Evaluating Speech Technology Integration in Children's Reading Education: A Study of TTS and STT Implementation in an Icelandic Educational Game

Alexander Guðmundsson

Faculty of Electrical and Computer Engineering
University of Iceland
2024

EVALUATING SPEECH TECHNOLOGY INTEGRATION IN CHILDREN'S READING EDUCATION: A STUDY OF TTS AND STT IMPLEMENTATION IN AN ICELANDIC EDUCATIONAL GAME

Alexander Guðmundsson

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Computer Science

Advisor
Hafsteinn Einarsson

Faculty Representative
XXNN3

M.Sc. Committee
Hafsteinn Einarsson
Steinn Guðmundsson

Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, October 2024

Evaluating Speech Technology Integration in Children's Reading Education: A Study of TTS and STT Implementation in an Icelandic Educational Game
Speech Tech Impact on Kids' Reading: A Gamified Study)
60 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Computer Science

Faculty of Electrical and Computer Engineering
School of Engineering and Natural Sciences
University of Iceland
Tæknigarður - Dunhagi 5, 107 Reykjavík
102, Reykjavik, Reykjavik Iceland

Telephone: 525 4000

Bibliographic information:
Alexander Guðmundsson, 2024, Evaluating Speech Technology Integration in Children's Reading Education: A Study of TTS and STT Implementation in an Icelandic Educational Game, M.Sc. thesis, Faculty of Electrical and Computer Engineering, University of Iceland.

ISBN XX

Printing: Háskólaprent, Fálkagata 2, 107 Reykjavík
Reykjavik, Iceland, October 2024

Abstract

This thesis examines the development and evaluation of a novel educational application that integrates Speech-to-Text (STT) and Text-to-Speech (TTS) technologies to support reading education among Icelandic children. Through a mixed-methods study involving nine participants aged 5-9 years, the research assessed both the technical feasibility of these technologies and their impact on user engagement. The TTS component demonstrated high effectiveness, with most participants achieving accuracy rates above 90% in listening comprehension tasks. However, the STT functionality showed significant limitations in accurately recognizing children's speech, particularly in the Icelandic language context. Sentiment analysis revealed positive emotional responses to the gamified learning environment, suggesting successful user engagement despite technical challenges. The study identified critical areas for improvement in speech recognition technology for minority languages while confirming the viability of TTS as a supportive tool for reading education. These findings contribute to the growing field of educational technology for minority languages and provide valuable insights for the future development of language learning applications. The research highlights the need for continued development of language-specific speech recognition models while demonstrating the potential of integrated speech technologies to enhance children's reading education.

Útdráttur

Þessi meistararitgerð skoðar þróun og mat á nýstárlegum námshugbúnaði sem samþættir talgreiningu og talgervingu til að styðja við lestrarnám íslenskra barna. Rannsóknin, sem byggir á blönduðum rannsóknaraðferðum með níu þátttakendum á aldrinum 5-9 ára, mat bæði tæknilega framkvæmanleika þessarar tækni og áhrif hennar á þátttöku nemenda. Talgervingarhluti forritsins virkaði vel en flestir þátttakendur náðu yfir 90% nákvæmni í verkefnum sem reyndu á hlustun. Hins vegar var talgreiningar hlutinn takmörkunum háður við að þekkja rétt tal barna. Tilfinningagreining leiddi í ljós að viðbrögð við leikjavædda námsumhverfinu voru að mestu jákvæð, sem bendir til árangursríkrar þátttöku notenda þrátt fyrir tæknilegar áskoranir. Rannsóknin greindi mikilvæg atriði til úrbóta í raddgreiningu fyrir tungumál eins og íslensku en staðfesti um leið notagildi talgervingartækninnar sem stuðningstækis við lestrarnám. Niðurstöðurnar leggja til mikilvægt framlag til vaxandi sviðs menntatækni fyrir tungumál sem töluð eru af fáum og veita verðmæta innsýn fyrir frekari þróun tungumálanámshugbúnaðar. Rannsóknin undirstrikar þörfina á áframhaldandi þróun sértækra talgreiningarlíkana en sýnir jafnframt fram á möguleika samþættrar raddtækni til að efla lestrarnám barna.

Acknowledgements

I would like to express my deepest gratitude to all those who supported me throughout the process of writing this thesis.

Special thanks go to my supervisor, Hafsteinn Einarsson, for his invaluable guidance, especially in helping me develop the project's scope, apply for funding, and structure the research, including setting up the longitudinal research plans, which, although not implemented due to technical challenges and delays, laid a solid foundation for the current case study.

I am particularly thankful to Helga Sigurmundsdóttir, who contributed significantly by conducting the individual case studies, and to the child participants and their parents for dedicating their time and effort to this research. I also extend my appreciation to the staff and teachers at Rimaskóli for their enthusiasm and willingness to assist in coordinating the research process, even though it had to be adapted due to logistical constraints.

I would also like to thank 100 Ord for providing the word-list that was used in the app development, and my employer, Opin Kerfi, whose technical infrastructure support was invaluable in the late stages of development.

Lastly, I owe a great debt of gratitude to my family: to my sister, who provided me with the laptop that made long development hours possible, and to my mother and brother who supported me emotionally through difficult times.

Disclaimer

In the preparation of this thesis, I utilized OpenAI's ChatGPT and QuillBot, both AI language models, to assist in various aspects of the research and writing process. The AI tools were employed in the following ways:

- **Idea Generation and Brainstorming**: ChatGPT was used to help generate ideas and perspectives during the initial stages of research, aiding in the development of research questions and methodology.

- **Language and Style Refinement**: Draft sections of the thesis were input into both ChatGPT and QuillBot to receive suggestions on improving clarity, coherence, and academic tone. These tools provided alternative phrasings and structural suggestions, which were carefully reviewed and incorporated where appropriate.

- **Literature Search Assistance**: ChatGPT assisted in identifying relevant literature by suggesting potential sources based on specific research topics and keywords. All suggested references were independently verified for accuracy and relevance before inclusion.

- **Interpretation of Results**: The AI tools helped explore different interpretations of the data collected. By providing summaries of findings to ChatGPT, it offered alternative perspectives and highlighted potential implications, which informed the analysis presented in this thesis. All interpretations were critically evaluated to ensure they aligned with the research objectives and data.

- **Proofreading and Editing Support**: QuillBot was used to help identify grammatical errors, improve sentence structure, and enhance the overall readability of the thesis draft. Final proofreading was conducted by me to ensure accuracy and consistency.

All content presented in this thesis is the result of my own research and analysis. The use of ChatGPT and QuillBot was limited to the support functions described above and did not replace my original work or critical thinking. I ensured that all AI-assisted inputs were critically evaluated and that the final decisions regarding content, interpretations, and conclusions were made independently.

This acknowledgment of AI assistance is provided to maintain transparency and uphold academic integrity. The use of ChatGPT and QuillBot adheres to the ethical guidelines and academic policies of the University of Iceland regarding the use of AI tools in scholarly work.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Introduction to the Problem

Iceland has been facing a significant decline in children's reading performance, with 25% of students falling below basic reading standards, according to the OECD, 2019, 2023. This decline in reading performance is a growing concern, especially compared to the OECD mean performance. Although the main reasons for this decline are not clear, several studies offer some possible reasons (Birgisdóttir, 2016; Ólafsdóttir, Sigurðsson et al., 2017). Iceland stopped standardized testing at the elementary school level in 2009 when reading performance started declining. Furthermore, there are concerns about the growing impact of modern technology, which provides more engaging distractions, such as games, social media, and video streaming, that might explain reduced interest in reading.

Although technology is often seen as a distraction from traditional learning, it also has the potential to serve as a powerful educational tool. Studies have shown that integrating educational technology into learning environments can enhance student engagement, improve learning outcomes, and foster personalized learning experiences (Kim and Frick, 2020; Kucirkova, 2020). According to Hamari et al. (2016), tools such as interactive reading applications and digital games have effectively enhanced children’s engagement and motivation in learning. Specifically, digital reading applications have effectively improved children's reading skills (Castek et al., 2020). However, the challenge lies in harnessing this potential to create engaging and educational learning experiences that turn technology's appeal into an advantage for reading development. Consequently, this research is driven by the need to explore how educational technology can be utilized to address the ongoing decline in reading performance among Icelandic students.

1.2 Purpose of the Study

This research investigates the potential of combining gamified learning with Speech-to-text (STT) and Text-to-Speech (TTS) technologies to enhance children’s reading skills. Building on the work of Þuríður Hilmarsdóttir Hilmarsdóttir in 2020, who developed an app for learning to read in her bachelor’s thesis, this study expands the application by incorporating gamified features tailored for Icelandic children aged five to ten years. The app is designed to make the process of learning to read more enjoyable, incorporating a point-based reward system that motivates children through consistent effort and progress.

Incorporating STT allows for rapid feedback on reading performance, while TTS technology is specifically used to facilitate exercise development and creation. This study was conducted in collaboration with *Helga Sigurmundsdóttir*, a local teacher, who worked closely with the children to assess their readiness for various tasks. A series of individual case studies were conducted, where each child engaged with the app in a single session. The main focus of the study was to understand the app's immediate impact on reading performance and emotional engagement across two types of tasks:

- **Listening tasks** (non-voice): The app's TTS system reads aloud a letter, word, or sentence, and the child selects the correct option from two choices.

- **Voice-based tasks**: The child reads aloud using the STT technology, which evaluates their pronunciation and accuracy.

1.3 Research Questions

The primary goal of this research is to assess the effectiveness of integrating **STT** and **TTS** technologies within a gamified reading app and study how children react to such an interface.

Main Research Question:

- *Are STT and TTS sufficiently mature technologies in Icelandic to be applied in a reading application for children?*

We also consider the following sub-question:

1. *What are the immediate effects of a single session with the gamified reading app on children's engagement levels, particularly comparing non-voice tasks (listening) with voice-based tasks (reading aloud)?*

1.4 Significance of the Study

Reading performance is fundamental to academic success and personal development. Finding **innovative ways** to improve reading skills in the digital age is crucial. This study aims to contribute to the field by studying whether automation through **STT** and **TTS** technologies is a feasible avenue to further develop apps that could improve reading performance. To our knowledge, this is the first time that these two technologies have been combined in an app to teach users to read in Icelandic.

2 Literature Review

2.1 Educational Technology and Reading Proficiency

Educational technology has significantly impacted various aspects of learning, particularly in enhancing student engagement. Integrating digital tools into education has shown promising results in improving student motivation and active participation. Studies have shown that integrating educational technology into learning environments can enhance student engagement, improve learning outcomes, and foster personalized learning experiences (Kim and Frick, 2020). Specifically, Kucirkova (2020) discusses how personalized learning with digital technologies can support reading development by catering to individual student needs.

2.1.1 Advanced Voice AI in Educational Technology

Recent advancements in voice AI technologies have transformed how students engage with educational tools. For instance, SoapBox Labs has developed an accurate voice recognition engine tailored for children's speech patterns. Integrated into platforms such as Scholastics' "Ready4Reading" program, this AI technology enhances phonics instruction and provides real-time feedback, increasing engagement through personalized feedback (Labs, 2023; Room, 2023). This innovation highlights the potential of voice AI in improving engagement and interaction during learning. In addition to SoapBox Labs, other voice AI tools like Google's "Read Along" app have demonstrated considerable promise in guiding and correcting children's reading aloud in real-time. Although these tools are primarily designed for English, they have established a precedent for the potential of voice AI to enhance engagement in educational applications (AI, 2023).

2.1.2 Icelandic-Specific TTS Development

Developing language models in STT and TTS has been crucial for Icelandic speakers. Projects funded by Almannarómur, in collaboration with Icelandic universities, have created open-source TTS systems for Icelandic, offering students high-quality learning experiences that are both linguistically and culturally relevant. Most systems have been based on data collected from the Samrómur project (Hedström et al., 2022).

However, challenges persist, particularly for minority languages such as Icelandic. These include phonetic accuracy and dialectal variations, which can impede the efficacy of STT and TTS systems. Research by Besacier et al. (2014b) highlights the need for more sophisticated language models to capture the nuances of minority languages better, ensuring that these tools are as effective in non-English contexts as they are in English-dominant ones.

2.2 Gamified Learning

Research by Hirsh-Pasek et al. (2015) underscores the significance of user-centered design in educational technology, especially in applications targeting children. Educational tools

designed with guided play and active engagement principles have proven to be more effective in maintaining student interest and fostering engagement. This brings us to the concept of gamification.

Gamification involves applying game design elements to non-game contexts and has been shown to increase motivation and engagement in learning. Deterding et al. (2011) highlight how gamification can enhance user engagement and motivation. Studies have demonstrated that gamification significantly improves user motivation by providing enjoyable and interactive experiences (Hamari et al., 2016). While Hamari et al.'s (2016) research spans various fields, the principles are particularly relevant in educational contexts, including reading applications, where gamified platforms can encourage students to practice more frequently.

Research by Pekrun (2006) and Ryan and Deci (2000) emphasizes the importance of positive emotions and intrinsic motivation in learning, suggesting that when students feel a sense of autonomy and competence, their motivation and engagement increase. Applying these principles, gamification elements that elicit positive emotions, such as curiosity and enjoyment, may lead to higher motivation and better learning outcomes.

