66], and *Littlebee*[1]. These games focus on facilitating speech practice as part of treatment and rehabilitation of speech disorders [4], while few support SLPs' assessments and diagnosis [23, 66], in which assessments are mainly based on the acoustic properties of speech. While viable, this method does not fit all speech disorders. For instance, dysarthria is a motor speech disorder that results from neurological damage affecting the muscles used in speech, causing difficulty in articulating words [22]. Observing and measuring facial features in children, including tongue movement, lip contraction, and other abnormalities in facial muscle strength, coordination, or symmetry, are key to identifying dysarthria and other motor speech disorders [1].

Here, we extend efforts around designing gamified tools supporting SLPs in assessing and diagnosing motor speech disorders by introducing *ArticuMotion*—a mobile app/game that invites children to perform speech tasks in a gamified setting. Our gamified solution is low-cost, non-invasive, and child-friendly and has promise as a diagnostic tool that SLPs can leverage to identify signs of dysarthria in young children by extracting features of facial movements (also called facial kinematics) through audio-visual data. Notably, our approach provides engaging experiences to young patients through gamification strategies.

Through participatory design (PD) [68], we engaged SLPs and computer vision engineers (CVEs) to co-design *ArticuMotion*. Our scope was to design a gamified app that engages children in clinical assessments of dysarthria while capturing videos of their facial features as they execute speech tasks. Our gamified solution (1) prioritizes engaging children in speech assessment tasks through interactive narratives and cartoon-style aesthetics while (2) embedding mechanisms for optimizing the capturing of visual data that potentially reveals signs of speech disorders based on facial landmarks (e.g., lips, chin, etc.). This entailed finding optimal strategies to reduce children's head movement while keeping the user interaction engaging. We then evaluated *ArticuMotion* in a user study with nine preschool children, where we assessed the UX and usability of *ArticuMotion*.

With our exploratory work, we contribute the following to HCI research that focuses on children's health: (1) an artifact called *ArticuMotion*, which leverages gamification for speech disorders assessments; (2) design recommendations and desiderata from SLPs for designing gamified assessments of speech disorders, including technical considerations for accommodating CVEs' needs while providing engaging, interactive experiences; (3) lessons learned and design considerations for assessing speech disorders with preschoolers via gamification. Our work provides further perspective on how gamified tools can meaningfully integrate with and extend clinical practices to foster children's well-being.

2 BACKGROUND AND MOTIVATION

Speech disorder encompasses many conditions that vary in severity, underlying causes, speech error characteristics, and response to treatment [76]. While recognizing that speech disorders are spread across and affect multiple age ranges, from childhood to adulthood, accurate diagnosis of speech disorders in young children is particularly important because (1) early childhood is a critical period for speech development and for maximizing the effectiveness of speech therapy [19], and (2) children with motor speech disorders need different types of intervention from children with linguistically-based speech disorders [22]. Specifically, children with motor speech disorders require intervention approaches that are high in intensity and target their underlying movement

[1]https://littlebeespeech.com/

deficits [55, 59], whereas phonologically-based approaches are more effective for children with non-motor speech disorders (e.g., [5]). Therefore, early and targeted treatment is critical for maximizing long-term speech outcomes. Left untreated, speech disorders can increase the risk of reading difficulties and poor academic performance, leading to low self-esteem, social exclusion, and jeopardizing children's well-being [69, 70].

In clinical practice, SLPs use a combination of formal and informal assessments to determine speech diagnoses. Standardized assessments of speech sound acquisition are commonly used and provide quantitative information about children's accuracy in producing speech sounds compared to age-based norms (e.g., The Goldman-Fristoe Test of Articulation-3 [62]). Scoring these tests requires SLPs to judge the accuracy of individual consonant and vowel sounds produced by the child. In addition, SLPs observe children's global speech symptoms across tasks of varying complexity, including single words, sentences, and conversational speech. These speech symptoms involve characteristics such as the child's speaking rate, intonation, nasality, voice quality, and overall articulatory precision. Currently, SLPs rely on auditory-perceptual judgment, where they listen and observe speech symptoms to determine a child's type of speech sound disorder and diagnosis [10, 22]. While viable, perceptual methods are often time-consuming, unreliable, and highly dependent on the clinician's experience [6, 36, 78]. This makes obtaining a clear-cut diagnosis difficult, especially because symptoms may overlap between speech disorders [19, 41]. Thus, objective tools are needed to improve the reliable and accurate diagnosis of speech disorders in children.

In response to the need for objective tools to aid speech diagnosis, computerized methods, such as acoustic analysis, have been developed to support SLPs [65]. Nowadays, acoustic analysis is accessible in clinical practices due to the availability of hardware and software (e.g., Praat[2]—a software package for speech analysis). These forms of acoustic analysis provide SLPs with valuable sources of information to infer precisely the characteristics of disordered speech [46, 48]. While acoustic analysis has shown useful for speech disorders broadly [48, 54], recent research used facial kinematics (i.e., the dynamic changes in facial features, such as the movement of the lips, jaw, and other facial muscles) as effective indicators of disordered speech in patients with dysarthria [1, 3]. However, pragmatic limitations, such as constraints on time and access to specialized equipment, have limited the clinical translation of this novel approach. Optical motion capture systems are highly accurate for face tracking, but those are expensive, time-consuming, and involve the placement of physical markers on the patient's face, which can be uncomfortable, especially for young children. Recent advances in computer vision may provide a suitable alternative that is accessible, contactless, and durable. Nonetheless, computer vision algorithms are known to be *"data hungry,"* i.e., they rely heavily on the availability of large volumes of data to derive clinically useful insights. Thus, many speech samples need to be collected and analyzed, resulting in lengthy sessions, which can be perceived as tiresome or boring. To that end, interactive systems can be integrated into speech assessment sessions, making these sessions more engaging and enjoyable for patients, especially children, an approach that has shown success in the context of speech practice (e.g., [23, 34]). In the following section, we survey prior work on interactive systems for speech therapy, explaining how such systems have been integrated into the practices of SLPs and what is the state-of-the-art. In that regard, we explain how our work, which aligns with gamified speech assessments, extends ongoing efforts that leverage interactive systems and technology-mediated solutions for assessing, diagnosing, and treating speech disorders.

[2]https://www.fon.hum.uva.nl/praat/

3 TECHNOLOGY-MEDIATED SOLUTIONS FOR SPEECH THERAPY

Interactive systems have been adopted to support and complement the workflow of SLPs. For example, telecommunication technologies and videoconferencing software are used to link therapists to their patients [74]. This enables SLPs to deliver consultations, assessments, and interventions remotely, thereby expanding access to therapy and reducing service inequities in rural areas [24, 28]. However, the telepractice of speech therapy introduces its own set of challenges. For instance, SLPs report that auditory-perceptual evaluation—a method already criticized for being subjective and unreliable [57], becomes even more difficult for teletherapy patients due to issues like poor Wi-Fi, poor auditory quality, and difficulty hearing/seeing the patient [30]. One approach to augment auditory-perceptual evaluation is to use computer-assisted speech analysis. For example, SLPs can use acoustic analysis software (e.g., *Praat*, *Voice Evalu8*, *Voice Analyzer*, and *Sonnetta*) on patient recordings [30] to calculate measures like pitch, jitter, breathing pauses, and voice breaks, which support SLPs with their diagnostic decision-making [54]. Such systems are primarily designed for SLPs and domain experts. Technology-mediated solutions have also been designed for patients, such as Computer-Based Speech Therapy (CBST) tools [13, 29]. According to Furlong et al. [29]:

> *"A CBST is a software offering predefined therapy tasks inclusive of instructional features (e.g., an animated talking tutor, the use of synthesized speech to provide models or instructions), motivational features (e.g., the use of animations, game-based activities) and quantitative features (e.g., the tracking of performance within and across therapy sessions), operating from a personal computer."* (p. 51)

While Furlong et al. [29] limit their definition of CBSTs to personal computers, advancements in mobile computing have made similar implementations possible on tablets and mobile phones. Chen et al. [13] proposes a broader definition of CBSTs that rather focuses on the target population:

> *"CBSTs are interactive computer programs that target a specific speech deficit based on a predefined therapy program, but not a program designed to facilitate improvement in speech or language skills in learning a non-native language, in the absence of a diagnosed speech or language deficit, or to tutor individuals whose abilities are already within the normal range."* (p. 100).

Table 1 shows existing CBST systems proposed in the literature that support speech therapy for children (see Attwell et al. [4] and Chen et al. [13] for a review).

Table 1: Systems designed to support speech therapy for children and their main characteristics.

Ref.	System	Purpose	Features	Population
[7]	Tingog	Training	Exemplifies correct speech with videos and text-to-speech, and uses speech recognition for feedback.	Children with repaired cleft palate aged 3 to 8 years old
[23]	SpokeIt	Training	Exemplifies correct speech with lip animation and uses speech recognition for feedback.	Children born with orofacial cleft
[32–34]	Apraxia World	Training	Integrates speech exercises into a platformer-style game, and provides feedback based on either automatic pronunciation evaluation or input from a human evaluator	Children with speech sound disorders
[58, 66]	Tabby Talks	Training	Embeds speech exercises into a "memory/concentration" game and uses speech recognition to assess speech errors and generate reports for SLPs.	Children with apraxia of speech
[21]	Lingo Dingo	Screening	A playful interface that prompts children to say specific words are recorded and passed to deep neural networks for classification.	Children ages 4 to 5 years old

Most CBSTs, as observed by Attwell et al. [4], have traditionally focused on supporting speech therapy for rehabilitation. These systems are designed to facilitate speech training, both in clinical settings and at home, which is a priority because, as demonstrated by Maas et al. [49], the effectiveness of speech rehabilitation requires intensive practice and a homework component where patients continue to engage with exercises at home. Thus, many CBSTs build on the literature around gamification [17, 37], defined as *"using elements of game design in non-game contexts, products, and services to motivate desired behaviors."* ([15], p.1). Leveraging the motivational affordances of games, many gamified applications are proposed in the literature to engage children in speech practice (e.g., [7, 23, 34, 66]). Few studies proposed augmenting their gamification strategies with speech recognition technology to provide feedback on users' speech [23, 34, 66]. For instance, *SpokeIt* [23] is a speech therapy game that exemplifies correct speech with lip animation effects (using Adobe Character Animator), then provides feedback on users' speech (using the speech recognition engine Pocketsphinx and RapidEars) to help users improve their speech production.