2.3 Speech-to-Text (STT) and Text-to-Speech (TTS) Technology

STT and TTS technologies have been transformative in education, particularly in improving engagement during language learning. These tools offer immediate, personalized feedback, which keeps students engaged and motivated during learning tasks. Marin et al. (2015) discuss how TTS technology enhances reading skills by providing real-time auditory feedback, allowing students to adjust their responses immediately. This immediate feedback can also contribute to increased engagement during reading activities. Bailey and Wolfson (2020) similarly highlight that advancements in speech-to-text technology have enhanced its accuracy and accessibility, consequently improving students' reading proficiency.

The development of specific STT and TTS models for children's speech, such as those by SoapBox Labs, has been pivotal in increasing engagement during reading tasks. By tailoring feedback to individual students, these systems ensure that students remain focused and engaged throughout their learning sessions (Labs, 2023). While STT and TTS technologies have advanced significantly, challenges remain for minority languages like Icelandic. Notably, issues such as phonetic accuracy and dialectal variations can hinder the effectiveness of these systems. The Language Technology Programme for Icelandic, as discussed by Steingrímsson et al. (2020), aims to address these challenges by developing advanced language resources and models specifically for Icelandic.

Sigurgeirsson et al. (2021) introduce Talrómur, a large Icelandic TTS corpus designed to improve the quality and accuracy of TTS systems for Icelandic, taking into account phonetic nuances and dialectal variations. Advancements in speech-to-text (STT) and text-to-speech (TTS) technologies have created new opportunities, particularly in education, prompting inquiries into the potential of automated tools to enhance reading proficiency beyond conventional approaches.

2.4 Assessment of Reading Skills

Assessing reading proficiency involves evaluating various components such as comprehension, fluency, and accuracy. Traditional methods, such as oral reading fluency tests, have been complemented by digital assessments that provide more detailed and immediate feedback.

Use of Digital Games for Assessment: Shute and Ke (2012) explore the potential of digital games in assessing and developing skills, including reading. They argue that games can provide a dynamic and interactive environment for assessment, offering immediate feedback and adapting to the learner's level. This approach aligns well with formative assessment principles, where ongoing feedback guides and improves learning. A more recent study by Ke (2019) emphasizes the effectiveness of digital games in educational assessments.

Measuring Reading Proficiency: Measuring reading proficiency requires a combination of quantitative and qualitative methods. Fuchs et al. (2001) underscore the necessity of both speed and accuracy in reading, which are essential for comprehension, in their discussion of the significance of oral reading fluency as an indicator of reading competence. Building on this, digital tools incorporating STT and TTS technologies can potentially enhance these traditional measures by providing detailed data on student performance. Recent advancements in digital assessment tools, as articulated by Dalton and Proctor (2021), have further improved the precision and usability of these measures.

In Iceland, reading performance has traditionally been measured using reading fluency tests in the lower grades of elementary school. In these tests, the student reads as many words as they can in a passage over a given period, and their fluency score is based on how many words they read. However, these tests are currently under review due to criticism regarding their inability to assess reading comprehension effectively.

2.5 Technology-Enhanced Reading Interventions

Integrating technology into reading interventions has shown to be effective in addressing reading difficulties. Various studies have documented the positive impact of technology on reading outcomes (Karam et al., 2018; Kim and Frick, 2020; Kucirkova, 2020).

Adaptive Learning Systems: Adaptive learning systems personalize the learning experience by adjusting content and feedback based on the learner's performance. These systems use algorithms to analyze student data and provide tailored interventions. Karam et al. (2018) state that adaptive learning systems can significantly improve reading skills by providing customized support and immediate feedback. A study by Kim and Frick (2020) further explores how digital tools can be designed to enhance student motivation and engagement in reading.

Adaptive learning technologies have shown significant promise in supporting diverse learners, including those with reading disabilities. By tailoring instruction to the specific needs of each learner, these technologies can help bridge the gap for students who struggle with traditional reading instruction. For instance, Kucirkova (2020) found that adaptive learning systems that incorporate multimodal feedback (visual, auditory, and kinesthetic) are particularly effective for students with dyslexia, offering them personalized support that addresses their unique learning challenges.

Interactive Reading Applications: Interactive reading applications combine various technological elements, such as SST, TTS, and gamification, to create an engaging learning environment. These applications support active learning and provide opportunities for repeated

practice. Research by Miller and Warschauer (2014) examines young children's internet use at home and school, providing insights into how digital access affects learning opportunities. While their study does not directly address interactive reading applications, it highlights the increasing role of technology in children's lives, suggesting potential avenues for educational interventions. Studies by Castek et al. (2020) further validate the effectiveness of these applications in improving reading skills.

2.6 Emotional and Motivational Aspects of Reading

The emotional and motivational aspects of reading are crucial for sustained engagement and improvement in reading skills. Pekrun (2006) discusses the role of emotions in learning, highlighting that positive emotions such as enjoyment and confidence can enhance motivation and performance.

Role of Emotions in Learning: Pekrun (2006) argues that emotions play a significant role in learning processes. Positive emotions like enjoyment and confidence are associated with higher motivation levels and better learning outcomes. Conversely, negative emotions can hinder learning by reducing engagement and persistence. Understanding students' emotional responses during reading activities can help design interventions that foster positive emotions and enhance learning (Pekrun, 2006). Research by Immordino-Yang and Gotlieb (2017) supports Pekrun's (2006) findings, emphasizing the integral role of emotions in learning and cognitive development.

Motivational Strategies in Reading Instruction: Motivational strategies in reading instruction aim to increase students' intrinsic motivation to read. According to Guthrie and Wigfield (2000), strategies such as providing choice, using interesting texts, and integrating social interactions can boost motivation and engagement in reading. Incorporating these strategies into technology-enhanced reading interventions can create a more motivating and effective learning environment. Studies by Kim and Frick (2020) further explore how digital tools can be designed to enhance student motivation and engagement in reading.

2.7 Summary

The integration of educational technology, particularly SST and TTS, offers significant potential for improving student engagement and motivation. Gamified learning platforms provide interactive and personalized experiences that increase student participation. Understanding the role of real-time feedback and emotional engagement is key to designing effective educational interventions. This literature review underscores the importance of combining technological innovations with motivational strategies to create engaging learning environments.

3 Methodology: Design and Development of the Reading Game Application

3.1 Introduction

This chapter discusses the design and development of two key applications: the Reading Game and the Listening Game. Both games aim to enhance children's literacy skills using interactive and adaptive technologies. This chapter will analyze the motivations behind various design choices, the algorithms and methods employed, and the iterative development process. Additionally, this chapter will discuss the evolution of the initial reading game, provide a breakdown of the updated features, and provide a detailed look into the design and logic of the Listening Game.

3.2 Evolution of the Reading Game Application

3.2.1 Initial Version

The first version of the reading game application set the foundation for future development, but it was limited in scope and flexibility. It primarily relied on pre-recorded voices for reading exercises[1], making it difficult to introduce new content without significant updates.

- **Challenges in Content Scalability:** The fixed nature of pre-recorded audio limited the app's ability to expand its library. Adding new reading tasks required extensive manual updates, which reduced the app's ability to grow dynamically.
- **Limited User Interaction:** The original version lacked engaging interactive elements, such as read-out-loud games, which restricted users to passive learning. The absence of speech-to-text (STT) and text-to-speech (TTS) features restricted users from engaging in reading exercises.
- **User and Score Management:** The initial version utilized a basic Firebase system for tracking user data and scores. While functional, it lacked the flexibility and robustness needed for scalable growth and personalized learning.

Figure 3.1 illustrates the initial system architecture.

[1] It may be noted that similar games such as "Lærum og leikum með hljóðin" have been made for Icelandic (Guðmundsdóttir, 2014) accessible at https://laerumogleikum.is/

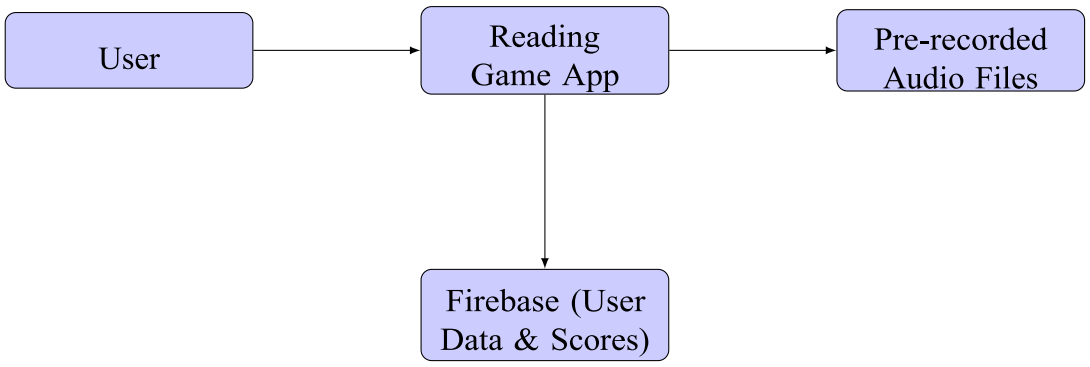

Figure 3.1: Old System Structure and Data Flow

3.2.2 Motivations for Upgrades

The motivation to enhance the reading game stemmed from the following objectives:

- **Scalability and Flexibility:** Transitioning to a dynamic content generation system where new tasks can be created and updated seamlessly without significant manual intervention.
- **Enhanced User Engagement:** Introducing features like STT, TTS, and real-time feedback to create a more interactive learning environment.
- **Improved Infrastructure:** Upgrading the backend infrastructure for better security, scalability, and performance using Amazon Amplify, GraphQL, and other modern cloud services.

3.3 Updated Features in the Reading Game

3.3.1 State Management with the BLoC Pattern

The introduction of the BLoC (Business Logic Component) pattern was a major upgrade from the previous set-state pattern. The BLoC pattern centralizes state management, ensuring efficient management of global states, especially when dealing with asynchronous operations like fetching data and processing user interactions.

- **Why BLoC?:** The BLoC pattern was chosen because it allows for a clear separation of UI and business logic. This approach makes the app's maintainability and scalability, especially as the complexity of the game logic increases.
- **Algorithm Overview:** Introducing features like STT, TTS, and real-time feedback to create a more interactive learning environment.
- **Improved Infrastructure:** The BLoC pattern operates based on streams and events. User actions trigger events that are processed in the BLoC layer, and corresponding states are emitted back to the UI. This event-driven architecture ensures that the state transitions are smooth and predictable.

As illustrated in Figure 3.2, the BLoC pattern in the Reading Game Application facilitates the flow of data and events between the serverless API, the BLoC layer, and the UI.

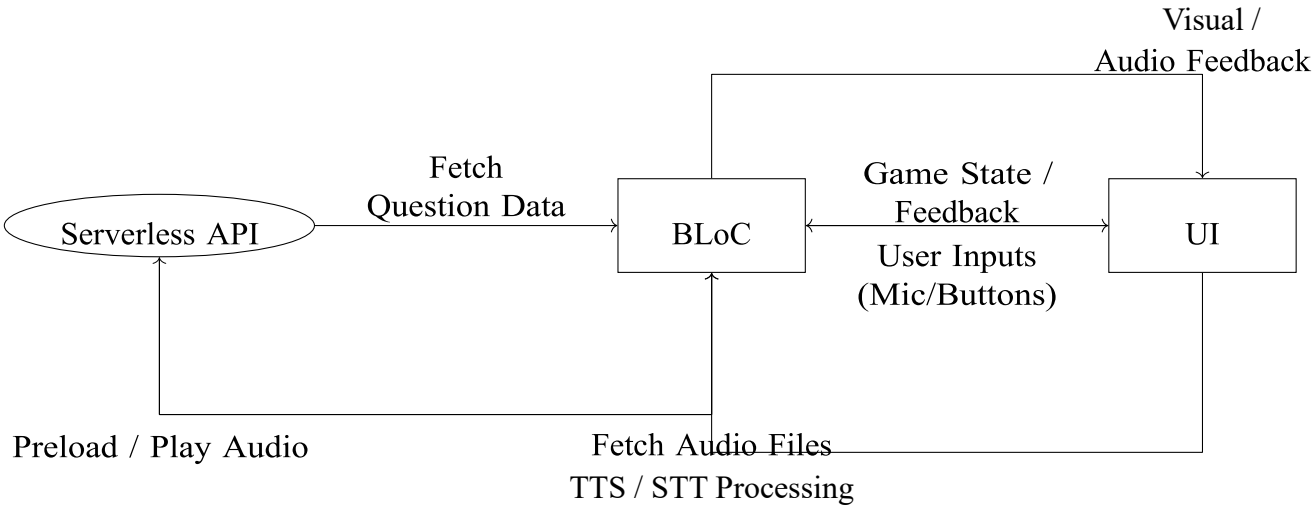

Figure 3.2: BLoC Pattern in the Reading Game Application

3.3.2 Back-end Enhancements

Authentication, Score, and User Data

The migration from Google Firebase to Amazon Amplify introduced significant security, scalability, and performance improvements.

- **Amazon Cognito for Authentication:** Cognito provides robust security features, including multi-factor authentication, password policies, and scalability to handle a growing number of users.
- **GraphQL for Data Management:** GraphQL allows for more efficient data querying and mutation, enabling real-time score updates and personalized user settings. The flexibility of GraphQL queries reduces data over-fetching, ensuring that only relevant data is retrieved.
- **Custom Score Tracking Algorithm:** A custom algorithm manages scores, dynamically updating user performance based on real-time interactions. The algorithm considers multiple factors, including accuracy, speed, and consistency, to determine the final score.

The back-end infrastructure with Amazon Amplify is depicted in Figure 3.3.

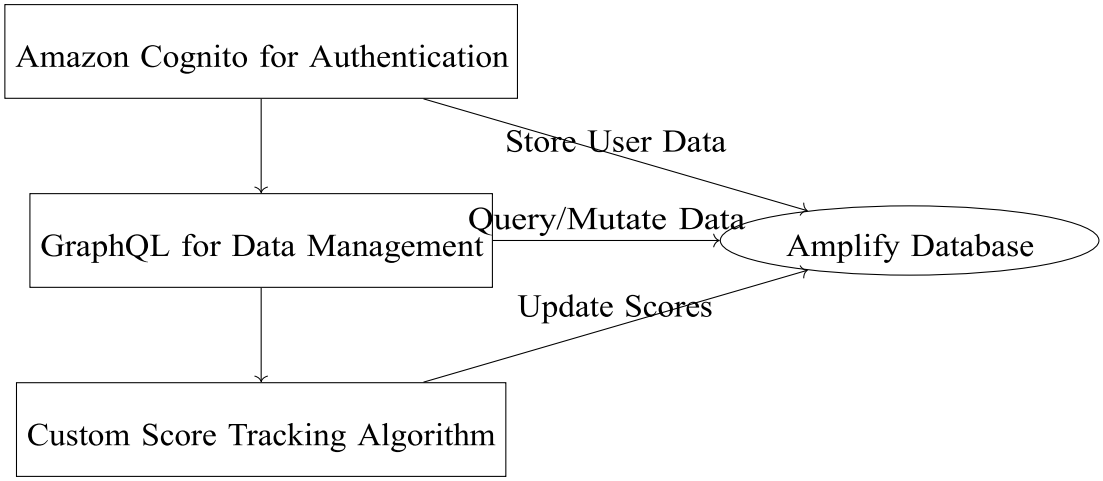

Figure 3.3: Back-end Infrastructure with Amazon Amplify

Dynamic Content Addition Through a Web Framework

A web framework was introduced to allow administrators to dynamically add new reading tasks using Amazon Polly. This framework automates the generation of TTS content, reducing the time and effort needed to update the app's content.

- **TTS Content Creation Workflow:** Admins input new content through a web form, which triggers a serverless function that generates the TTS audio using Amazon Polly. The audio is stored in Amazon S3, where it is readily accessible by the app during gameplay.
- **GraphQL for Data Management:** GraphQL allows for more efficient data querying and mutation, enabling real-time score updates and personalized user settings. The flexibility of GraphQL queries reduces data over-fetching, ensuring that only relevant data is retrieved.
- **Automation and Caching:** The generated content is cached to reduce the need for repeated fetching from S3, improving performance during game sessions.

This dynamic content addition workflow is illustrated in Figure 3.4.

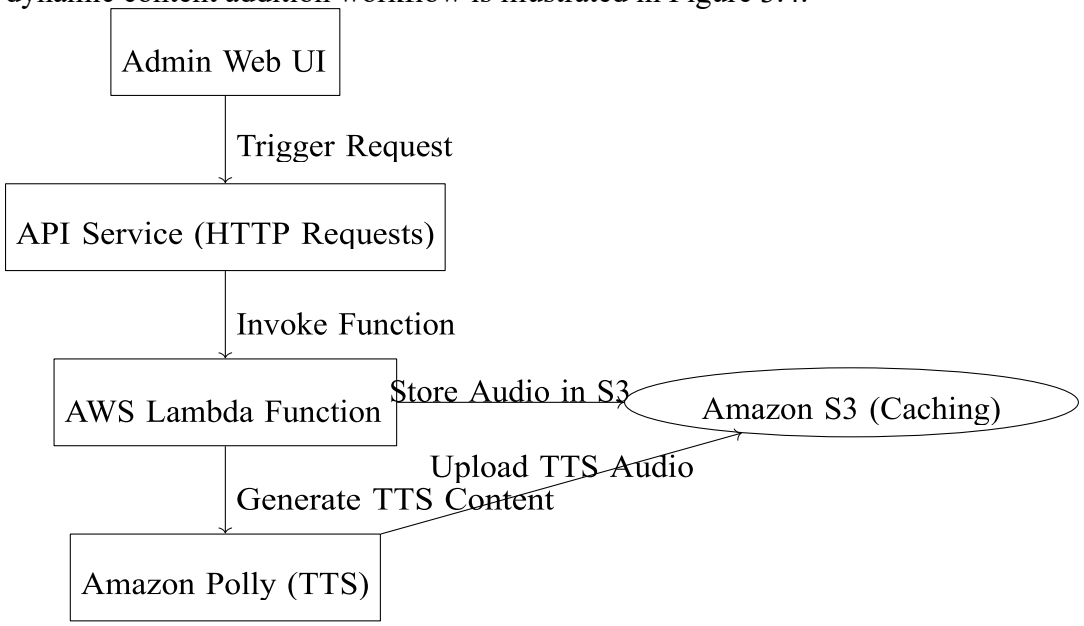

Figure 3.4: Dynamic Content Addition Workflow

3.3.3 Custom SAMPA Rules and Icelandic Letter Mappings

Mapping Icelandic Letters

The SAMPA mappings and comments presented in this thesis are grounded in previous research in Icelandic language phonetics and computational linguistics. These mappings are particularly informed by studies such as *Talrómur: A Large Icelandic TTS Corpus* (Sigurgeirsson et al., 2021), which provides detailed phonetic data for text-to-speech systems in Icelandic.

Additionally, Besacier et al. (2014a) offer a comprehensive framework for adapting phonetic systems like SAMPA to under-resourced languages, emphasizing their importance for speech recognition and language modeling. The vowel and consonant mappings presented in this study align with these foundational principles.

The implementation of these mappings leverages practical insights from phonetic transcription and corpus-based language modeling to ensure consistency with Icelandic phonological rules and speech synthesis applications. Table 3.1 below lists the Icelandic vowels and their corresponding SAMPA mappings. Table 3.2 shows detailed consonant mappings.

Table 3.1: Icelandic Vowels and Their SAMPA Mappings

Vowel	SAMPA Equivalent	Comments
a	/a/	Standard mapping
á	/au/	Diphthong /au/
e	/E/	Open-mid front vowel //
é	/e:/	Long close-mid front vowel
i	/I/	Near-close front vowel //
í	/i:/	Long close front vowel /i/
o	/O/	Open-mid back vowel //
ó	/Ou/	Diphthong /ou/
u	/Y/	Close front rounded vowel /y/
ú	/u:/	Long close back vowel /u/
y	/I/	Near-close front vowel //
ý	/i:/	Long close front vowel /i/
æ	/aE/	Diphthong /ai/
ö	/9/	Open-mid front rounded vowel
:	/:/	Indicates vowel length

Table 3.2: Icelandic Consonants and Their SAMPA Mappings

Consonant	SAMPA Equivalent	Comments
b	/b/	Standard mapping
d	/d/	Standard mapping
ð	/D/	Voiced dental fricative /ð/
f	/f/	Standard mapping
g	/G/	Voiced velar fricative //
h	/h/	Standard mapping
j	/j/	Palatal approximant
k	/k_h/	Aspirated voiceless plosive
l	/l/	Standard mapping
m	/m/	Standard mapping
n	/n/	Standard mapping
p	/p_h/	Aspirated voiceless plosive
r	/r/	Alveolar trill
s	/s/	Standard mapping
t	/d/	Mapped to /d/ for correction
v	/v/	Standard mapping
x	/x/	Voiceless velar fricative
þ	/T/	Voiceless dental fricative

Playback Examples for Word and Sentence Pronunciation

Figure 3.5 illustrates the application of SAMPA rules during TTS processing. The process involves converting input text into SAMPA transcriptions before passing it to the TTS engine, ensuring accurate pronunciation.

Table 3.3: Playback Examples for Words and Sentences

Category	Text	Key
Word	Vörubíll	"v9rYbi:dl"
Word	Blæjubíll	"blaEjYbi:dl"
Simple Sentence	Halló heimur	"hallOu heimur"
Simple Sentence	Palli hjólar	"p_hallI hjólar"
Medium Sentence	Villi og Kalli kölluðu á Silla	"vIllI og k_hallI kölluðu á Silla"
Medium Sentence	Það er halli í brekkunni	"Það er hadlI í brekkunni"

Playback of Word and Sentence with and without Key

Below are playback examples demonstrating how the pronunciation differs with and without the custom SAMPA keys, featuring Dora and Karl's voices.

Word: Vörubíll

- **Dora Playback Without Key**: Play Default
- **Dora Playback With Key**: Play Custom

- **Karl Playback Without Key**: Play Default
- **Karl Playback With Key**: Play Custom

Sentence: Halló heimur

- **Dora Playback Without Key**: Play Default
- **Dora Playback With Key**: Play Custom
- **Karl Playback Without Key**: Play Default
- **Karl Playback With Key**: Play Custom

Sentence: Palli hjólar

- **Dora Playback Without Key**: Play Default
- **Dora Playback With Key**: Play Custom
- **Karl Playback Without Key**: Play Default
- **Karl Playback With Key**: Play Custom

Sentence: Villi og Kalli kölluðu á Silla

- **Dora Playback Without Key**: Play Default
- **Dora Playback With Key**: Play Custom
- **Karl Playback Without Key**: Play Default
- **Karl Playback With Key**: Play Custom

Sentence: Villi ætlar að fara í sund

- **Dora Playback Without Key**: Play Default
- **Dora Playback With Key**: Play Custom
- **Karl Playback Without Key**: Play Default
- **Karl Playback With Key**: Play Custom

Sentence: Það er halli í brekkunni

- **Dora Playback Without Key**: Play Default
- **Dora Playback With Key**: Play Custom
- **Karl Playback Without Key**: Play Default
- **Karl Playback With Key**: Play Custom

Additionally, the system recognizes and correctly processes special tags such as <prosody> for adjusting pitch, volume, and speed and <break> for inserting pauses, ensuring that the spoken output is as natural and accurate as possible.

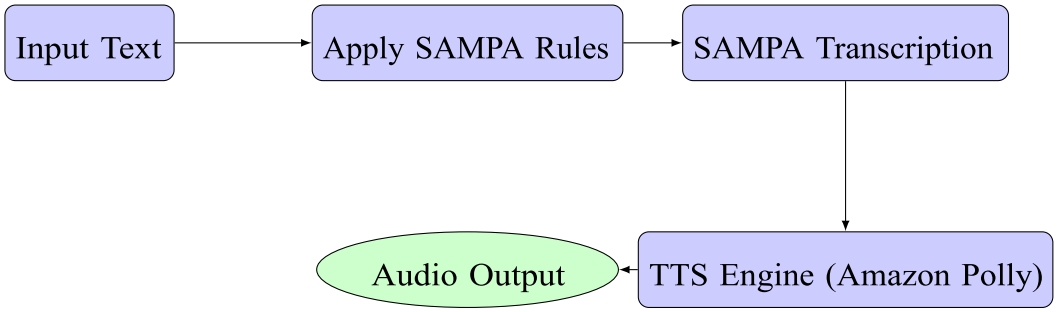

Figure 3.5: Application of SAMPA Rules during TTS Processing

Manual Adjustments in the Web Framework

To complement the automated SAMPA rules, the web framework allows for manual interventions further to refine the pronunciation and timing of the spoken content.

- **Adding SAMPA Sounds:** Administrators can manually input SAMPA transcriptions between words to correct specific pronunciation issues. This feature is particularly useful for fine-tuning the speech output for words or phrases that the automated system might not handle perfectly.
- **Inserting Pauses:** A customizable pause can be inserted between words or sounds to enhance the naturalness of the spoken content. This is especially helpful for ensuring clarity in longer phrases or sentences.

The modal interface for adding SAMPA sounds and pauses is shown in Figure 3.6.

Figure 3.6: Modal Interface for Adding SAMPA Sounds and Pauses

3.3.4 Text-to-Speech and Speech-to-Text Integration

In the process of integrating Icelandic language support, significant challenges arose due to the specific phonetic characteristics of Icelandic, which standard TTS engines like Amazon Polly struggled to accurately reproduce. To address these challenges, a comprehensive set of SAMPA (Speech Assessment Methods Phonetic Alphabet) rules was implemented.

Enhanced sTTS with Amazon Polly

The app's TTS feature leverages Amazon Polly's advanced neural voices, particularly Icelandic voices: Dóra and Karl. The integration of these voices enhances authenticity and engagement for Icelandic-speaking users.

- **Caching and Preloading Strategy:** Audio files are preloaded to minimize latency during gameplay. The caching mechanism reduces data fetch times, ensuring smooth audio transitions.
- **Serverless Architecture:** The TTS processing is handled by an AWS Lambda function, which automates the uploading and management of audio files.

As illustrated in Figure 3.7, the TTS processing workflow incorporates SAMPA integration to improve pronunciation accuracy.