Our work extends prior CBSTs in two ways: (1) where prior work has mostly focused on using games and gamification to engage children in speech practice, we are concerned with augmenting and supporting SLPs during speech assessment when working with children, and (2) while prior work on CBSTs proposes auto-identification of speech errors based on acoustic properties of the speech signal [12, 34, 66], we incorporate speech signal with visual data that can potentially reveal signs of speech disorders based on facial landmark movements.

Beyond interactive design for children, a few studies proposed facial tracking technology [1, 26, 27], where measures of the movement of facial features, such as lip corners and the vertical movement of the upper and lower lips and the jaw, provided substantial improvement compared to audio-only modeling of speech articulation. In other words, visual cues have shown the potential to provide an additional channel by which SLPs can better understand, evaluate, and treat disordered speech [1]. However, this approach has yet to be seamlessly incorporated into interactive systems for speech assessment, given that the acquisition of facial movement data requires the use of motion capture systems, which can be expensive and inaccessible. We explore an alternative solution by co-designing an interactive system for pediatric speech assessment based on computer vision algorithms and 2D videos embedded in a gamified experience—*ArticuMotion*. Given the exploratory nature of our work, we position *ArticuMotion* as a gamified research tool [16] and acknowledge that further work is needed to establish its clinical utility in assessment and diagnostics.

4 ARTICUMOTION DESIGN

Our design process took approximately eight months between January and August 2023. Following Participatory Design (PD) approaches [68], we co-designed *ArticuMotion* with speech-language pathologists (SLPs) and computer vision engineers (CVEs). Through PD, we elicit expertise and insights from these stakeholders to inform a user-centered design of *ArticuMotion* that considers the needs of healthcare providers (SLPs), engineers that support them (CVEs), and their patients (children with speech disorders). The steps we followed in our design process were inspired by the Design Thinking Model [9] (see Figure 2), starting with an observation study, where two researchers with a background in HCI and Game Design (GD) analyzed six video recordings of SLPs administering speech assessment to preschool children. This allowed us to empathize and identify common themes and challenges in the context of pediatric speech assessment. Next, we organized PD sessions with SLPs, CVEs, GDs, where we used methods such as divergent brainstorming [45] and prototyping [68] to co-create two game concepts, which we iteratively developed and refined into the polished *ArticuMotion* prototype.

We evaluated *ArticuMotion* with nine preschool children for UX and usability and used their insights to inform our iterative process. We elected to involve children in the evaluation of *ArticuMotion* rather than the generation of our design because, as prior work suggests [39], existing co-design techniques often work best with children older than our target age group. More importantly, our design problem is highly specialized; thus, to ensure alignment of our design with domain-specific requirements (e.g., designing around established speech assessment tasks), we involved domain experts in co-creating our design solution and used input from children to refine it. Next, we provide detailed descriptions of our co-design activities. Section 5 introduces *ArticuMotion*, and the user study is discussed in Section 6.

4.1 *Empathize:* Observation Study

In the initial stage of our design process, we observed six recorded videos of preschool children (ages 3 to 6) interacting with SLPs to understand better the context of pediatric speech assessment and the practices of SLPs. The video recordings featured children performing speech tasks such as the Arizona 4 test—a standardized measure of articulation and phonology [71]; the TACL 4—a standardized measure of auditory comprehension [11]; and the Diadochokinesis (DDK) task—a common research and clinical tool used to test oromotor skills by measuring how quickly a patient can accurately repeat a series of rapid, alternating sounds, e.g., "puh-puh-puh" [47]. Our supplementary material has further details and

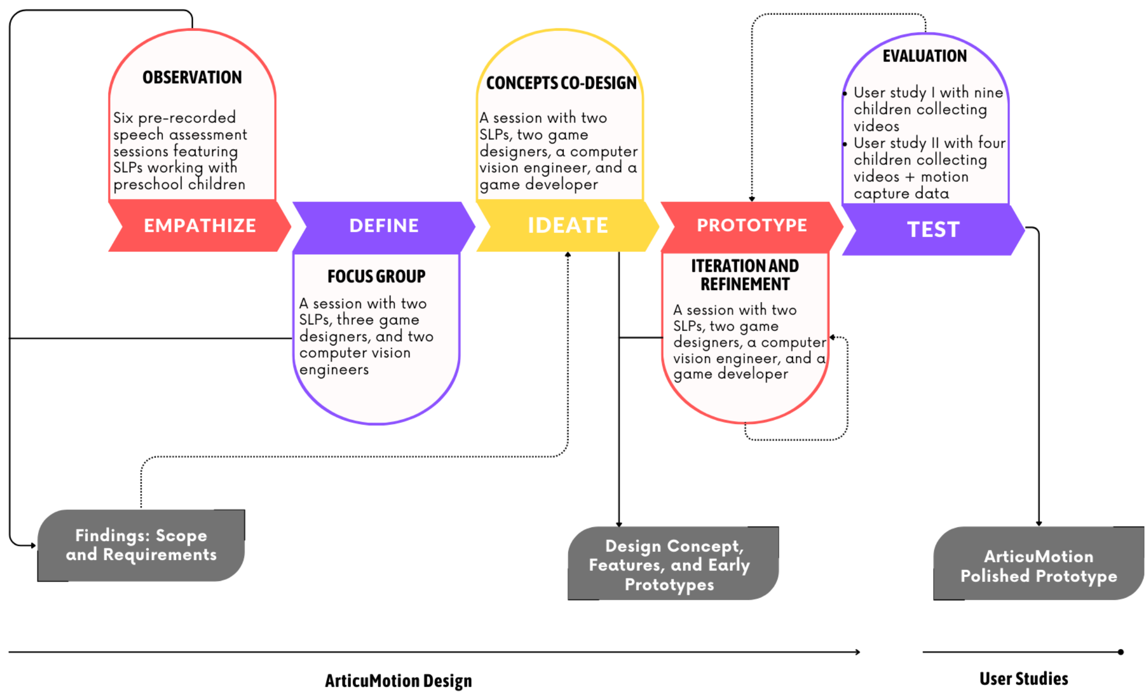


Figure 2: A diagram describing the design and evaluation cycle for *ArticuMotion*

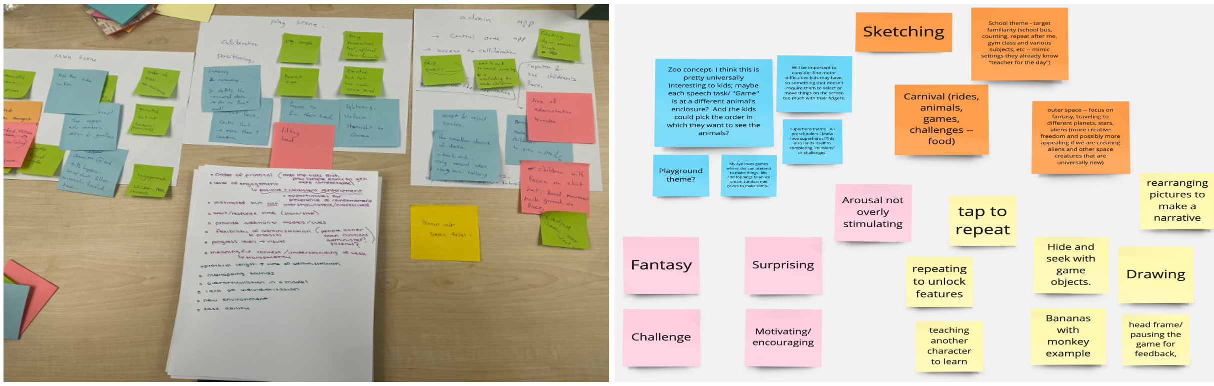
descriptions of all the tasks included in the video recordings. Each session lasted approximately two hours, and participants were classified by SLPs as (1) two typically developing, (2) two with suspected speech disorders, and (3) two with dysarthria.

To analyze the data, the two authors with a background in HCI and GD observed the video recordings and took notes. The two authors used Microsoft Excel to code video recordings and observation notes. Following a flexible inductive coding approach [8], we coded instances of interaction between SLPs and children, practices and approaches employed by SLPs, challenges, needs, and pain points encountered by both SLPs and children during the speech assessment tasks. For instance, if we observed the SLP instruct the child to repeat the task, we would label this as "Need for Repetition." Similarly, if an SLP had to demonstrate the task for a child to elicit their response, we would label this as "Leading by Example." When participants showed signs of disengagement, tiredness, or boredom (e.g., rolling eyes, asking, "Is this the last one?", "Are we done?"), we would label this as "Lack of Engagement." All codes were organized into themes that captured the challenges, needs, and practices of SLPs administering speech assessments to preschool children. To validate our findings, we organized a focus group discussion with our stakeholders, where we discussed and refined the themes that emerged from our observation study. We aimed to confirm that our interpretations were grounded in the stakeholders' perspectives and uncover additional challenges, needs, and requirements they might have.

4.2 *Define:* Focus Group

We invited two SLPs and two CVEs, with two authors as HCI/GD researchers (detailed in Table 2), to a focus group discussion for approximately two hours. The session followed a hybrid format, where participants attended in person or virtually using Microsoft Teams. We paired the participants based on their backgrounds and gave them time to think and reflect on the question before we engaged in group discussion. During the session, we validated our findings from the observation study and explored any additional challenges, needs, or requirements that SLPs and CVEs might have to support pediatric speech assessment based on video data. As the facilitator, one author started the sessions with two guiding questions for SLPs: (1) How would you describe your best experience administering speech assessment for preschool children? (2) How would you describe your worst experience administering speech assessment for preschool children? Similarly, we invited CVEs to share their experiences working with 2D videos and facial tracking algorithms and asked: (1) How would you describe the ideal data for facial tracking algorithms in terms of its attributes? (2) How would you describe your worst experience working with data for facial tracking algorithms? Furthermore, SLPs were prompted to reflect on their practices and experiences working with children, express expectations and ideal scenarios, and identify specific goals, desired features, and effective strategies.