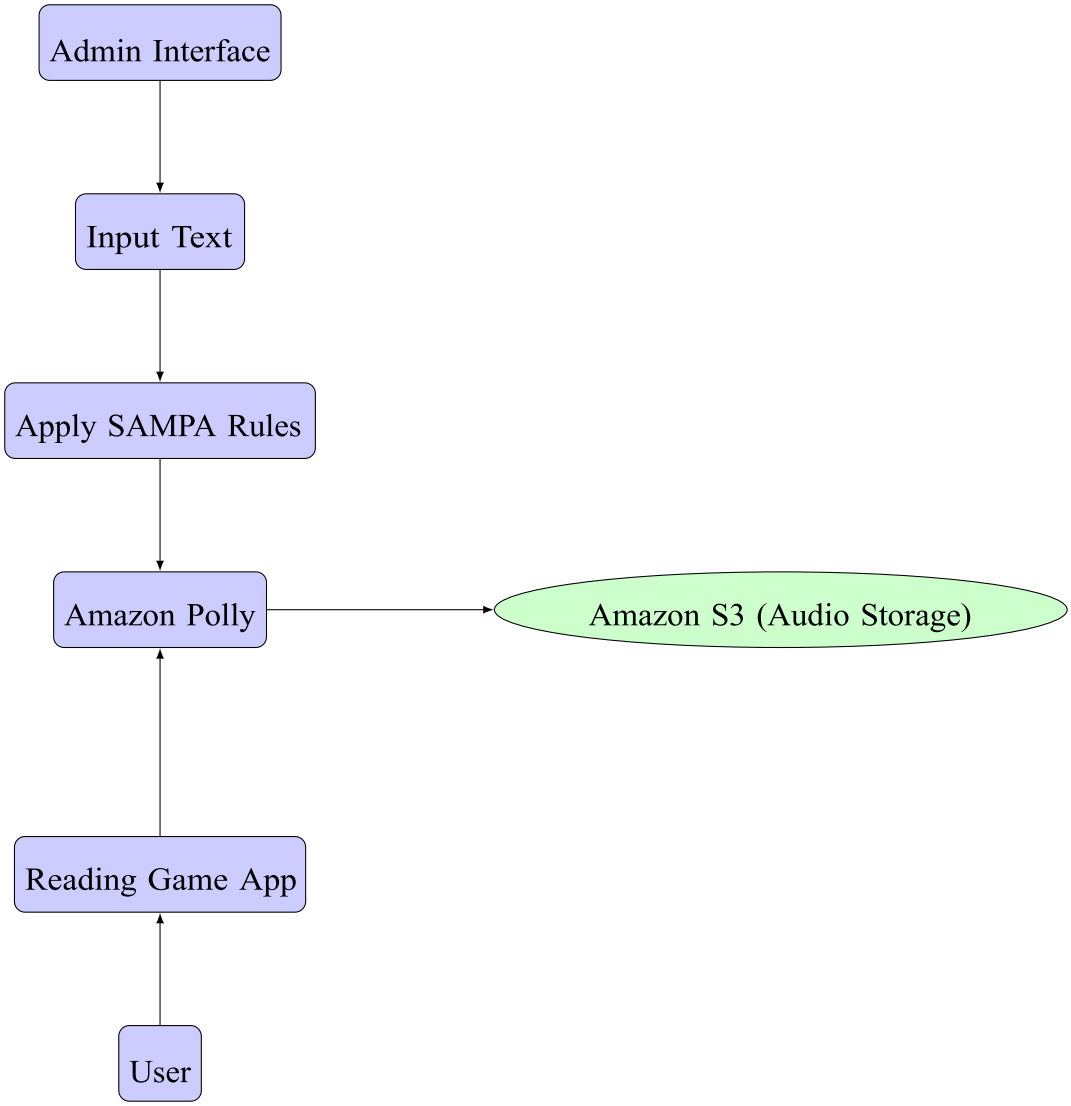

Figure 3.7: TTS Processing Workflow with SAMPA Integration

Advanced STT Functionality with Google Speech API

The STT feature is designed to provide real-time feedback and interaction. Google's Speech API was chosen for its high accuracy and low latency. The STT error correction and fallback strategy is depicted in Figure 3.8

- **Error Handling and Correction Algorithms:** The STT system is equipped with error-handling mechanisms that account for common speech recognition errors, especially in cases involving young users. The system can identify and correct errors using context-aware algorithms.

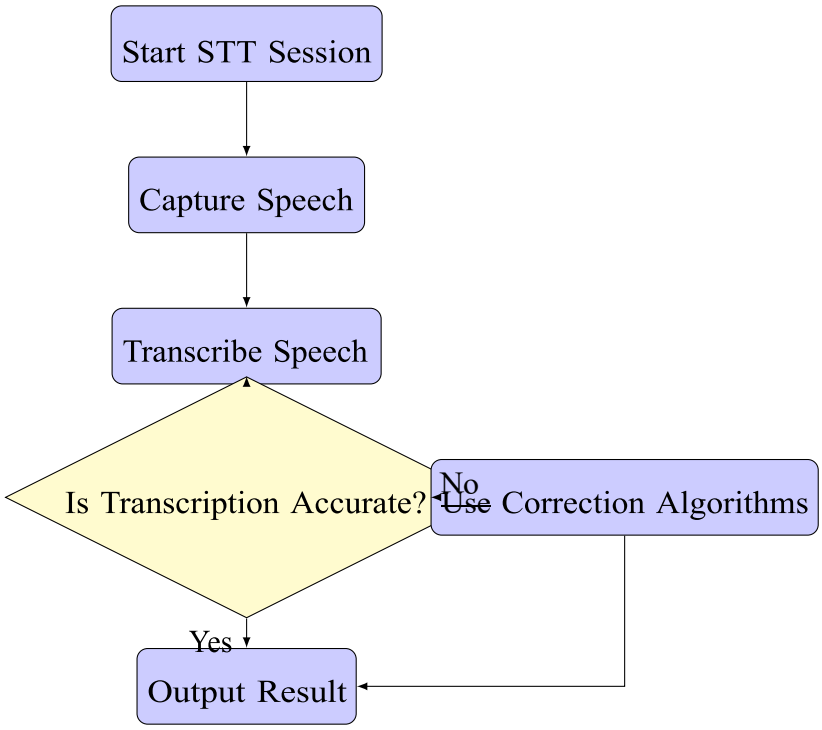

Figure 3.8: STT Error Correction and Fallback Strategy

User-Centric Settings and Personalization

The app offers various customization options to enhance the user experience, as shown in Figure 3.9.

- **AI Voice Selection:** Users can choose between different voices (e.g., Karl and Dora) for an immersive reading experience.
- **Accessibility Features:** Options such as manual error handling for children with reading disabilities ensure inclusivity.
- **Cloud-Synced Preferences:** User preferences are stored in the cloud, allowing seamless access across multiple devices.

Figure 3.9: Settings Screen

3.4 Text-to-Speech (TTS) Game Logic and Workflow

3.4.1 Game Structure Overview

The TTS game focuses on listening comprehension. The user listens to an audio prompt (such as a letter, word, or sentence) and selects the correct answer from two options displayed on the screen. The game is structured to support different levels of difficulty and dynamically adapts based on user input.

Figure 3.10 depicts the TTS game logic and workflow.

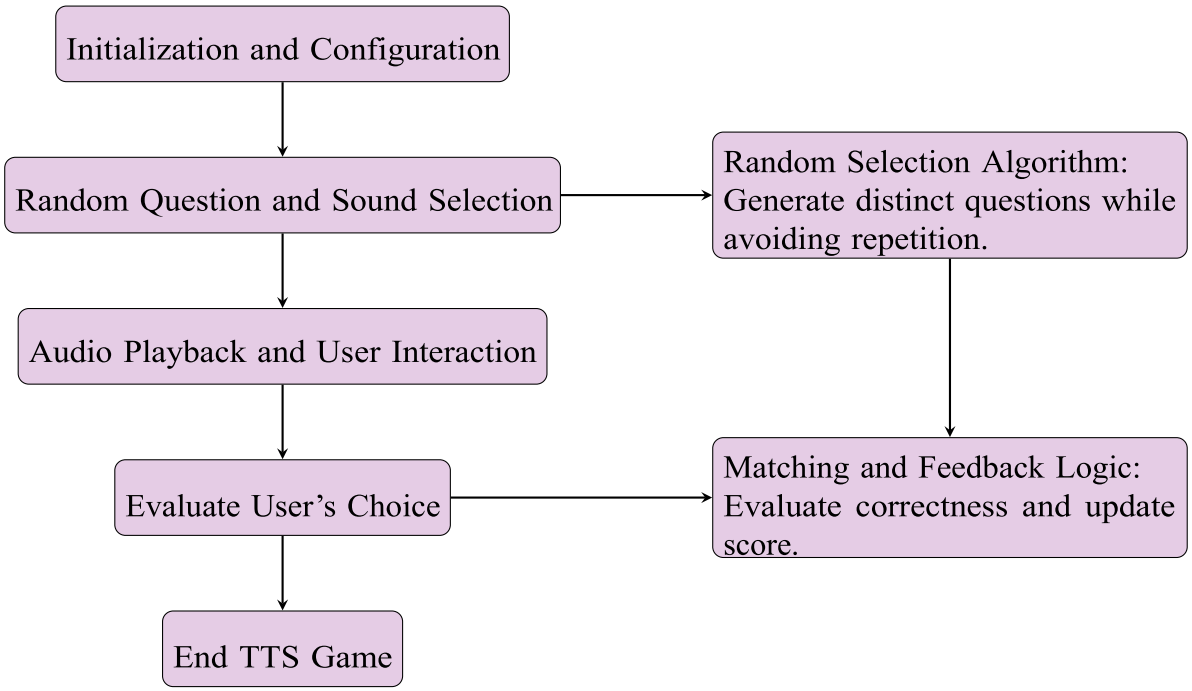

Figure 3.10: TTS Game Logic and Workflow

3.4.2 Core Mechanics

Initialization and Configuration

The game's configuration depends on the user's chosen settings, such as game type and difficulty level. During initialization, the game retrieves and prepares the relevant content for the session. This content is preloaded to ensure smooth playback and minimal latency during interactions.

Random Question and Sound Selection

Questions and their corresponding audio files are randomly selected from a preloaded question bank. The selection process ensures that each round presents a unique challenge.

Mathematically, let Q be the set of all available questions:

$$Q = \{q_1, q_2, \ldots, q_n\}$$

The game selects two distinct questions, q_i and q_j, such that $i \neq j$, ensuring that the user is always presented with different options. The corresponding audio files are also preloaded, allowing instant playback when the game starts.

Audio Playback and User Interaction

Once the answer options are displayed, the app plays one of the audio files. The user listens and selects the correct option. The system tracks which sound was played and compares it against the user's choice.

The process can be represented as:

1. Randomly choose a question and play its associated audio file.
2. Present two options on the screen, one corresponding to the played audio, for the user to select.
3. Evaluate whether the user selected the correct option.

3.4.3 Detailed Explanation of Key Algorithms

1. **Random Selection Algorithm:** The algorithm ensures that questions and their corresponding sounds are randomly selected while avoiding repetition within a session.

   Mathematically:

   • For each game iteration, generate a random index r such that $1 \leq r \leq n$ and $r \neq i$.

2. **Matching and Feedback Logic:** Once the user selects an answer, the system checks if it matches the correct answer and plays a corresponding sound based on correctness. The score is updated accordingly.

3.4.4 Sequence of Events in the TTS Game

The game follows a step-by-step sequence, as outlined below:

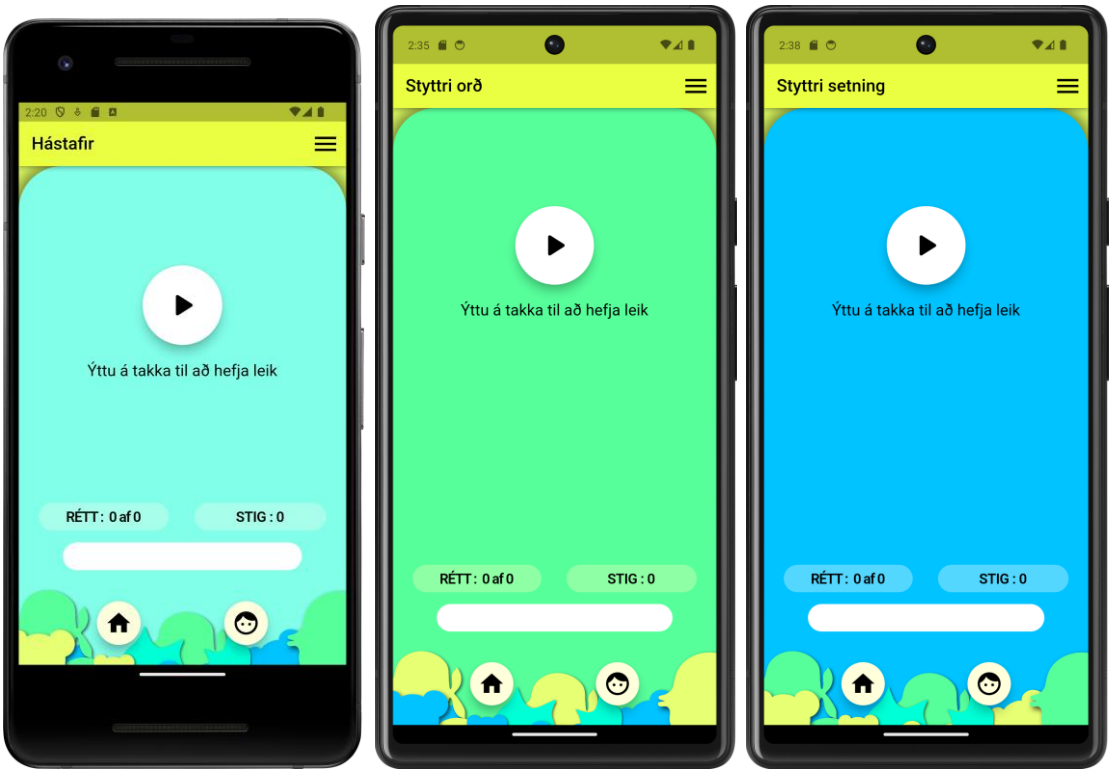

Figure 3.11: Ready State Screens for Letters, Words, and Sentences in the TTS Game

Ready State:

As shown in Figure 3.11, the game begins in a ready state, awaiting the user to initiate a session. The interface is pre-configured with the game type, difficulty level, and other user-specific settings.