Similar inquiries were made to CVEs regarding data and algorithmic limitations. HCI/GD researchers were encouraged to share ideas and strategies for designing engaging interactive



(a) Post-it Notes from the Focus Group Session **(b) Online Miro Board from the Concept Co-design Session**

Figure 3: Pictures from Our Participatory Design Sessions

experiences. As advised by Peterson and Barron [60], we used post-it notes to generate information, share a common understanding, and sort ideas (see Figure 3a). We used physical post-its for in-person attendees and *Miro* [3] for virtual attendees. In addition, after acquiring participants' verbal consent, we recorded the session for transcription and further analysis. Similar to our approach described in Section 4.1, we used thematic analysis, particularly inductive open coding [8], to analyze our data such that two authors with a background in HCI and GD codded instances where our stakeholders described expectations of the system, attributes of ideal data and scenarios, criteria for successful completion of assessment tasks, and pain points encountered by both SLPs and CVEs in their workflow. For example, when SLPs mention a challenge they often encounter, such as "when [children] don't understand the task," we would label this as "Need for a Meaningful Context." Similarly, when CVEs state, "we couldn't capture the facial landmarks" because of the child's movement, we would label this as "Need to Maintain Front-facing view." Furthermore, descriptions like "not silly" and "kind of getting to the right level of engagement" would be labeled as "Characteristics of Optimal Experience." To ensure that our themes are grounded in the expertise of our stakeholders, we used expert validation, where we shared our themes with the stakeholders for input and further refinement, which resulted in three main themes.

Table 2: Description and Details of Co-design Participants.

ID	Background	Experience Level	Gender
S1	Expert in pediatric motor speech disorders	5+ years	Female
S2	Graduate student in speech-language pathology	Up to 3 years	Female
V1	Expert in computer vision engineering	5+ years	Female
V2	Postdoctoral researcher in computer vision engineering	5+ years	Female
G1	Graduate student in computer science	3 to 5 years	Female
G2	Graduate student in interaction design	3 to 5 years	Female
G3	Expert in game design	5+ years	Male

[3]https://miro.com/

4.3 Findings

We identified three themes emerging from the observation study and focus group: (1) *Practices and approaches of SLPs administering pediatric speech assessment*, which discusses experiences working with children, specifically what works and what does not work when administering speech assessment to preschoolers; (2) *Functional requirements for technology-mediated pediatric speech assessments*, which describes SLPs' and CVEs' functional requirements for a pediatric speech assessment solution based on 2D videos; and (3) *Crafting the user experience*, which describes the rather intangible aspects of interactive design where gratification, stimulation, and meaningfulness of the experience are discussed.

4.3.1 Practices and Approaches of SLPs Administering Pediatric Speech Assessment. S2 emphasized the importance of establishing rapport with children before attempting to administer speech assessment: *"We'll often do something, um, low pressure to start to kind of help them [children] get comfortable. So something like play sample like we play with the play basket, something like that."* This helps foster a sense of trust and connection and creates a positive environment, reducing anxiety and laying the ground for an effective session. In addition, S2 stressed the importance of offering *"positive and consistent reinforcement,"* to encourage children to complete the assessment tasks. Additionally, offering rewards is another technique that SLPs use to motivate children to perform the tasks. In particular, S2 mentioned that stickers often worked as a reward mechanism, where children are given stickers when they complete a task. An interesting remark by S2 was that providing children with a "personalized" preference is often effective: *"We usually use sticker charts and then within that an opportunity for preference of reinforcement. So, having different things that'll be motivating based on what the kid likes. And it's not even a huge range that we have. We have a few different sticker boards. It's like trains and castles. Yeah, just a few differences."*

In addition to (tangible) rewards, S2 mentioned embedding the task into a *"a game, a competition"* of who can perform the task better: *"I bet you cannot do it as many as I can"*, and then use over-exaggeration as an intangible reward *"Ohh my gosh, you could do more than me. Things like that. Kind of boost* *their ego a bit."* Furthermore, SLPs highlighted the importance of progress indicators, where children are given feedback on how many more tasks or attempts are needed to complete the assessment successfully. For example, S2 mentioned that she would count on her fingers to represent the remaining tasks to children visually. Furthermore, S2 mentioned that children can be shy or uncomfortable during the assessment; in such cases, SLPs would ask a parent to step in and participate in the session, which they found would boost the child's engagement with the tasks, e.g., *S2: "sometimes we tag team it, so a parent would come in and help with the task [...] when the kid is really shy, and they don't want to do it, then it's like 'let's hear mommy do it first."* In other cases, bringing something personal and familiar in the session, e.g., *S2: "sometimes they [children] bring their favorite toy"* would help encourage children to engage with SLPs.

Lastly, given that children have a range of speaking abilities, various scaffolding techniques are used to elicit the desired response. These techniques include (1) *direct instruction*, i.e., explicitly asking children to pronounce letters, words, or phrases (e.g., *"Can you say Buy Bobby a Puppy?"*); (2) *modeling*, where SLPs demonstrate the desired response, (e.g., *"Watch me do it"*); (3) *collaboration*, where SLPs assist children in completing the task (e.g., *"Let's do it together"*); (4) *competition*, where SLPs turn the task into a competition (e.g., *"I bet you can not do it as many as I can"*); and (5) *guided-instruction*, where SLPs break down the task into sub-tasks and guide children through each step (e.g., *"Can you say, Buy. Bobby. A Puppy?"*) Below, we provide the retrieved list of SLPs' successful pediatric speech assessment practices:

- Establish a rapport using a low-pressure activity, e.g., a play segment.
- Include a familiar person, like a parent, in the session to boost the participation of shy children.
- Use reinforcement and rewards to motivate children to complete the tasks, e.g., stickers.
- Offer personalized rewards, e.g., choosing from a range of different themes of stickers.
- In addition to tangible rewards, intangible rewards such as recognition, praise, and over-exaggeration can boost ego and provide encouragement.
- Provide feedback and progress indicators, e.g., counting how many repetitions/tasks are needed to achieve reward.
- Use various scaffolding techniques such as direct instruction, modeling, collaboration, competition, and guided instruction.

4.3.2 Functional Requirements for Technology-Mediated Pediatric Speech Assessment. Our stakeholders specified that a good assessment system needs to be portable. Thus, hand-held devices such as iPads are preferred over desktops. This is also consistent with prior work, which shows that the Apple iPad is the most common hardware tool that SLPs use for integrating apps in speech therapy [20]. Thus, we decided to build an iOS app. Based on the needs and challenges described by our stakeholders, we established two key goals for the system: (1) to facilitate engaging children in speech assessment tasks and (2) to capture accurate data (2D videos) of children as they perform these tasks. Based on these two objectives, SLPs and CVEs provided functional requirements to guide the development of such solutions. For example, SLPs discussed features that would facilitate the administration of speech assessment from a pragmatic perspective, such as flexibility of administration, which entails allowing SLPs to pause, repeat, and skip tasks as needed. SLPs explained that children have different speaking abilities; for instance, children with speech disorders might need more time to formulate a response compared to typically developing children. Thus, the system must allow SLPs control and flexibility over attempts and repetitions, order of tasks, and time designated for each task. Furthermore, the system must allow SLPs to skip tasks if needed. Because SLPs use different techniques for scaffolding, such as instruction and collaboration, the system must include a mechanism for turn-taking to prevent an overlap between the sounds of SLPs and children, for example, when SLPs model a task or give instruction. In addition, SLPs suggested that a "wait-time" feature is needed in scenarios where SLPs need to step back and give teaching cues and scaffolding between attempts.

From the perspective of engineers, requirements for the accuracy of captured data were discussed. CVEs indicated that the resolution of the videos is an important indicator of the quality of data (e.g., 100 frames/sec). CVEs also mentioned that a big challenge they often face when working with facial detection algorithms is movements. Children often move, turning their faces away from the camera or putting their hands in front of their mouths, which blocks facial landmarks such as lips and chin. These movements can lower the accuracy of captured data when working with facial tracking algorithms. As such, CVEs explained that obtaining (and maintaining) an optimal position where children sit still and have their facial landmarks visible in each frame is highly challenging. Hence, the system must have a mechanism for constant assessment and correction of position. Additionally, CVEs mentioned that preprocessing and video segmentation are time-consuming activities that cause a high computational burden. Accordingly, one of the requirements from the engineering perspective is to have videos of each speech assessment task recorded separately instead of recording a full session. Furthermore, CVEs sought a feature to keep frames of interest only (i.e., a portion of the video recording where children are performing the task) and discard those that are irrelevant to the task (e.g., a portion of the video recording where SLPs are giving instruction).

Lastly, both CVEs and SLPs agreed that the model behind the pediatric speech assessment tool we are designing should be descriptive (i.e., provides measurements such as speed of movement or range of movement) as opposed to a classifying model (i.e., groups the children into categories of typically developing versus children with motor-speech disorders) in the words of S1:

> *"From a clinical perspective, nobody's going to be like, 'I did this app and the kid has dysarthria,' you know what I mean? Even without enough data to build the classifying model, just being able to get data about the individual child is meaningful. So, you know, outputting range of movement, speed of movement, and then some of like the acoustic parameters and things like that, that gives us a profile of that individual child."*

This conforms with the technical constraints emphasized by CVEs, which underscore the scarcity of data that accurately represents the diverse speech patterns observed in children across different developmental stages, which hinders the development of a classification model.

4.3.3 Crafting the User Experience. While the priority for a speech assessment solution is to obtain accurate data, our analysis revealed that working with children necessitates careful consideration of the context in which speech assessment tasks are embedded; as a conclusion from GD1: *"I think the worst experience would be when the [speech assessment] task doesn't make sense for the participant like they don't understand."* Hence, SLPs were enthusiastic about delivering speech assessment via an interactive medium, as explained by S1:

> *"I think that this is something that is actually potentially a big advantage of doing this [speech assessment tasks - Ed.] within an app because usually the way we do this as part of an assessment is like 'okay, here is the next thing we are gonna do. Now we are gonna say this sentence,' or 'now I wanna see how long you can say an 'aa' sound.' Some of these tasks may seem sort of unusual or not something that a child is typically used to doing; sometimes, they are put off by being asked to do something like that."*

Accordingly, SLPs emphasized the significance of embedding speech assessment in a meaningful context, which may help facilitate the administration of speech assessment to children. S2 described an optimal context as one that keeps children *"motivated but not overstimulated or overexcited because that might not be the most naturalistic speech."* On the other extreme, if the assessment feels boring to children, according to S1, they *"might be sort of mumbling or, if they're nervous, they're really quiet, and so we [SLPs] can tell we're not getting their best speech because they're talking quietly or mumbling or whatever."* As such, we concluded by identifying the target gratifications for our game experience as reasonably challenging and not too overly stimulating. Based on these insights, we aimed to build a gamified app that delivers a captivating narrative with intuitive rules and game objectives that are not too hard to achieve [72]. We document requirements and game design features, showcasing the perspectives of SLPs and engineers in Table 3.

Table 3: Stakeholder Requirements and Game Design Features.

ID	Stakeholder	Requirement	Game Design Feature
S1	SLP	Flexibility for different abilities	Initial time settings
S2	SLP	Flexibility of administration	Control panel to pause, redo, skip tasks
S2	SLP	Establish a rapport with a low-pressure activity	Character selection screen
S2	SLP	Give positive and consistent reinforcement	Affirmation and encouragement in narration
S1	SLP	Provide progress indicators	Energy bar corresponding to task progress
V1	CVE	Task completion indicators	Reward for progress through the narrative
S2	SLP	Give positive reinforcement	Rewards and badges
V2	CVE	Consistent assessment and correction of head position	Star mechanism for feedback on head position
V2	CVE	Video segmentation	Recording each task separately

4.4 *Ideate:* Concept Co-design

Participants in the second design session were two SLPs, one CVE, and three HCI researchers/GDs. One author acted as a facilitator and presented similar applications in the context of speech therapy for inspiration. Next, the facilitator initiated a brainstorming exercise, where participants brainstormed themes for our game. Figure 3b shows themes emerging from our discussion. We used the Mechanics, Dynamics, and Aesthetics (MDA) framework to guide our concept design [40]. Similar to the prior session, we obtained video recordings and transcripts for further analysis. Concepts co-design was done iteratively, where we co-created two game concepts, one featuring a space adventure and the other a zoo exploration journey. We chose these two themes because SLPs described them as universal and highlighted how children are familiar with both concepts: *"The zoo idea is great because I think that's something that even pretty young kids are familiar with, or even if they haven't been to a zoo for some reason, they're familiar with animals."* As advised by S1, we developed our concepts around two speech assessment tasks. The first task is the DDK [47], which is a maximum performance task, where children are asked to repeat the syllable "Pa" as many times as they can in one breath. The second task is to repeat the phrase "Buy Bobby a Puppy" five times [42]. Figure 4 shows early design concepts. After the design sessions, we continued engaging with our stakeholders for feedback on *ArticuMotion* throughout its development.

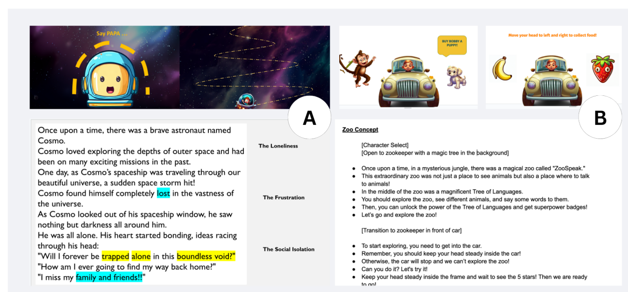


Figure 4: Narrative design and early prototypes: (A) Space adventure and (B) Zoo exploration.

4.5 *Prototype:* Iterations and Refinement

We created the early prototypes of *ArticuMotion* with Microsoft PowerPoint, gathering feedback from SLPs and CVEs. Art assets, including character design, environment, and in-game elements like badges, were crafted with text-to-image Microsoft Designer[4] and edited as needed using Adobe Photoshop. Discussions around early prototypes helped refine the game design, mechanics, and setting details. Upon finalizing our game design concept, we initiated the implementation phase, where an interactive prototype of *ArticuMotion* was developed using the Unity game engine. The details are described in Section 5. The initial interactive prototype included basic functionalities of both the space and zoo concepts. The narrative was incorporated as an on-screen script, and the SLP would have to navigate through the dialogue by clicking on the iPad. The space and zoo concepts were also given narrative endings indicated by a bell chime to inform the child that the experience had ended.

[4]https://design.microsoft.com

5 ARTICUMOTION

ArticuMotion is a gamified speech assessment tool that incorporates speech assessment tasks in an interactive narrative game. Specifically, two speech assessment tasks (i.e., DDK and "Buy Bubby a Puppy") are contextualized in the storyline. Players progress through the gameplay by completing the speech tasks. In designing *ArticuMotion*, we considered personalization and player agency, where players can select a character from diverse gender and race representation. Players control the character so that they are presented with challenges (i.e., speech tasks), and their actions advance the story. Furthermore, SLPs have a control panel that can be used to skip and repeat tasks, calibrate head position for enhanced facial landmarks visibility, and adjust the time duration per task. Moreover, as advised by SLPs, regardless of the player's performance, the game advances just by completing the tasks.

In the opening scene, SLPs are prompted to select a concept (i.e., zoo or space) and input the time duration for the speech tasks embedded in the game (Figure 5 F). In the following scene, the child is prompted to select a character from six different options (Figure 5 A). Head calibration is done through the metaphor of putting on a helmet (Figure 5 B) in the space concept and riding a safari car in the zoo concept. The core mechanics in *ArticuMotion* are designed to positively encourage children to sit still and keep their heads within the frame of interest to ensure maximum visibility of facial landmarks, such that if they move their heads outside the frame, the game will pause and prompt them to move their heads inside. In the space concept, this is contextualized in the act of keeping a helmet on for the space adventure to continue. Similarly, in the zoo concept, children must stay in a safari car while their avatar, associated with their head, is displayed on the front window, mimicking a driving scene. Moreover, we created a staring mechanism to provide real-time feedback to children about the position of their heads (exemplified in Figure 5 B). These features were intended to ensure proper data collection. Moreover, an energy bar is displayed to estimate the duration of a task (Figure 5 D, G, I).

5.1 Development Tools

ArticuMotion was developed using the Unity game engine with version 2022.3.9f1. This was the latest long-term support version of Unity, released at the start of development, and offered the most features and game engine stability. We utilized Unity's AR Foundation and Apple ARKit[5] integration packages for facial recognition and NatML's VideoKit SDK[6] for data collection through video recording due to the package's ease of use and active developer support during the development of *ArticuMotion*. Additionally, we used the software Xcode[7], which Unity requires to build and deploy apps to iOS devices. One challenge we faced during development was mediating camera usage between face recognition and video recording. During initial development, video recordings from the app became corrupted and were not properly saved during gameplay. The cause of this issue was that both the ARFoundation and VideoKit SDKs were attempting to use the camera simultaneously, and the camera could not feed the two information simultaneously. This was solved by prioritizing the facial recognition feature. The camera feed was outputted to a Unity camera through this package, which we then recorded instead of recording the camera feed directly.

[5] https://developer.apple.com/documentation/arkit/
[6] https://www.natml.ai/videokit
[7] https://developer.apple.com/xcode/

5.2 Data Collection Metrics

We established various metrics and features that would allow us to capture usable data. *ArticuMotion's* gameplay would stop if the children moved their heads outside the frame of interest—front-facing the center of the iPad's camera, and would resume when their heads became realigned. This box for bounding the children's heads was a 600 by 600-pixel box in the center of the iPad's screen. Additionally, the game would pause if the children's heads exceeded a 28.1-degree rotation in any direction from facing the screen. To measure the success of speech tasks, we check to see if the microphone input meets or exceeds -22 decibels (dB). These metrics were initially chosen arbitrarily and then iterated on through pilot testing and feedback from SLPs and CVEs until they reached a point where they were not disruptive to the children's gameplay experience and captured usable data.

5.3 Space Treasure Hunt

This story revolves around a child astronaut discovering a treasure map of the galaxy. The goal of the game is to collect hidden treasures in space. After choosing a character, the game narrator asks the player to fix their head within the spaceship and try to keep it steady to get four or five stars. To explore outer space, the narrator instructs the player to "put on a space helmet" and initiate a spaceship's engine by saying "Papa" as many times as possible in one breath—the DDK task. After successfully filling the energy bar, the player moves to the next scene, where they encounter a friendly alien named Bobby on a distant planet. To receive a green gem, the player repeats the phrase "Buy Bobby a Puppy". Successful completion allows the player to collect the final gem and return to Earth.

5.4 Zoo Exploration Adventure

After selecting one of six characters, a Zoo keeper narrates a magical story about the colorful tree of languages. The player aims to reach the magical tree and gain strength and kindness badges. The zoo keeper guides players to ride a safari car, keep their heads in the car's front window frame, and earn stars for maintaining the correct position. After getting stars (above four), the player starts the car engine by loudly saying the first task, "Papa" as many times as possible in one breath—the DDK task. When the energy bar is full, the vehicle moves, advancing the character to the next level. Next, the player interacts with a monkey named Bobby, who asks the player to repeat task two—repeating the phrase "Buy Bobby a Puppy" five times. After filling this level's energy bar, the player reaches the shimmering tree of language within the forest and earns kindness and strength badges.

6 EVALUATION STUDY

We evaluated *ArticuMotion* in a user study with nine preschool-aged children. We focused on understanding how these nine children experience and interact with *ArticuMotion* and if its design keeps them engaged during speech assessment. Throughout the study,