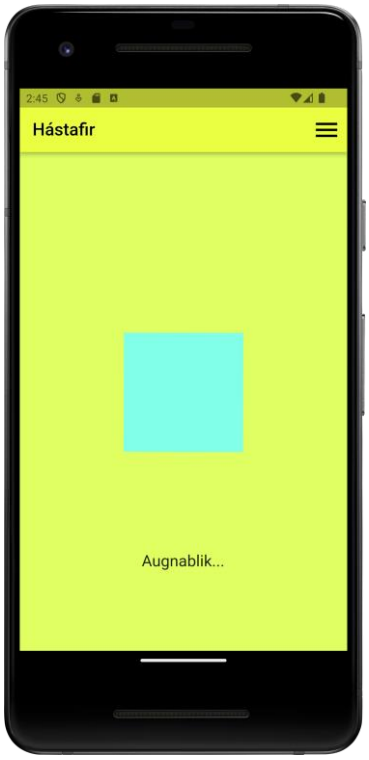

Figure 3.12: User Presses Play Button in the TTS Game

User Presses Play Button:

As depicted in Figure 3.12, the user starts the game by pressing the play button. This triggers the game's preloading and configuration process, which involves setting up all necessary resources like questions, audio files, and algorithms based on the selected settings.

Content Preloading:

The game preloads audio files (for both AI_DORA and AI_KARL) and questions. Preloading minimizes latency during gameplay, enabling smooth and instantaneous playback of sounds and transitions between questions.

Question Selection and Playback:

The system randomly selects two questions and plays one of the associated audio prompts. The user hears either AI_DORA or AI_KARL read the letter, word, or sentence. The correct audio prompt (e.g., AI_DORA) is predesignated as either the top or bottom option.

User Input and Evaluation:

As shown in Figure 3.13, the user can interact with the game in several ways:

> **Pressing the Lady Icon (AI_DORA):** Replays the letter, word, or sentence with AI_DORA's voice.

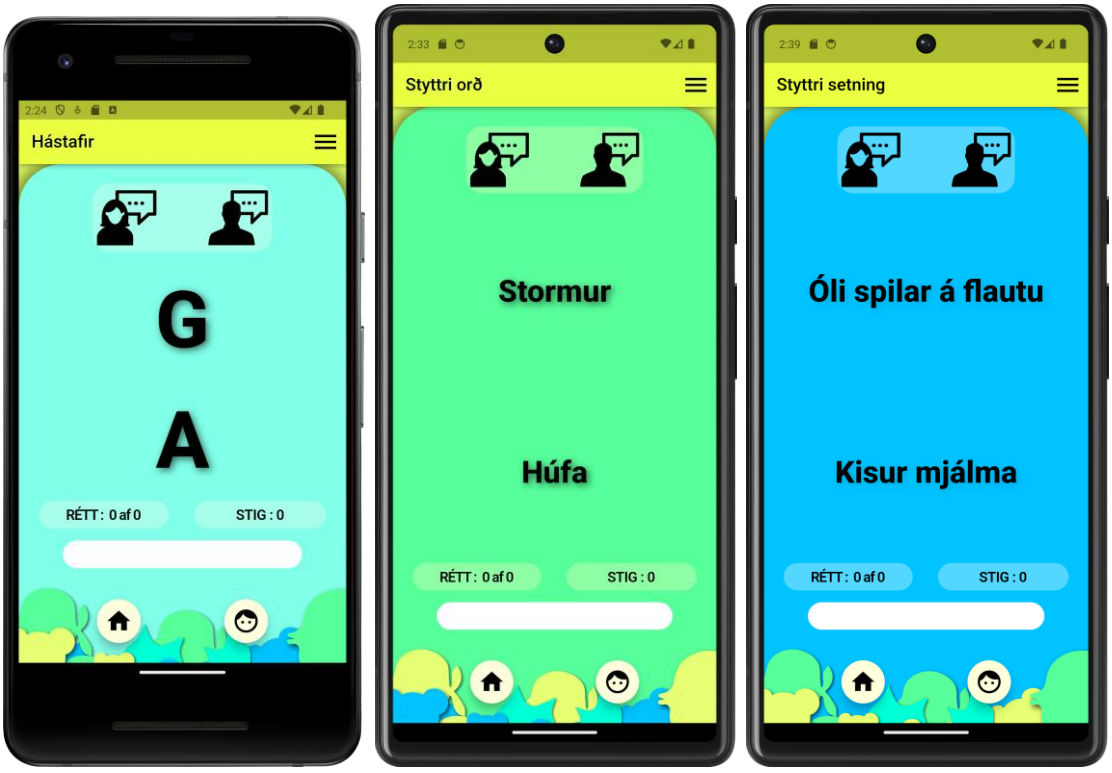

Figure 3.13: Question Selection and Playback in the TTS Game

- **Pressing the Man Icon (AI_KARL):** Replays the letter, word, or sentence with AI_KARL's voice.
- **Pressing the Top Option:** Selects the top letter, word, or sentence as the answer.
- **Pressing the Bottom Option:** Selects the bottom letter, word, or sentence as the answer.

The user's choice is immediately evaluated to determine if it is correct.

Feedback and Scoring:

If the Correct Option is Pressed: The app plays a positive feedback sound (e.g., a correct sound). The user earns one star, which is added to their total score.

Figure 3.14 depicts the correct answer.

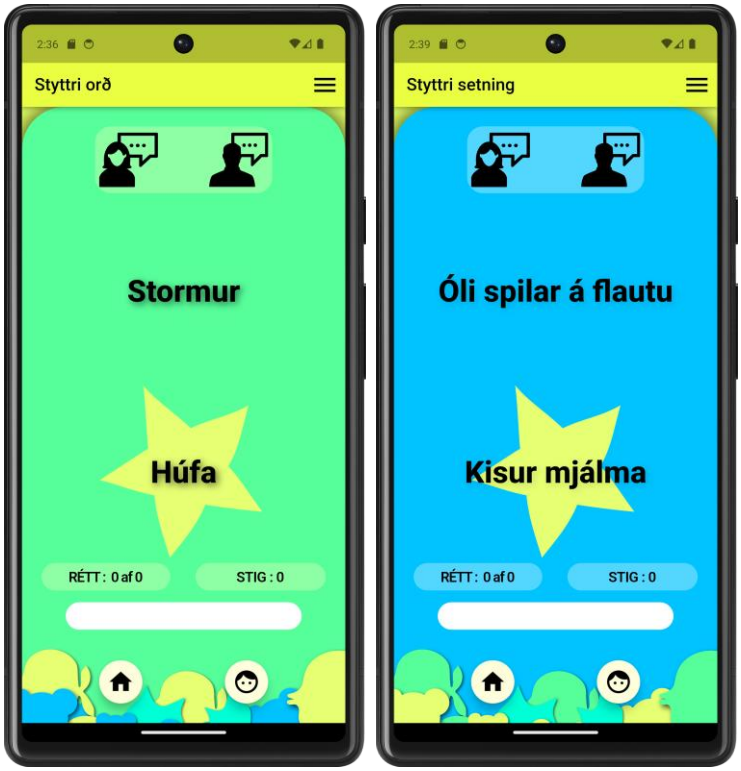

Figure 3.14: Question Answered Correctly in the STT Game

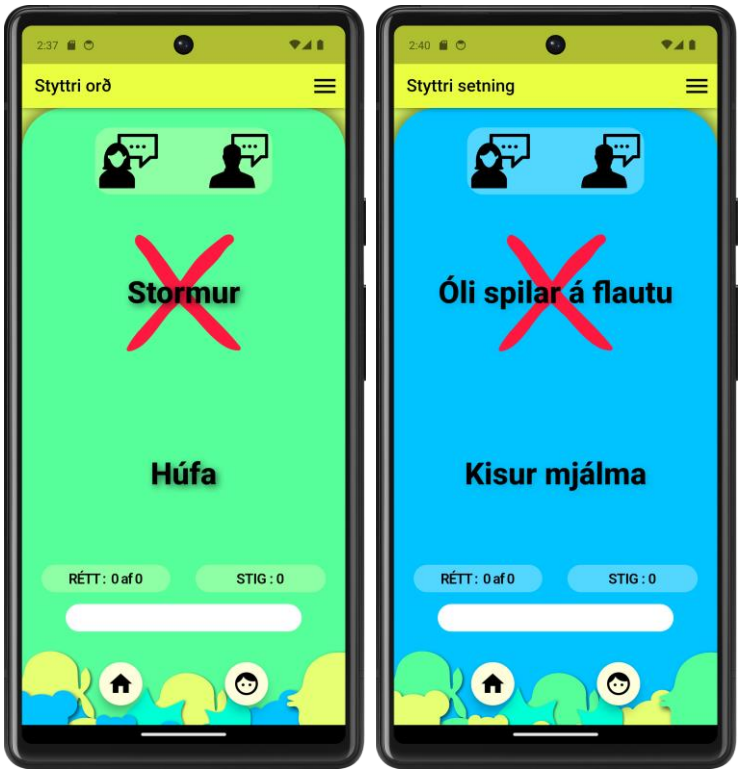

Figure 3.15: Question Answered Incorrectly in the STT Game

If the Incorrect Option is Pressed: The app plays a negative feedback sound (e.g., an incorrect sound). The user loses one star, reducing their score.

Figure 3.15 depicts the incorrect answer.

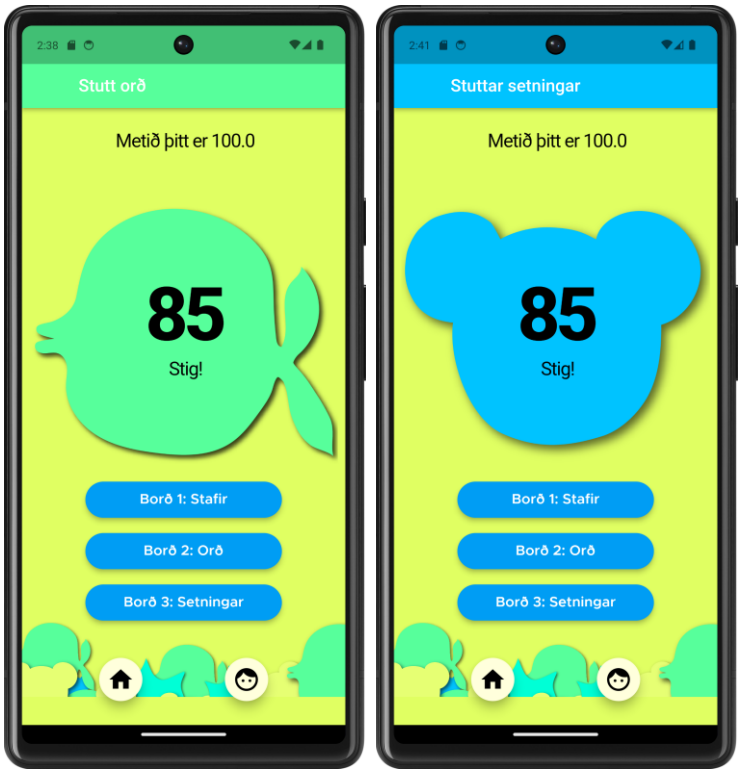

Figure 3.16: TTS Question Game Finished

Check for Completion (10 Stars):

After each round, the game checks if the user has earned a total of 10 stars:

- If the user has 10 stars, the game ends, and the final score is displayed, as shown in Figure 3.16
- If the user does not have 10 stars, the game returns to the Question Selection and Playback phase, where a new question is selected and played. Figure 3.17 presents a representation of the whole TTS workflow.

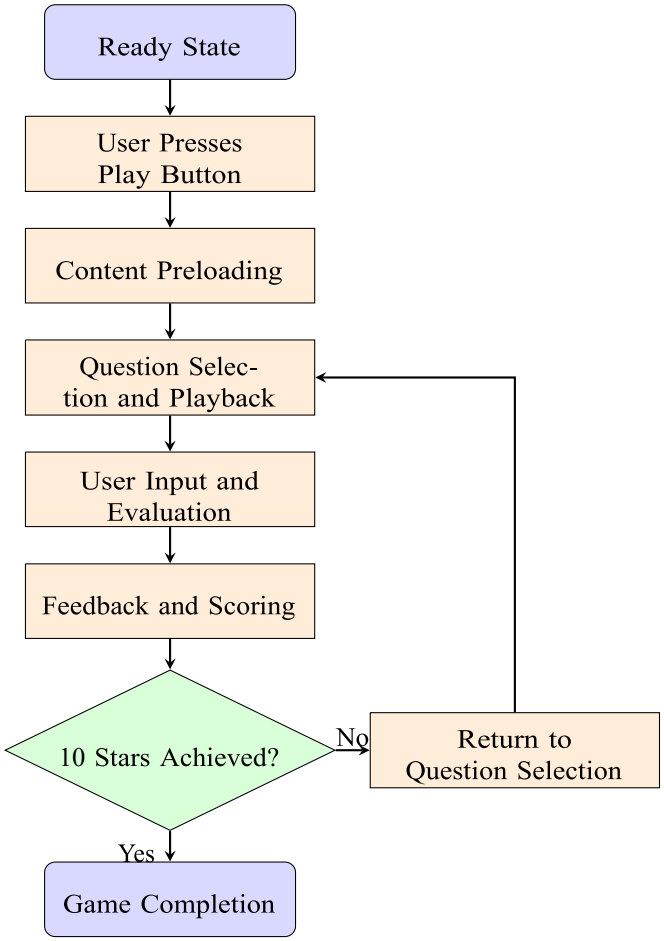

Figure 3.17: TTS Game Logic and Workflow

3.5 Speech-to-Text (STT) Game Logic and Workflow

3.5.1 Overview

The STT game aims to help users improve their reading and pronunciation skills. Users speak a word or sentence into the app, and the system evaluates their speech in real-time against the expected answer. The game design incorporates real-time speech recognition, comparison algorithms, scoring mechanisms, and feedback loops. The system handles different levels of difficulty, from recognizing letters to full sentences.

Figure 3.18 illustrates the workflow for STT and evaluation:

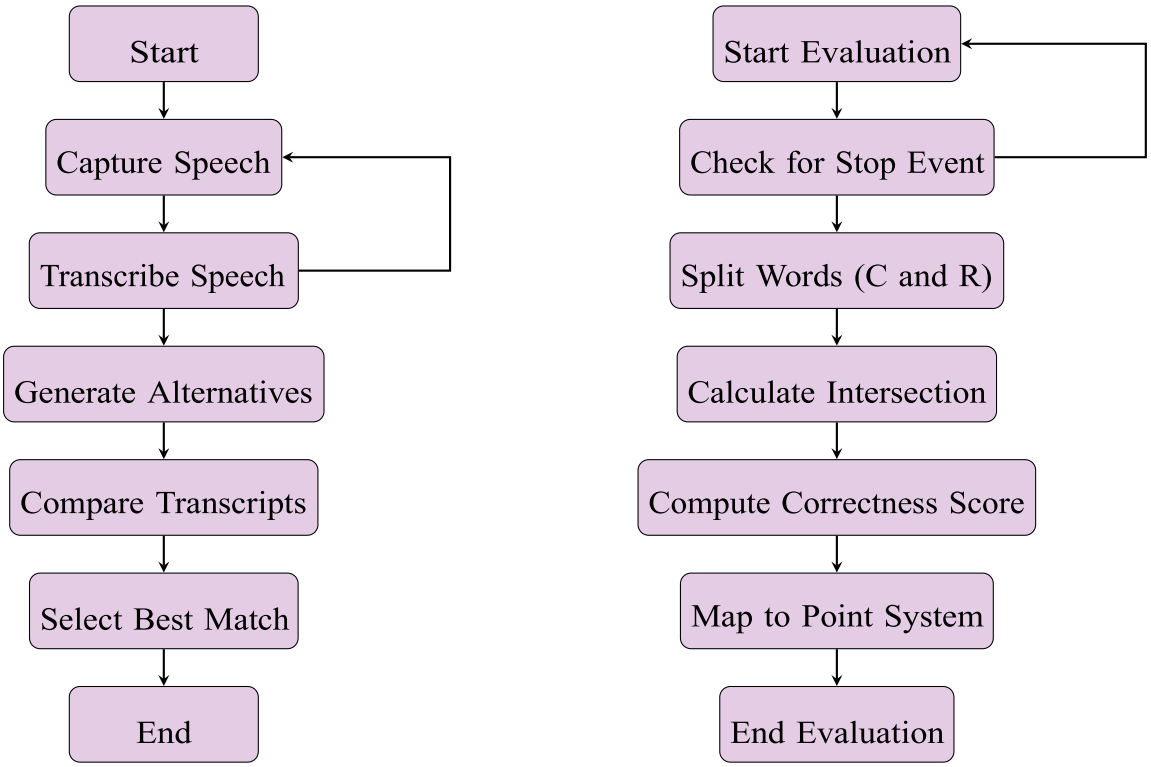

Figure 3.18: The Speech-to-Text and Evaluation Workflows for the Reading Game Application.

3.5.2 Core Mechanics

Capturing Real-Time Speech

The game uses Google's Speech API to capture and transcribe speech input in real-time. When the user starts speaking, the system continuously listens to the audio and provides a stream of recognized words. These recognized words are then compared to the correct answer.

The transcription process is represented as follows:

$$\text{Transcript} = \{W_1, W_2, \ldots, W_n\}$$

where each W_i represents a word in the user's speech.

Word Matching with the [bestLastWord] Algorithm

The primary goal is to determine how close the user's spoken words are to the expected answer. The [bestLastWord] algorithm compares the transcribed words (including alternatives) against the correct answer to find the best match.

Mathematically, we represent this process as follows:

1. Let Q be the correct answer.
2. Let $T = \{T_1, T_2, \ldots, T_k\}$ be the set of alternative transcripts returned by the API.
3. For each transcript T_i, calculate the difference $d(Q, T_i)$ using a comparison function f.

The difference d is determined by:

$$d(Q, T_i) = \sum_{j=1}^{m} |Q_j - T_{ij}|$$

where Q_j and T_{ij} are corresponding character codes in the correct answer and the transcript, respectively.

The algorithm selects the transcript with the smallest difference:

$$\text{BestMatch} = \min_i \{d(Q, T_i)\}$$

This ensures that even minor recognition errors can be corrected by selecting the best match from the alternatives.

Final Evaluation with the *[doneListener]*

Once the user stops speaking, the evaluation phase begins. The recognized words are compared against the correct answer. The comparison is more complex for longer phrases, where partial correctness is considered. The scoring process considers the number of matching words and their sequence.

The correctness score S is calculated as follows:

$$S = \frac{|C \cap R|}{|C|}$$

- C is the correct answer split into words.
- R is the recognized answer split into words.

Where $|C \cap R|$ represents the number of words correctly matched between the correct and recognized answers. The score is then mapped to a point system, allowing for partial correctness.

3.5.3 Post-Processing and Scoring

The scoring system is designed to be flexible, taking into account the user's accuracy at varying levels of difficulty. The evaluation method adjusts based on whether the task involves recognizing letters, words, or sentences.

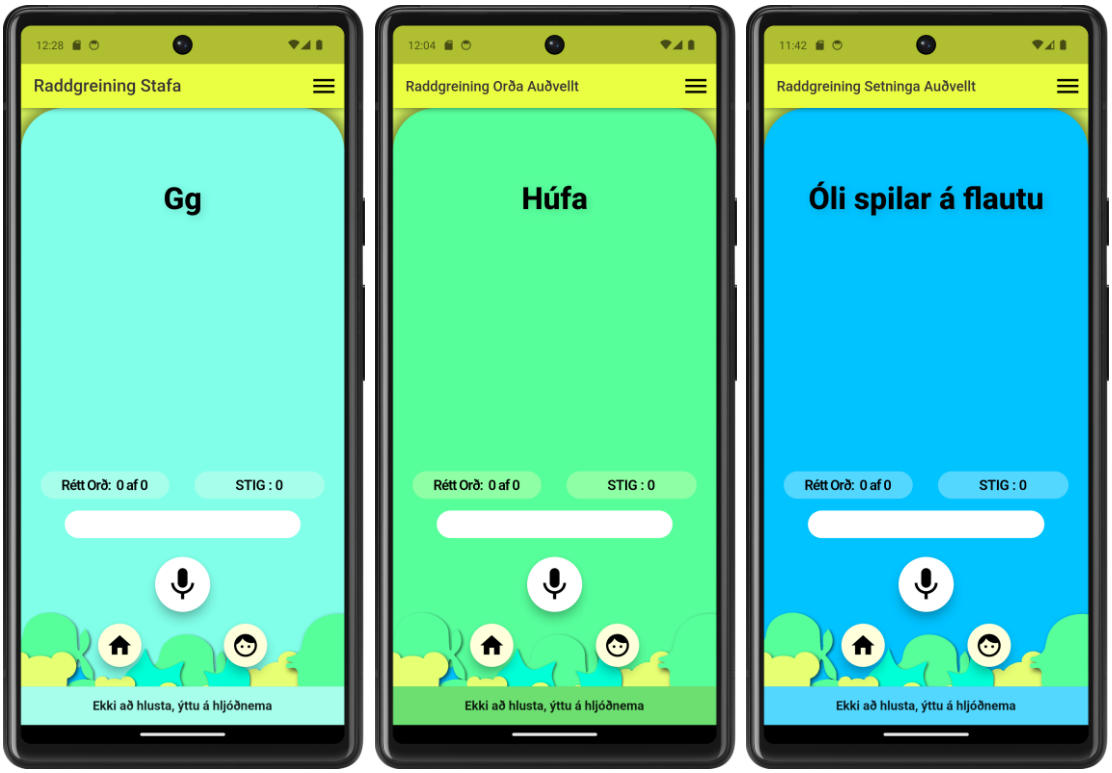

Figure 3.19: STT Ready State

Level-Specific Logic

For simpler levels, like recognizing individual letters, the evaluation focuses on the first letters match. For more complex levels, such as sentences, the evaluation considers word order and correctness.

Manual Corrections and Result Saving

If the automated system detects an error, the app allows for manual corrections. This is especially useful for edge cases where speech recognition might misinterpret the user's input. Manually corrected results are flagged and saved for future analysis.

Sequence of Events in the STT Game

The Speech-to-Text (STT) game follows a detailed sequence of actions from start to finish. Below is a step-by-step breakdown of the game flow:

Ready State

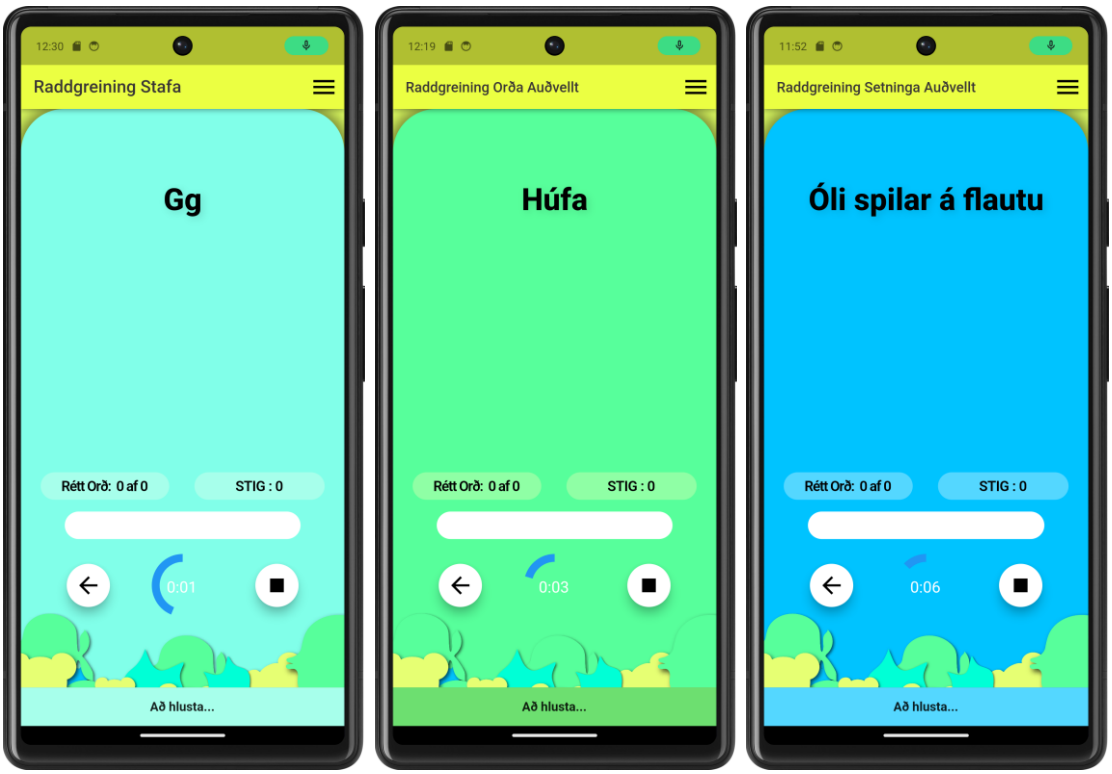

Figure 3.20: STT Listening State

The game begins in a ready state, awaiting user input. The game is preconfigured based on the selected settings (e.g., game type, difficulty). In this state, all required resources like questions, audio files, and algorithms are initialized and preloaded.

Figure 3.19 shows the game's ready state.

Description of Buttons:

- **Microphone Button:** Tapping this button starts the speech recognition process.
- **Home Button:** Allows the user to quit the game and return to the home screen.

- **Profile Button:** Allows the user to access the profile page and view high scores.

User Presses the Mic Button

The user initiates speech recognition by pressing the microphone button, as shown in Figure 3.20. This action initiates speech capture, enabling the app to begin listening to the user's input.

Description of Buttons:

- **Stop Button:** Allows the user to stop recording at any time.
- **Retry Button:** Appears if no input is detected, enabling the user to try again.