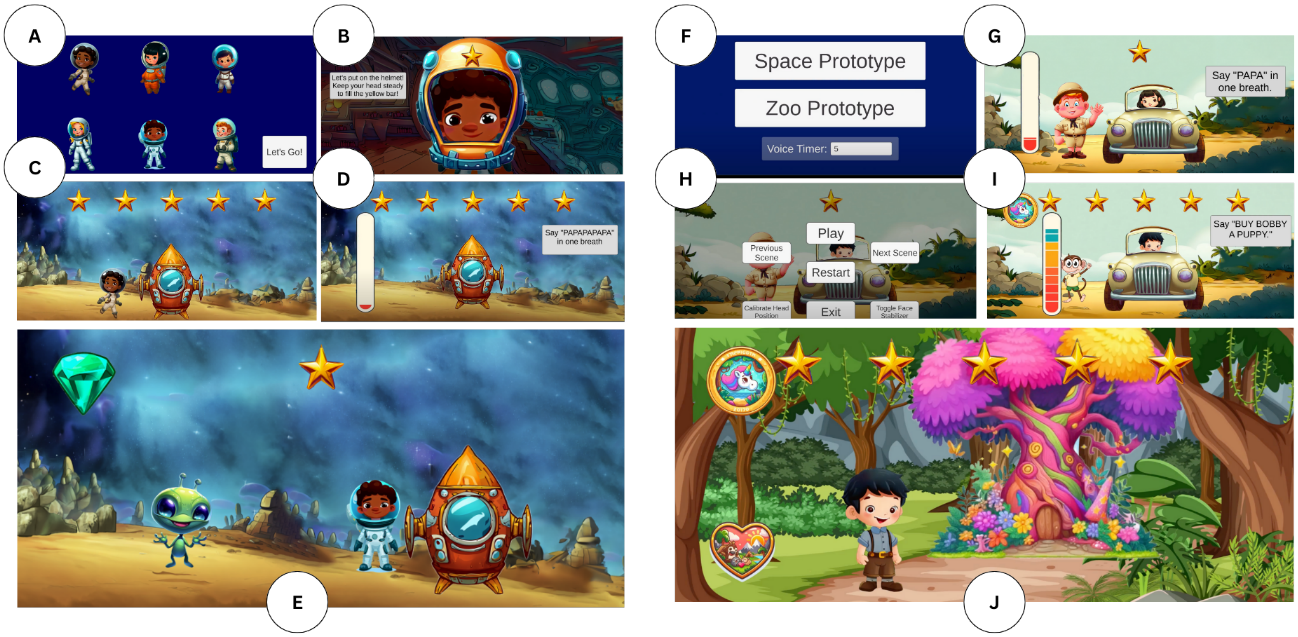


Figure 5: Screenshots from *ArticuMotion*. The main screen (F) enables SLPs to choose between two concepts: *Space* (A, B, C, D, E) or *Zoo* (G, H, I, J). Both themes have the same functionalities. (A) The character selection screen from space, where children choose their in-game avatar; (B) The head calibration screen, leveraging the metaphor of wearing a helmet for a space adventure; when five stars appear on the screen, the helmet is calibrated. (C, D) Speech Task 1 in space and (G) In zoo; in both, children are prompted to say "PAPA/" in one breath to (C, D) Fill up an energy bar to blast off their spaceship or (G) Start a car engine. (I) In speech task 2, children are prompted to say "Buy Bobby a Puppy" five times. Rewards: (E) Win a green gem or (J) Badges for the colorful tree of languages.

we sought to explore the UX and identify potential usability issues of *ArticuMotion* together with SLPs and use the insights emerging from the study to inform its future design iterations.

6.1 Procedure

We contacted the childcare center associated with our university and sought their assistance in recruiting participants. The director was responsible for disseminating flyers with study descriptions and parental consent forms. After receiving written parental consent, we coordinated with the director to set up our study in a room at their location. Figure 6a shows the study setup, where we ran *ArticuMotion* on an iPad Pro 11 installed on a tripod. Children did not actively hold the iPad; instead, they were seated on a chair, front-facing the iPad to view the screen, and SLPs interacted directly with the game (i.e., touching through the screen and using the control panel options).

We had two SLPs alternating and running the sessions such that in each session, an SLP sat beside the participant and walked them through the play-testing protocol, which included (1) obtaining consent, (2) a pre-interview, (3) playing *ArticuMotion*, and (4) a post-interview, as described in Appendix A. Three research team members were present—two of whom observed and took notes, while the third provided technical support as needed. Children were aware of the presence of the research team, as we were introduced to them when they first arrived in the room. However, after the introduction, we maintained a position behind a wall, concealing our presence from the direct view of the children. This decision was made to avoid any tension resulting from having multiple unfamiliar faces around the children. Therefore, only the SLPs engaged directly with the children. In addition to field notes, we collected 2D videos of children as they performed the speech assessment tasks embedded in *ArticuMotion*. These videos were recorded within our game via the iPad camera. Furthermore, we collected videos using an external camera that captured facial expressions and interactions between SLPs and children. All videos were stored in a password-protected server and transcribed for further analysis. We used a flexible coding approach [8] for our analysis, similar to our method in Section 4.1.

6.2 Participants

We recruited nine English-speaking children (six boys and three girls) aged three to five years (M = 3.98, SD = 0.69). Table 4 provides an overview of the participants' information. Play-testing sessions were carried out individually and typically lasted 12 minutes on average. All children played *the Space Treasure Hunt* and *zoo Exploration Adventure* in sequence, where the order was randomized. Figure 6b features one participant playing *the Space Treasure Hunt* game.

Table 4: Participants in the evaluation study. *NA* = did not explicitly say.

Participants ID	Age	Gender	Prior iPad Experience
P1	4	M	YES
P2	3	M	NA
P3	4	F	YES
P4	4	M	NA
P5	4	M	NA
P6	5	F	NA
P7	4	M	YES
P8	4	F	YES
P9	3	M	YES

6.3 Findings

Our findings suggest that *ArticuMotion* was an enjoyable experience for the children. Of the nine children queried about their opinions of whether they enjoyed the game, seven responded affirmatively, stating that they found the game *"fun"* (P8), *"nice"* (P1), and *"kinda cool"* (P7), while the remaining two did not provide an answer. Further, five children expressed interest in playing *ArticuMotion* again. As mentioned, all children played the two themes in *ArticuMotion*, where each theme had two speech tasks, for a total of four tasks per child. The overall task completion rate [64] was 86%. When asked about their preference, six out of nine children said they favored the Space Treasure Hunt over the Zoo Exploration Adventure, one was neutral, and two preferred the Zoo Exploration Adventure. When asked what they liked about the game, most said *"the space alien"*.

6.3.1 Capitalizing on Existing Familiarity and Affinity. *ArticuMotion* appears to leverage pre-existing mental models based on knowledge, understanding, and positive feelings that children already have about the use of certain technology in their daily life, which is consistent with the device, aesthetics, and overall design of *ArticuMotion*. These results are consistent with other UX studies showing how pre-existing mental models impact how individuals may understand and perceive novel technology [75]. For example, the children were familiar with the technology used to display the game and already enjoyed playing games on it—*"I have my own [iPad] at home, and I love to play games on the iPad"* (P1) and *"I like PBS Kids. It has 100 games in it"* (P7). Furthermore, our analysis shows that the themes around which we built our game are already attractive to children, as exemplified by a conversation between SLP1 and P2:

> **SLP1:** *"Do you like animals?"*
> **P2:**...nods yes...
> **SLP1:** *"What's your favorite animal?"*
> **P2:** *"I love green snakes!"*
> **SLP1:** *"Have you seen snakes at the zoo before?"*
> **P2:** *"Yeah!"*

Two children participating in our study showed particular interest in Space Treasure Hunt and generally appreciated its aesthetics and subjects (i.e., the spaceship, the astronaut, the stars):

> **P1:** *"[I like - Ed.] the space alien"*
> **P3:** *"[I like - Ed.] the space game because I want to be an astronaut and a doctor when I grow up"*

Two other children were more attracted by the AI-generated aesthetics of Zoo Exploration Adventure, particularly its colors, the nature depicted in it, and the general cartoony vibe:

> **P2:** *"The tree, I like green trees and purple ones and green ones and blue ones and...and yellow ones!"*
> **P6:** *"I like the tree with all the colors!"*.

6.3.2 Different Engagements. During the study, we observed how *ArticuMotion* would engage children differently based on their preferences of aesthetics and narrative. While this is not inherently surprising, such individual preferences also impacted how children would engage with the underlying assessment of speech disorder embedded in the game's mechanics, including the speech tasks. For instance, children sometimes would be automatically compelled to perform speech tasks without SLPs needing to engage or provide cues, which resulted in effortless data gathering. In fact, P1 attempted to say "pa pa" right after the task was presented on screen by *ArticuMotion* and did not seem to need to follow guidance or instructions from SLPs directly. For other children, the game seemed to provide visual aid, where SLPs would leverage the very aesthetics and animation of the game to guide children in performing the required task:

> **SLP1:** *"Look at this spaceship...it will launch when you say pa pa!"*

Occasionally, SLPs played *ArticuMotion* in front of the children to demonstrate the task and have children engage with the game by prompting them to imitate their actions:

> **SLP2:** *"As fast and as many times as you can on one breath...like this, watch me...[performs the task - Ed.]...and then it is your turn"*

Further, SLPs used *ArticuMotion* as a means to complete the speech tasks with the children collaboratively and generally mixed-and-matched different approaches based on children's responses, showing how a gamified tool like *ArticuMotion* could be differently integrated within speech disorders assessment. These approaches resonate with the SLPs' desiderata emerging in PD and focus groups. These results suggest that *ArticuMotion* may serve as a valuable complementary tool for SLPs to support speech assessments and their more orthodox approaches. Overall, we find that *ArticuMotion* provided an immersive and engaging UX to both children and SLPs and has the potential for seamless integration of assessment tasks, speech data extraction, and patient engagement.

6.3.3 "A Sugar-Coated Pill". In *ArticuMotion*, we prioritized embedding the pediatric and technical requirements of SLPs and CVEs. However, we also sought to engage children in meaningful and enjoyable player/UX experiences. This was achieved by seamlessly integrating the speech assessment tasks provided by SLPs with an interactive narrative and game mechanics that need optimizing and not compromising (1) the visibility of facial features

(a) Study Setup

(b) A Session Featuring the Space Game

Figure 6: Evaluating the usability and user experience of *ArticuMotion.*

while (2) minimizing potential distracting factors and disruptive movements of children as they perform these tasks. In that respect, the results of our user study show that *ArticuMotion* successfully engaged SLPs and children in interactive, technology-mediated experiences that concealed both the clinical and engagement objectives.

Mirroring the concept of a "sugar-coated pill," *ArticuMotion* adeptly disguised the sometimes tedious speech treatment tasks within its captivating game mechanics, transforming essential therapy into an immersive and enjoyable experience for its young users. This immersion was often characterized by children's natural predisposition to enter a "magic circle" of make-believe, in which SLPs sometimes participated to create further connections with their young patients. This is exemplified by how children oftentimes naturally immersed themselves in a role-playing game while engaging with *ArticuMotion.* For instance, P3 pretended to be putting on an "actual helmet" when instructed to sit still and then enacted preparing herself for space exploration in an imaginary cockpit. During the space exploration, P1 said *"It's really windy there, though. At Saturn,"* showing how the game sparked imaginative play and distracted the child from the tediousness of performing repetitive tasks. In a reaction of both surprise and joy, after completing the Space Treasure Hunt, P8 exclaimed: *"I could fly."* The children also appreciated that the game allowed them to choose a fitting avatar, providing a broad range of haircuts, looks, skin color, and gender. P4, for example, remarked how his gender aligning with the gender of his avatar made the player experience believable and consistent: *"I want to be this one because I am a boy!".*

Overall, the environment provided by *ArticuMotion* acted as a believable context where children could intuitively and effortlessly perform the required tasks, while its game design favored identity-building, engagement, collaboration, and SLPs in conducting their speech disorder assessments.