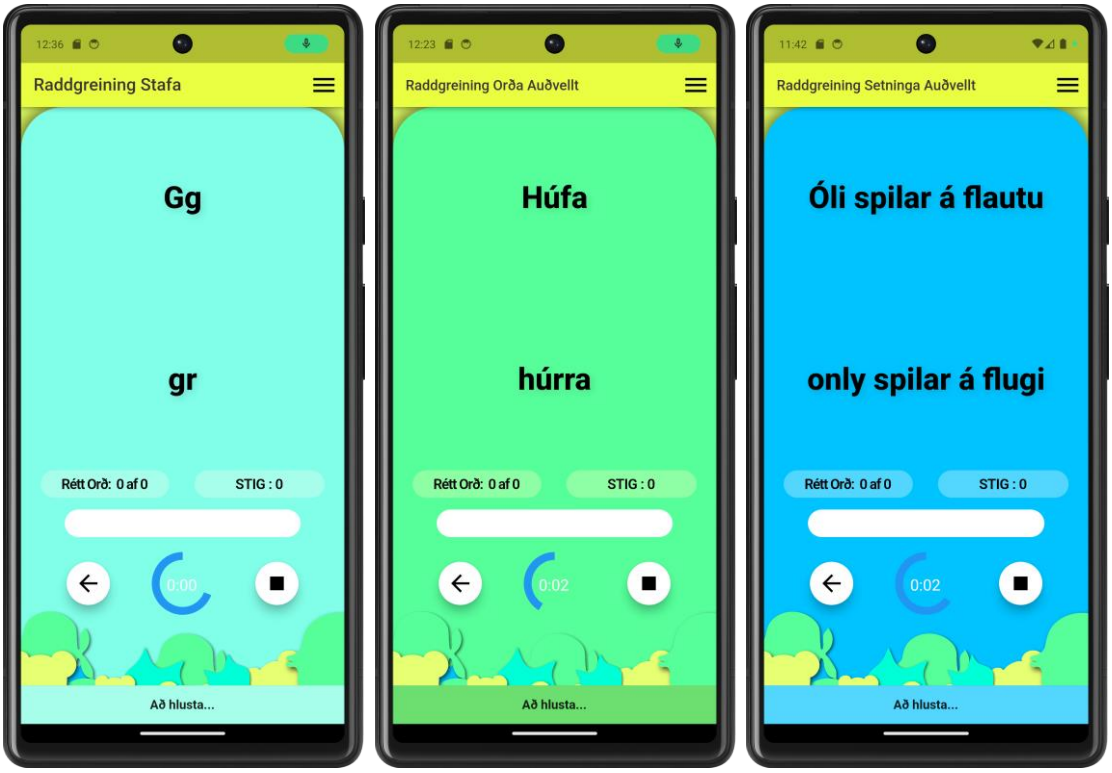

Figure 3.21: STT Conditional Check State

Start Speech Recognition and Real-Time Word Matching

The app begins capturing and transcribing the user's speech using the Google Speech API. The recognized words are streamed back in real-time for processing. As the user speaks, the recognized words are continuously compared to the expected answer using the bestLastWord method. The algorithm compares the primary and alternative transcripts from the Google Speech API to select the best match.

The word matching is represented in Figure 3.21 and Figure 3.22

Figure 3.22: STT Progress State

Conditional Checks

The app monitors the user's behavior, such as pauses, timer expiration, or if no valid words have been recognized. If these conditions are met and no meaningful input is captured, the game returns to the ready state, allowing the user to retry. Otherwise, if the user stops speaking, the timer expires, or the stop button is pressed, the app moves to the next stage.

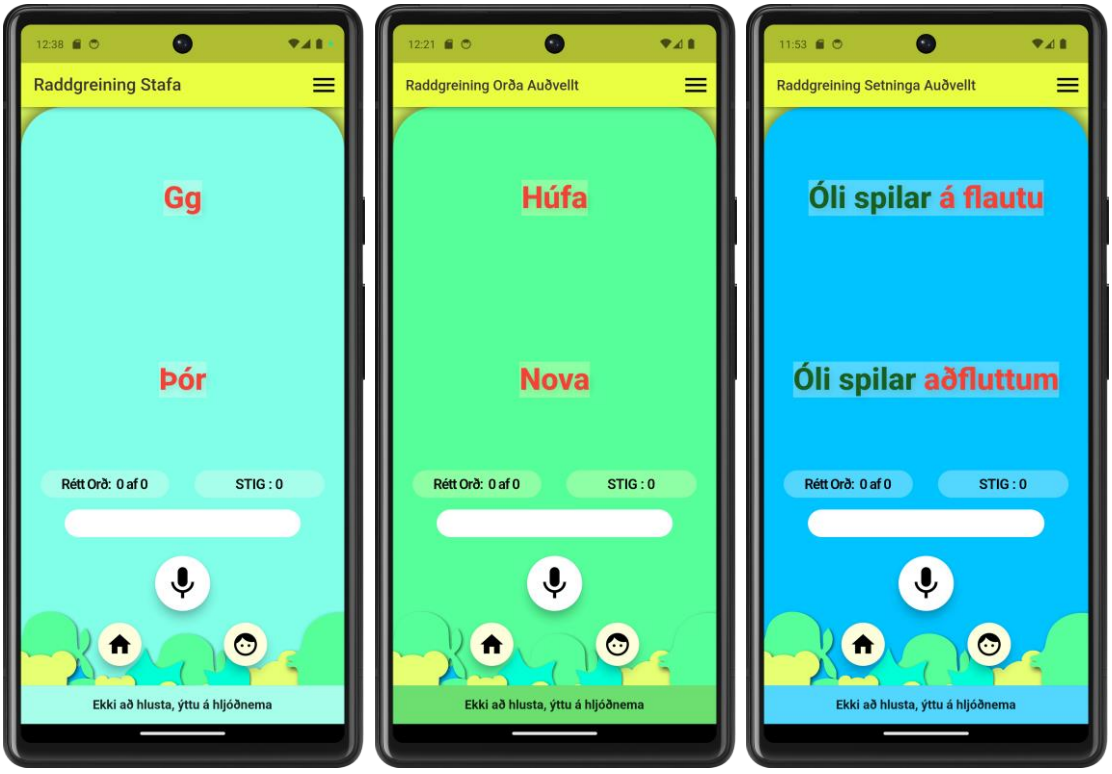

Figure 3.23: STT Transcript Done

Process Final Speech Result and Compare with Expected Answer

Once the speech session ends, the app finalizes the recognition process. The doneListener is triggered, which processes the final recognized words and prepares them for evaluation. The recognized words are then evaluated against the correct answer using the isCorrect method. The evaluation logic is tailored based on the specific game type:

- **Letter recognition**: the focus is on accurately matching individual characters.
- **Words and sentences**: the evaluation considers word order, word frequency, and partial correctness by comparing each word to the corresponding parts of the expected answer.

Figures 3.23 and 3.24 illustrate the representation of incorrect or partly correct answers.

In this system, the AI and custom algorithms described above select the most similar option regardless of confidence level.

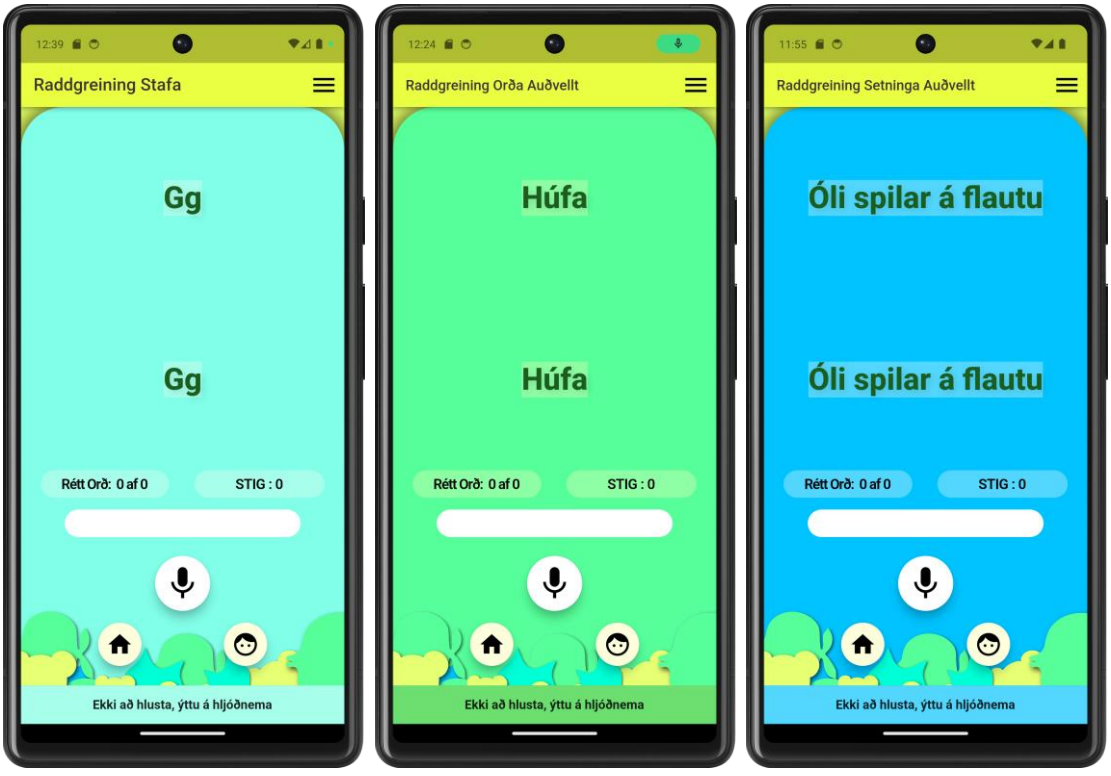

Figure 3.24: STT Correct

Assign Points and Play Feedback Sounds

Points are assigned according to the scoring system depending on whether the user's answer is correct. Feedback sounds are played to reflect the outcome of positive sounds for correct answers and negative sounds for incorrect answers. The scoring considers factors like accuracy, number of correct words, and overall correctness.

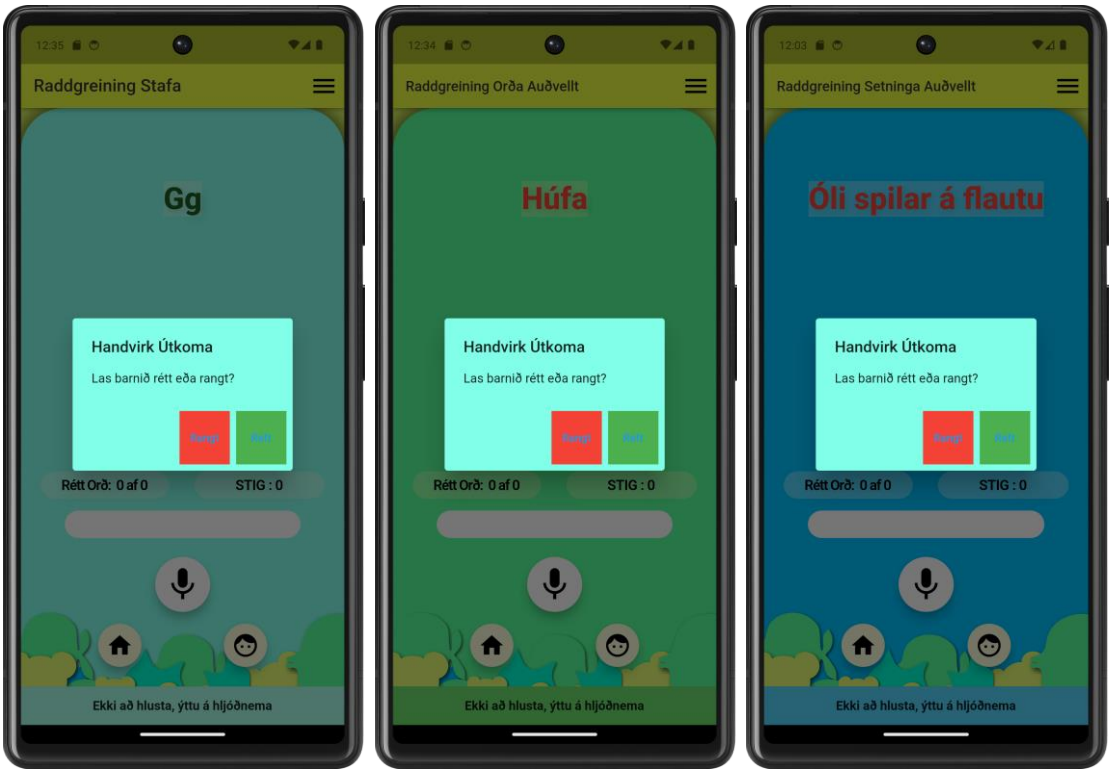

Figure 3.25: STT Manual Fix, only works if it is toggled on in the settings screen

Manual Correction (if enabled)

If manual correction is enabled, the user (or an administrator) can override the automatic evaluation. For example, if the app incorrectly marks a user's correct answer as wrong, the correction updates the score and the classification (e.g., Manual_Correct).

Figure 3.25 illustrates the representation of manual correction.

Figure 3.26: STT Game Done

Save Result (if enabled)

The final results, including recognized words, the correct answer, points, and any relevant metadata (e.g., correction type, alternative transcripts), are saved for further analysis. These results can be used to track user progress, gain insights into the game's effectiveness, and improve future iterations of the speech recognition algorithm.

Check if 10 Questions are Answered

The game checks whether the user has answered 10 questions. If 10 questions have been completed, the game transitions to the finish state. If fewer than 10 questions have been answered, the app returns to the ready state and prepares for the next question.

Finish State

Once all questions are answered, the game session ends, and the results are displayed as shown in Figure 3.26. Depending on the game configuration, the app either transitions to a results screen showing the user's score and feedback or saves the session results.

Figure 3.27 presents a visual representation of the whole STT workflow.