7 DISCUSSION

This effort aimed to build a tool to support SLPs in assessing motor speech disorders in early childhood. Speech assessment, much like speech practice, may involve repetition as SLPs seek to uncover patterns and variations in speech production, identify speech errors, and track progress or changes over time. This repetitiveness makes it difficult for SLPs to keep young patients engaged. Here, we explored how we can leverage the benefits of gamification to motivate and engage patients in speech practice [32, 35, 73].

However, designing for assessment poses another unique challenge—different from practice, assessments need to be accurate, precise, and reliable. In our scope of speech assessment for motor speech disorders, motion capture systems are highly accurate instruments to obtain measures of facial movements [1, 3], but those are not child-friendly.

7.1 Artifact Contribution, i.e., *ArticuMotion*

With *ArticuMotion,* we contribute an artifact [77] that enables contactless, non-invasive extraction of facial movement data based on videos of children as they perform speech tasks in a gamified sitting. In positioning games as assessments, however, a conundrum that we, as designers, faced was balancing the functional (e.g., maximizing the visibility of facial landmarks and minimizing children's movement) and clinical (e.g., uttering "pa pa") requirements with designing an enjoyable experience for children. This conundrum resembles well-known design tensions in serious game design—the art and practice of creating games for non-entertainment purposes [38].

To solve the above in our context, we embed these requirements in a fantasy context that makes it meaningful. We borrowed the metaphor of putting on a helmet in the Space Treasure Hunt game and riding a safari car in the Zoo Exploration adventure to gauge a front-facing view. To maintain that optimal position, we emphasized these requirements in context: "Remember we can't breathe in outer space without a helmet" and "The car will stop if you move." Similarly, This metaphorical context seems to have successfully created parallels between an unusual task (e.g., uttering "pa pa" as many times as you can in one breath) and a responsive environment, where this action is paralleled to the "blast of a spaceship" and "start up a fire engine." Based on our design and evaluation of *ArticuMotion*, we offer design considerations for speech assessment with preschool children next.

7.2 Designing for Aesthetic Experiences and Moving Beyond Badges

In *ArticuMotion*, we employed generative AI to create vibrant art assets in the style of popular cartoons. As suggested in co-design sessions, badges and reward systems were designed to boost extrinsic motivation and signal task completion. Surprisingly, when we asked children what they liked in *ArticuMotion*, none of the children mentioned the badges or in-game rewards. Instead, children highlighted in-game characters (e.g., alien and astronaut), animation (e.g., colorful tree), and progression (e.g., the ending). Our results align with the idea that rewards function as progress indicators and enhance user performance without affecting intrinsic motivation [53]. Furthermore, children expressed excitement when verbally encouraged and applauded by SLPs for completing the tasks, consistent with a previous study [14] suggesting that positive feedback improved both the choices users made and how interested they felt about what they were doing. Therefore, our findings suggest that the game's appeal and engagement were more related to overall aesthetics and visual elements than to extrinsic rewards like badges, which helped the children become intrinsically motivated by the game. AI-generated art seems appealing to children, therefore offering a viable route for rapid prototyping and content creation for game designers. We acknowledge that this could impact artists, which calls for further study.

7.3 Designing for Substitutivity

Substitutivity is a well-understood concept in HCI [18]. This idea is often implemented in the context of allowing equivalent input and output values to be substituted for each other, e.g., text and speech interfaces. For preschoolers, who are generally unable to read, substitutivity becomes even more needed. From our user study, we found the on-screen script impractical as it required that SLPs read the script for children, which can cause children to look away from the screen during tasks, hindering the visibility of facial landmarks in the video recording. Thus, in the polished version of *ArticuMotion,* we added audio narration to relieve SLPs from having to read the on-screen narrative.

We also added a reattempt speech tasks function to collect multiple data sets from a child if needed. Furthermore, in our study, we see this idea of substitutivity moving beyond the input/output modality to the game's context. As described, we embedded the same speech tasks in two different themes (space and zoo). We find that children have different preferences, which is to be expected. More importantly, we find that changing the context can conceal the repetitiveness of the tasks (e.g., P1 asking, "What's the next one called?" upon finishing both concepts). Given that we used generative AI to create art assets in *ArticuMotion,* our exploration suggests a potential future where rapid prototyping and dynamic generation of the context within which health assessment is embedded might be possible and worthwhile.

7.4 Designing for Personalization

SLPs in our PD suggested various techniques for scaffolding in speech assessment, such as prompts, cues, modeling, and collaboration, see Section. 4.3.1. Furthermore, SLPs underscored the importance of tailoring these techniques to a child's specific needs and abilities. In *ArticuMotion*, this is realized through the flexibility of administration, where SLPs can pause a task to give further cues and instruction, repeat and skip tasks as needed, and dynamically set the time duration for speech tasks. Furthermore, *ArticuMotion* offers personalization based on preference, where children can choose their character and the theme they want to play— Space Treasure Hunt game or Zoo Exploration Adventures for practicing similar tasks.

7.5 Limitations and Future Work

This research is a first step toward developing a low-cost, non-invasive, child-friendly speech assessment tool that could be clinically used by SLPs for objective diagnosis of motor speech disorders in young children. While our user study demonstrates the promising potential of *ArticuMotion*, it is important to acknowledge certain limitations that impact the generalizability of our findings. First, *ArticuMotion* was tested only with preschool children of typical development. As a result, the insights gained from our user studies might not fully capture the nuances and specific needs of preschoolers with motor speech disorders.

Additional research efforts are needed to evaluate *ArticuMotion* with children of diverse speech abilities and characteristics. Furthermore, five out of nine children who participated in our user study explicitly indicate prior experience using an iPad; children unfamiliar with tablets/iPads may have different expectations and experiences when using *ArticuMotion.* Additionally, our user study evaluates the user experience of *ArticuMotion.* In future research, we intend to validate the data captured within *ArticuMotion*, where we estimate measures of lip movements from *ArticuMotion's* 2D videos and compare those to corresponding 3D optical motion data. This would enable us to evaluate the effectiveness of *ArticuMotion* in estimating measures of facial movement and the robustness of our approach as compared to motion capture systems. Lastly, a clinical trial is needed to evaluate the clinical utility of *ArticuMotion.*

8 CONCLUSION

We presented *ArticuMotion,* an app that prompts children to perform speech assessment tasks in a gamified setting. *ArticuMotion* was co-created with SLPs and CVEs. This is a first step towards developing a non-intrusive and child-friendly speech assessment tool that could clinically be used to support objective assessment of motor speech disorders based on measures of facial movements. If successful, this tool or others with similar intentions will have profound implications because early diagnosis of speech disorders can lead to effective intervention, improved long-term functional communication outcomes, and a better quality of life for children.

9 SELECTION AND PARTICIPATION OF CHILDREN

Ethical approval was obtained from the Institutional Research Board at Northeastern IRB 18-05-28. To recruit children, we contacted the childcare center associated with our university. Nine English-speaking children (six boys and three girls) aged three to five years were recruited. Before the study, and in coordination with the director of the childcare center, we sent copies of the study description and consent forms to parents. Upon receiving written parental consent, children were invited, one by one, to a room at the childcare center where the study was set up. The center's director, with whom the children were familiar, assumed the role of guiding each child into the study setting, acquainting the child with the study team, and accompanying them back to their classrooms. She was present for the entire study duration, where a session lasted 12 minutes on average. A child-friendly language was used to explain the research study and procedures. Although parental consent had already been obtained before our user study, we designed the study protocol such that we explained the study objective, solicited children's agreement to engage in the study, and clarified that they were free to drop out of participation if they so wished, as described in Appendix. A. This consent protocol was designed to respect children's autonomy. All recordings were stored in a password-protected server accessible only to the study team.

ACKNOWLEDGMENTS

We are deeply grateful to Alyssa Sims and Jamie Klag for generously dedicating their time and providing invaluable insights that greatly enriched this work. We further extend our gratitude to the participants in our co-design sessions, the staff at Northeastern Childcare Center, the children, and their parents, without whom this study would not have been possible. This research was funded by an internal seed grant from Northeastern University. Furthermore, the first author acknowledges financial support from King Saud University and the Saudi Arabian Cultural Mission to the United States.

REFERENCES

[1] Kristen M. Allison, Ignatius S. B. Nip, and Panying Rong. 2022. Use of Automated Kinematic Diadochokinesis Analysis to Identify Potential Indicators of Speech Motor Involvement in Children With Cerebral Palsy. *American Journal of Speech-Language Pathology* 31, 6 (Nov. 2022), 2835–2846. https://doi.org/10.1044/2022_AJSLP-21-00241 Publisher: American Speech-Language-Hearing Association.

[2] ASHA American Speech-Language-Hearing Association. 2024. Speech Sound Disorders-Articulation and Phonology. https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology/

[3] Ioanna Anastasopoulou, Pascal van Lieshout, Douglas O. Cheyne, and Blake W. Johnson. 2022. Speech Kinematics and Coordination Measured With an MEG-Compatible Speech Tracking System. *Frontiers in Neurology* 13 (2022). https://www.frontiersin.org/articles/10.3389/fneur.2022.828237

[4] Geertruida Aline Attwell, Kwabena Ebo Bennin, and Bedir Tekinerdogan. 2022. A Systematic Review of Online Speech Therapy Systems for Intervention in Childhood Speech Communication Disorders. *Sensors* 22, 24 (Jan. 2022), 9713. https://doi.org/10.3390/s22249713 Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

[5] Elise Baker and Sharynne McLeod. 2011. Evidence-Based Practice for Children With Speech Sound Disorders: Part 1 Narrative Review. *Language, Speech, and Hearing Services in Schools* 42, 2 (April 2011), 102–139. https://doi.org/10.1044/0161-1461(2010/09-0075) Publisher: American Speech-Language-Hearing Association.

[6] Irene Velsvik Bele. 2005. Reliability in perceptual analysis of voice quality. *Journal of Voice: Official Journal of the Voice Foundation* 19, 4 (Dec. 2005), 555–573. https://doi.org/10.1016/j.jvoice.2004.08.008

[7] Gayle Marielle A. Belen, Kim Benedict P. Cabonita, Vanessa A. dela Peña, Hayfah A. Dimakuta, Fritz G. Hoyle, and Ralph P. Laviste. 2018. Tingog: Reading and Speech Application for Children with Repaired Cleft Palate. In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*. 1–6. https://doi.org/10.1109/HNICEM.2018.8666309

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa.

[9] Tim Brown and Jocelyn Wyatt. 2010. Design Thinking for Social Innovation By. *Stanford Social Innovation Review* 8 (July 2010), 30–35. https://doi.org/10.1596/1020-797X_12_1_29

[10] Kate Bunton, Raymond D. Kent, Joseph R. Duffy, John C. Rosenbek, and Jane F. Kent. 2007. Listener agreement for auditory-perceptual ratings of dysarthria. *Journal of speech, language, and hearing research: JSLHR* 50, 6 (Dec. 2007), 1481–1495. https://doi.org/10.1044/1092-4388(2007/102)

[11] Elizabeth Carrow-Woolfolk. 2014. TACL4 Test for Auditory Comprehension of LanguageFourth Edition. https://www.proedinc.com/Products/12700/tacl4-test-for-auditory-comprehension-of-languagefourth-edition.aspx

[12] Sofia Cavaco, Isabel Guimarães, Mariana Ascensão, Alberto Abad, Ivo Anjos, Francisco Oliveira, Sofia Martins, Nuno Marques, Maxine Eskenazi, João Magalhães, and Margarida Grilo. 2020. The BioVisualSpeech Corpus of Words with Sibilants for Speech Therapy Games Development. *Information* 11, 10 (Oct. 2020), 470. https://doi.org/10.3390/info11100470 Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[13] Yi-Ping Phoebe Chen, Caddi Johnson, Pooia Lalbakhsh, Terry Caelli, Guang Deng, David Tay, Shane Erickson, Philip Broadbridge, Amr El Refaie, Wendy Doube, and Meg E. Morris. 2016. Systematic review of virtual speech therapists for speech disorders. *Computer Speech & Language* 37 (May 2016), 98–128. https://doi.org/10.1016/j.csl.2015.08.005

[14] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125, 6 (1999), 627–668. https://doi.org/10.1037/0033-2909.125.6.627 Place: US Publisher: American Psychological Association.

[15] Sebastian Deterding. 2012. Gamification: designing for motivation. *Interactions* 19, 4 (July 2012), 14–17. https://doi.org/10.1145/2212877.2212883

[16] Sebastian Deterding, Alessandro Canossa, Casper Harteveld, Seth Cooper, Lennart E. Nacke, and Jennifer R. Whitson. 2015. Gamifying Research: Strategies, Opportunities, Challenges, Ethics. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 2421–2424. https://doi.org/10.1145/2702613.2702646

[17] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining Gamification, Vol. 11. 9–15. https://doi.org/10.1145/2181037.2181040

[18] Alan Dix, Janet E. Finlay, Gregory D. Abowd, and Russell Beale. 2003. *Human-Computer Interaction* (3rd edition ed.). Pearson, Harlow, England ; New York.

[19] Barbara Dodd. 2013. *Differential Diagnosis and Treatment of Children with Speech Disorder.* John Wiley & Sons. Google-Books-ID: efOrRCx4H88C.

[20] Yao Du, Kathryn Lubniewski, Lori Price, Grace Breslin, Paula Thomson, Natashka Jinadasa, and Nikita Soni. 2023. "They Can't Believe They're a Tiger": Insights from pediatric speech-language pathologist mobile app users and app designers. *International Journal of Language & Communication Disorders* 58, 5 (2023), 1717–1737. https://doi.org/10.1111/1460-6984.12898 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1460-6984.12898.

[21] Andreas Duenser, Daniel Smith, Jill Freyne, Elizabeth Gilleran, Kristy de Salas, and Hilary Cinis. 2018. Engaging assessments: interface design of a referral support tool for adults and children. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 42–52. https://doi.org/10.1145/3292147.3292166

[22] Joseph R. Duffy. 2013. *Motor speech disorders substrates, differential diagnosis, and management* (3rd ed. ed.). Elsevier Mosby, St. Louis, Mo.

[23] Jared Duval, Zachary Rubin, Elena Márquez Segura, Natalie Friedman, Milla Zlatanov, Louise Yang, and Sri Kurniawan. 2018. SpokeIt: building a mobile speech therapy experience. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3229434.3229484

[24] Marge Edwards, Arlene Stredler-Brown, and K. Todd Houston. 2012. Expanding Use of Telepractice in Speech-Language Pathology and Audiology. *Volta Review* 112, 3 (2012), 227–242. https://doi.org/10.17955/tvr.112.3.m.704 Publisher: Alexander Graham Bell Association for the Deaf.

[25] Charlotta Elo, Mauri Inkinen, Eveliina Autio, Tanja Vihriälä, and Johanna Virkki. 2022. The Role of Games in Overcoming the Barriers to Paediatric Speech Therapy Training. In *Games and Learning Alliance (Lecture Notes in Computer Science)*, Kristian Kiili, Koskinen Antti, Francesca de Rosa, Muhterem Dindar, Michael Kickmeier-Rust, and Francesco Bellotti (Eds.). Springer International Publishing,

Cham, 181–192. https://doi.org/10.1007/978-3-031-22124-8_18

[26] Olov Engwall. 2005. Introducing visual cues in acoustic-to-articulatory inversion. 3205–3208. https://doi.org/10.21437/Interspeech.2005-846

[27] Olov Engwall, Olle Bälter, Anne-Marie Öster, and Hedvig Kjellström. 2006. Designing the user interface of the computer-based speech training system ARTUR based on early user tests. *Behaviour & Information Technology* 25, 4 (July 2006), 353–365. https://doi.org/10.1080/01449290600636702 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01449290600636702.

[28] Glenn Craig Fairweather, Michelle Ann Lincoln, and Robyn Ramsden. 2016. Speech-language pathology teletherapy in rural and remote educational settings: Decreasing service inequities. *International Journal of Speech-Language Pathology* 18, 6 (Nov. 2016), 592–602. https://doi.org/10.3109/17549507.2016.1143973 Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/17549507.2016.1143973.

[29] Lisa Furlong, Shane Erickson, and Meg E. Morris. 2017. Computer-based speech therapy for childhood speech sound disorders. *Journal of Communication Disorders* 68 (July 2017), 50–69. https://doi.org/10.1016/j.jcomdis.2017.06.007

[30] Shirley Gherson, Raquel Tripp, Deanna Goudelias, and Aaron M. Johnson. 2023. Rapid Implementation of Teletherapy for Voice Disorders: Challenges and Opportunities for Speech-Language Pathologists. *Journal of Voice* (Aug. 2023). https://doi.org/10.1016/j.jvoice.2023.06.024

[31] F. P. Glascoe. 1991. Can clinical judgment detect children with speech-language problems? *Pediatrics* 87, 3 (March 1991), 317–322.

[32] Adam Hair, Kirrie J. Ballard, Constantina Markoulli, Penelope Monroe, Jacqueline Mckechnie, Beena Ahmed, and Ricardo Gutierrez-Osuna. 2021. A Longitudinal Evaluation of Tablet-Based Child Speech Therapy with Apraxia World. *ACM Transactions on Accessible Computing* 14, 1 (March 2021), 3:1–3:26. https://doi.org/10.1145/3433607

[33] Adam Hair, Constantina Markoulli, Penelope Monroe, Jacqueline McKechnie, Kirrie J. Ballard, Beena Ahmed, and Ricardo Gutierrez-Osuna. 2020. Preliminary Results From a Longitudinal Study of a Tablet-Based Speech Therapy Game. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382886

[34] Adam Hair, Penelope Monroe, Beena Ahmed, Kirrie J. Ballard, and Ricardo Gutierrez-Osuna. 2018. Apraxia world: a speech therapy game for children with speech sound disorders. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, Trondheim Norway, 119–131. https://doi.org/10.1145/3202185.3202733

[35] Sadrieh Hajesmaeel-Gohari, Saeideh Goharinejad, Elaheh Shafiei, and Kambiz Bahaadinbeigy. 2023. Digital games for rehabilitation of speech disorders: A scoping review. *Health Science Reports* 6, 6 (June 2023), e1308. https://doi.org/10.1002/hsr2.1308

[36] Benjamin Halberstam. 2004. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL; journal for oto-rhino-laryngology and its related specialties* 66, 2 (2004), 70–73. https://doi.org/10.1159/000077798

[37] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences*. 3025–3034. https://doi.org/10.1109/HICSS.2014.377 ISSN: 1530-1605.

[38] Casper Harteveld. 2011. *Triadic Game Design: Balancing Reality, Meaning and Play* (1st ed.). Springer Publishing Company, Incorporated.

[39] Alexis Hiniker, Kiley Sobel, and Bongshin Lee. 2017. Co-Designing with Preschoolers Using Fictional Inquiry and Comicboarding. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5767–5772. https://doi.org/10.1145/3025453.3025588

[40] Robin Hunicke, Marc Leblanc, and Robert Zubek. 2004. MDA: A Formal Approach to Game Design and Game Research. *AAAI Workshop - Technical Report* 1 (Jan. 2004).

[41] Seigel Jenya Iuzzini, Kristen M. Allison, and Ruth Stoeckel. 2022. A Tool for Differential Diagnosis of Childhood Apraxia of Speech and Dysarthria in Children: A Tutorial. *Language, Speech, and Hearing Services in Schools* 53, 4 (Oct. 2022), 926–946. https://doi.org/10.1044/2022_LSHSS-21-00164 Publisher: American Speech-Language-Hearing Association.

[42] Seigel Jenya Iuzzini, Tiffany P. Hogan, and Jordan R. Green. 2017. Speech Inconsistency in Children With Childhood Apraxia of Speech, Language Impairment, and Speech Delay: Depends on the Stimuli. *Journal of Speech, Language, and Hearing Research* 60, 5 (May 2017), 1194–1210. https://doi.org/10.1044/2016_JSLHR-S-15-0184 Publisher: American Speech-Language-Hearing Association.

[43] Jenya Iuzzini-Seigel, Kristen M. Allison, and Ruth Stoeckel. 2022. A Tool for Differential Diagnosis of Childhood Apraxia of Speech and Dysarthria in Children: A Tutorial. *Language, Speech, and Hearing Services in Schools* 53, 4 (Oct. 2022), 926–946. https://doi.org/10.1044/2022_LSHSS-21-00164

[44] Sophie Jullien. 2021. Screening for language and speech delay in children under five years. *BMC Pediatrics* 21, 1 (Sept. 2021), 362. https://doi.org/10.1186/s12887-021-02817-7

[45] Emmanuel M. Kalargiros and Michael R. Manning. 2015. Divergent Thinking and Brainstorming in Perspective: Implications for Organization Change and Innovation. In *Research in Organizational Change and Development*. Research in Organizational Change and Development, Vol. 23. Emerald Group Publishing Limited, 293–327. https://doi.org/10.1108/S0897-301620150000023007

[46] Raymond D. Kent, L. Pagan-Neves, K. Hustad, and H. Wertzner. 2009. CHAPTER 5 Children ' s Speech Sound Disorders : An Acoustic Perspective. https://www.semanticscholar.org/paper/CHAPTER-5-Children-%E2%80%99-s-Speech-Sound-Disorders-%3A-An-Kent-Pagan-Neves/efa33d92f5b70d1039dfa65d2e04f533ab797dc5

[47] Mónica Lancheros, Daniel Friedrichs, and Marina Laganaro. 2023. What Do Differences between Alternating and Sequential Diadochokinetic Tasks Tell Us about the Development of Oromotor Skills? An Insight from Childhood to Adulthood. *Brain Sciences* 13, 4 (April 2023), 655. https://doi.org/10.3390/brainsci13040655

[48] Julie M. Liss, Laurence White, Sven L. Mattys, Kaitlin Lansford, Andrew J. Lotto, Stephanie M. Spitzer, and John N. Caviness. 2009. Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of speech, language, and hearing research: JSLHR* 52, 5 (Oct. 2009), 1334–1352. https://doi.org/10.1044/1092-4388(2009/08-0208)

[49] Edwin Maas, Christina Gildersleeve-Neumann, Kathy J. Jakielski, and Ruth Stoeckel. 2014. Motor-based intervention protocols in treatment of childhood apraxia of speech (CAS). *Current Developmental Disorders Reports* 1, 3 (Sept. 2014), 197–206. https://doi.org/10.1007/s40474-014-0016-4

[50] Maura R. McLaughlin. 2011. Speech and language delay in children. *American Family Physician* 83, 10 (May 2011), 1183–1188.

[51] Sharynne Mcleod and Elise Baker. 2014. Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics & Phonetics* 28, 7-8 (2014), 508–531. https://doi.org/10.3109/02699206.2014.926994

[52] Sharynne McLeod, Linda J. Harrison, Lindy McAllister, and Jane McCormack. 2013. Speech sound disorders in a community study of preschool children. *American Journal of Speech-Language Pathology* 22, 3 (Aug. 2013), 503–522. https://doi.org/10.1044/1058-0360(2012/11-0123)

[53] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Do points, levels and leaderboards harm intrinsic motivation? an empirical analysis of common gamification elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications (Gamification '13)*. Association for Computing Machinery, New York, NY, USA, 66–73. https://doi.org/10.1145/2583008.2583017

[54] W. Minnema and H.-C Stoll. 2008. Objective assessment of voice disorders with Praat. 22 (July 2008), 24–29.

[55] Elizabeth Murray, Patricia McCabe, and Kirrie J. Ballard. 2014. A Systematic Review of Treatment Outcomes for Children With Childhood Apraxia of Speech. *American Journal of Speech-Language Pathology* 23, 3 (Aug. 2014), 486–504. https://doi.org/10.1044/2014_AJSLP-13-0035 Publisher: American Speech-Language-Hearing Association.

[56] National Institute on Deafness and Other Communication Disorders NIDCD. 2016. Quick Statistics About Voice, Speech, Language | NIDCD. https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language

[57] Jennifer Oates. 2009. Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP)* 61, 1 (2009), 49–56. https://doi.org/10.1159/000200768

[58] Avinash Parnandi, Virendra Karappa, Youngpyo Son, Mostafa Shahin, Jacqueline McKechnie, Kirrie Ballard, Beena Ahmed, and Ricardo Gutierrez-Osuna. 2013. Architecture of an automated therapy tool for childhood apraxia of speech. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2513383.2513450

[59] Lindsay Pennington, Naomi K. Parker, Helen Kelly, and Nick Miller. 2016. Speech therapy for children with dysarthria acquired before three years of age. *Cochrane Database of Systematic Reviews* 7 (2016). https://doi.org/10.1002/14651858.CD006937.pub3 Publisher: John Wiley & Sons, Ltd.

[60] Elizabeth R. Peterson and Kevin A. Barron. 2007. How to Get Focus Groups Talking: New Ideas that will Stick. *International Journal of Qualitative Methods* 6, 3 (Sept. 2007), 140–144. https://doi.org/10.1177/160940690700600303 Publisher: SAGE Publications Inc.

[61] Patricia A. Prelock, Tiffany Hutchins, and Frances P. Glascoe. 2008. Speech-Language Impairment: How to Identify the Most Common and Least Diagnosed Disability of Childhood. *The Medscape Journal of Medicine* 10, 6 (June 2008), 136. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491683/

[62] Ronald Goldman, PhD and Macalyne Fristoe, PhD. 2015. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Developmental-Early-Childhood/Goldman-Fristoe-Test-of-Articulation-3/p/100001202.html

[63] Sara Rosenbaum, Patti Simon, Committee on the Evaluation of the Supplemental Security Income (SSI) Disability Program for Children with Speech Disorders and Language Disorders, Board on the Health of Select Populations, Youth Board on Children, Institute of Medicine, Division of Behavioral and Social Sciences Education, , and Engineering National Academies of Sciences. 2016. Childhood Speech and Language Disorders in the General U.S. Population. In *Speech and Language Disorders in Children: Implications for the Social Security Administration's Supplemental Security Income Program*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK356270/

[64] Jeff Sauro and James R. Lewis. 2005. Estimating Completion Rates from Small Samples Using Binomial Confidence Intervals: Comparisons and Recommendations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 49, 24 (Sept. 2005), 2100–2103. https://doi.org/10.1177/154193120504902407 Publisher: SAGE Publications Inc.

[65] Jordanna S. Sevitz, Brianna R. Kiefer, Jessica E. Huber, and Michelle S. Troche. 2021. Obtaining Objective Clinical Measures During Telehealth Evaluations of Dysarthria. *American Journal of Speech-Language Pathology* 30, 2 (March 2021), 503–516. https://doi.org/10.1044/2020_AJSLP-20-00243

[66] Mostafa Shahin, Beena Ahmed, Avinash Parnandi, Virendra Karappa, Jacqueline McKechnie, Kirrie J. Ballard, and Ricardo Gutierrez-Osuna. 2015. Tabby Talks: An automated tool for the assessment of childhood apraxia of speech. *Speech Communication* 70 (June 2015), 49–64. https://doi.org/10.1016/j.specom.2015.04.002

[67] Laura Sices, Chris Feudtner, John McLaughlin, Dennis Drotar, and Michelle Williams. 2003. How do primary care physicians identify young children with developmental delays? A national survey. *Journal of developmental and behavioral pediatrics: JDBP* 24, 6 (Dec. 2003), 409–417. https://doi.org/10.1097/00004703-200312000-00002

[68] Jesper Simonsen. 2004. Participatory IT Design. Designing for Business and Workplace Realities. MIT press. *Bødker, K., F. Kensing and J. Simonsen (2004). Participatory IT Design. Designing for Business and Workplace Realities. MIT press.* (Jan. 2004). https://www.academia.edu/38351927/Participatory_IT_Design_Designing_for_Business_and_Workplace_Realities_MIT_press

[69] M. Snowling, D. V. Bishop, and S. E. Stothard. 2000. Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 41, 5 (July 2000), 587–600. https://doi.org/10.1111/1469-7610.00651

[70] Margaret J. Snowling, D. V. M. Bishop, Susan E. Stothard, Barry Chipchase, and Carole Kaplan. 2006. Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 47, 8 (Aug. 2006), 759–765. https://doi.org/10.1111/j.1469-7610.2006.01631.x

[71] Janet B. Fudala and Sheri Stegall. 2017. Arizona™-4 Arizona Articulation and Phonology Scale, Fourth Revision. https://www.wpspublish.com/arizona-4-arizona-articulation-and-phonology-scale-fourth-revision

[72] Katie Salen Tekinbas and Eric Zimmerman. 2003. *Rules of Play: Game Design Fundamentals*. MIT Press. Google-Books-ID: YrT4DwAAQBAJ.

[73] Carolyne Alphonsus Tommy, Jacey Lynn Minoi, and Chin Saw Sian. 2019. Assessing Fun and Engagement in Mobile Applications for Children with Speech Delay. *Applied Mechanics and Materials* 892 (2019), 79–87. https://doi.org/10.4028/www.scientific.net/AMM.892.79 Publisher: Trans Tech Publications Ltd.

[74] Michael P. Towey. 2012. Speech Telepractice: Installing a Speech Therapy Upgrade for the 21st Century. *International Journal of Telerehabilitation* 4, 2 (Dec. 2012), 73–78. https://doi.org/10.5195/IJT.2012.6112

[75] Giovanni Maria Troiano, Esben Warming Pedersen, and Kasper Hornbæk. 2014. User-defined gestures for elastic, deformable displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2598153.2598184

[76] R. Waring and R. Knight. 2013. How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language & Communication Disorders* 48, 1 (Jan. 2013), 25–40. https://doi.org/10.1111/j.1460-6984.2012.00195.x

[77] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research contributions in human-computer interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069

[78] B. J. Zyski and B. E. Weisiger. 1987. Identification of dysarthria types based on perceptual analysis. *Journal of Communication Disorders* 20, 5 (Oct. 1987), 367–378. https://doi.org/10.1016/0021-9924(87)90025-6

A USER STUDY PROTOCOL

A.1 Consent

We are going to play some games on the iPad today. Does that sound okay to you? Your grown-ups said it is okay for you to help us, but if you don't want to play, you don't have to, and nobody will be upset. And if you want to take a break, we can do that.

- Does that sound good to you?
- Are you ready to get started?
- Do you have any questions?

Participants were then asked if they wanted to write, spell, or say their names, which were documented on a consent form.

A.2 Pre-interview

- How old are you? When is your birthday?
- We are going to play some games on the iPad now. Does that sound okay to you?
- Do you play on the iPad at home? What is your favorite game/show?

A.3 Post-interview

- Can I ask you a couple of questions about the game?
- What did you think about these games?
- Which one was your favorite?
- Can you tell me what you liked about it?
- Would you want to play it again?