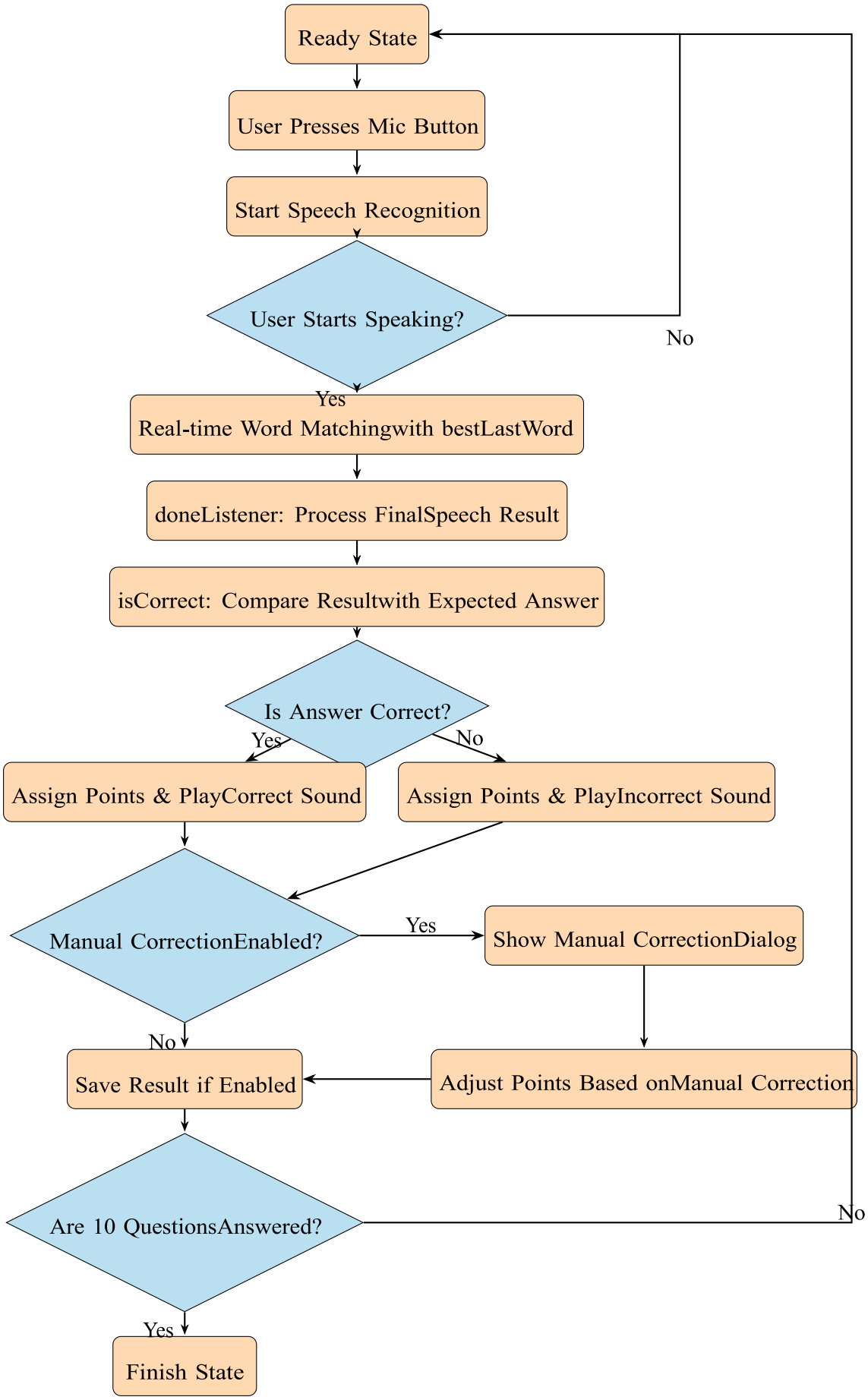

Figure 3.27: The Speech-to-Text and Evaluation Workflows for the Reading Game Application

3.5.4 Justification for Technology Choices

Selection of Google Speech API for STT

When the development of *Lesapp* commenced over three years ago, the options for Icelandic STT were extremely limited. The Google Speech API (Google Cloud, 2023) was chosen based on several factors:

Availability of Icelandic Language Support: At the time, Google was among the few providers offering Icelandic STT services, which was essential for our application's core functionality.

Existing Libraries and Documentation: Google provided developer-friendly libraries and extensive documentation, which facilitated the integration process with our application's framework. This was crucial given the project's limited resources and the need for efficient development cycles.

Accuracy and Performance: Preliminary testing indicated that Google's STT offered reasonable accuracy for Icelandic speech recognition, especially compared to other available technologies at the time. While not perfect, it was sufficient for the application's initial requirements.

Alternative STT solutions for the Icelandic language were either unavailable or lacked the required accuracy and developer support. For example, open-source STT engines such as Mozilla's DeepSpeech (Mozilla, 2023) lacked trained models for Icelandic, and developing custom models was beyond the scope and resources of this project.

Selection of Amazon Polly for TTS

For the TTS component, Amazon Polly (Amazon Web Services, 2023) was selected due to:

High-Quality Icelandic Voices: Amazon Polly offered natural-sounding Icelandic voices, namely *Karl* and *Dóra*, which were essential for creating an engaging and authentic user experience for children. The quality of these voices was superior to other options available at the time.

Scalability and Reliability: Amazon's infrastructure ensured that the TTS services were scalable and capable of meeting the application's performance requirements, including real-time feedback and simultaneous user interactions.

Integration Capabilities: The availability of APIs and comprehensive documentation facilitated the seamless integration of Amazon Polly into our existing application architecture with minimal overhead.

Initially, we planned to use the TTS solution from Tiro (Tiro, 2023), a local provider specializing in Icelandic language technologies. However, during testing, we encountered several issues:

Performance Limitations: Tiro's TTS did not meet performance expectations in terms of responsiveness and naturalness of speech, which are critical for maintaining user engagement in a reading application for children.

Integration Challenges: Difficulties arose in seamlessly integrating Tiro's solution into our application's workflow, partly due to less comprehensive documentation and support compared to Amazon's services.

Given these challenges, we decided to proceed with Amazon Polly to ensure the application met its user experience and technical reliability objectives.

3.5.5 Challenges and Limitations

While the selected technologies enabled us to develop the core functionalities of *Lesapp*, we encountered several challenges:

Recognition Accuracy with Children's Speech: The Google Speech API sometimes struggled to accurately recognize children's speech patterns in Icelandic, affecting the effectiveness of the STT features. Children's pronunciation can vary significantly, and the lack of specialized models for young Icelandic speakers posed a limitation.

Customization of TTS Pronunciation: Although Amazon Polly provided high-quality voices, certain Icelandic words and phonemes required custom handling. We implemented custom SAMPA rules (as detailed in subsection 3.3.3) to improve pronunciation accuracy, which added complexity to the development process.

4 Methodology: Research Design and Data Collection

4.1 Research Design

This study adopts a mixed-methods approach, integrating both qualitative and quantitative data to comprehensively evaluate the effectiveness of integrating Speech-to-Text (STT) and Text-to-Speech (TTS) technologies within a gamified reading app for Icelandic children.

The qualitative data, comprising video recordings and annotated comments, provide insights into the participants' experiences, perceptions, and emotional responses during their interaction with the app. The quantitative data consists of performance scores from the reading comprehension games, detailed logs of participant actions during the game, and sentiment indicators derived from the video recordings.

4.1.1 Participants

Nine students participated in the study, with ages ranging from 5 to 9 years. The participants were selected using a convenience sampling method through family relationships and acquaintances, focusing on students who were available and willing to participate. While not random, this approach allows for an initial exploration of the app's effectiveness across various age groups and reading abilities. The participants' connections to the researcher may introduce potential biases, as elaborated in the Limitations section.

4.1.2 Ethical Considerations

Prior to the study, informed consent was obtained from the parents or guardians of the participants. Participants were assured of confidentiality and informed of their right to withdraw from the study at any time without any consequences. All collected data were anonymized, and participants were assigned identifiers to protect their identities.

4.2 Interaction Process and Task Flow

Each session with a participant followed a structured process guided by a facilitator who is a primary school teacher. This ensured consistency and enabled the collection of comprehensive data on user interactions. The facilitator dynamically adapted the selection of tasks based on the participant's performance and comfort level, providing a tailored experience.

The session began with a brief interaction between the facilitator and the student to establish rapport and ease the participant into the study. This initial engagement was crucial for putting

the student at ease and establishing a positive tone for the upcoming tasks. The tasks were progressively increased in complexity, divided into the following categories:

4.2.1 Listening Tasks (Non-Voice):

- **Letter Comprehension:** Participants were presented with uppercase and lowercase letters. The app's Text-to-Speech (TTS) system vocalized a letter, and the participant selected the correct letter from two options.
- **Word Comprehension:** Similar to the letter task but using words of varying difficulty levels (easy and medium) instead of letters.
- **Sentence Comprehension:** Participants listened to sentences and selected the correct one from two options.

4.2.2 Voice-Based Tasks (Speech):

- **Letter Pronunciation:** Participants read aloud letters displayed on the screen, and the app's Speech-to-Text (STT) system evaluated their responses.
- **Word Reading:** Participants read words aloud, and the app provided feedback based on STT analysis.
- **Sentence Reading:** Participants read sentences aloud, and the app assessed their reading accuracy.

4.2.3 Selection of Words and Sentences

The words and sentences were extracted from a book designed to teach children how to read in the early grades of primary school. Words and sentences were categorized into two levels of difficulty: **easy** and **medium,** based on the reading level of the book from which they were taken. This categorization allowed for a gradual increase in complexity, helping participants build confidence with simpler tasks before progressing to more challenging ones.

4.3 Data Collection Methods

The collected data comprised four main components: performance scores, sentiment indicators, comments, and video recordings. Each component provided a different perspective on the participants' interaction with the app.

4.3.1 Performance Scores

Performance scores represent the quantitative outcomes of the participants' interactions with the app. These scores were used to evaluate user performance and included:

- Correct and incorrect responses.

- Validity of responses (e.g., whether the response was registered correctly by the app or affected by technical issues).

4.3.2 Sentiment Indicators

To capture sentiment data, we developed a custom observational coding scheme to accurately reflect participants' unique expressions. Participant responses were categorized as positive or negative, a common approach when performing sentiment analysis. While this method introduces some subjectivity, we applied consistent criteria to minimize bias, with a trained annotator reviewing the videos to assign sentiment scores. This approach allowed for a culturally relevant interpretation of the children's emotions, ensuring subtle expressions were captured and providing insights into user satisfaction and areas for app improvement.

To quantify sentiments, the following observational coding scheme was developed:

- **Positive Indicators:** Smiling, laughing, expressing verbal satisfaction (e.g., "This is fun," "I like this"), enthusiastic gestures, and sustained engagement.

- **Negative Indicators:** Frowning, sighing, expressing verbal frustration (e.g., "This is confusing," "I don't understand"), dismissive gestures, and signs of disengagement.

Specific emotional indicators were defined and coded as follows:

- **Frustration:** Characterized by verbal protests, repeated clicking behaviors, agitated movements, or vocal outbursts in response to system delays or unexpected feedback. Frustration was labeled as a negative emotional indicator.

- **Engagement:** Demonstrated through sustained attention, active participation, verbal and non-verbal responses to content, and voluntary interaction with learning materials. Engagement was labeled as a positive emotional indicator.

- **Confidence:** Exhibited through assertive responses, upright posture, willingness to attempt challenges, and positive self-commentary. Confidence was labeled as a positive emotional indicator.

- **Success Satisfaction:** Evidenced by smiles, positive verbal expressions, seeking approval, and animated reactions to achievements. Success satisfaction was labeled as a positive emotional indicator.

- **Joy/Amusement:** Manifested through spontaneous laughter, positive verbal expressions, and animated positive reactions during activities. Joy/Amusement was labeled as a positive emotional indicator.

- **Uncertainty:** Displayed through hesitant behavior, verbal expressions of doubt, and decreased participation. Uncertainty was labeled as a negative emotional indicator.

- **Disappointment:** Shown through visible dejection, verbal disagreement with feedback, and decreased engagement following unexpected outcomes. Disappointment was labeled as a positive emotional indicator.

An annotator reviewed the videos and assigned these sentiment indicators, which were then used to compute an overall **Sentiment Score** for each participant using the following formula:

$$\text{Sentiment Score} = \frac{\text{Positive Sentiment Count} - \text{Negative Sentiment Count}}{\text{Total Sentiment Indicators}}$$

4.3.3 Comments

Comments were recorded at specific time points during the videos to provide additional context or clarification for particular interactions or technical issues encountered.

4.3.4 Video Recordings

Sessions were video-recorded to capture both the screen and the participants' reactions. This allowed for a comprehensive analysis of user interactions, behaviors, and emotional responses.

4.4 Data Classification and Calculation Methods

To accurately analyze the collected data, detailed classification systems were established for both the TTS and STT games. This section outlines the methods used to classify responses and calculate performance metrics.

4.4.1 Text-to-Speech (TTS) Game Classification

In the TTS games, participants listened to audio prompts and selected the correct answer from two options. Responses were classified into four categories:

1. **Correct:** The participant selected the correct answer independently.
2. **Incorrect:** The participant selected the incorrect answer independently.
3. **Error-Human:** The facilitator provided hints, pointed to the correct answer, or answered for the participant. These instances were excluded from performance calculations.
4. **Error-App:** Technical issues occurred, such as the audio not playing, forcing the participant to guess. These instances were recorded to assess the app's reliability.

Calculations for the TTS game were performed as follows:

- **Total Valid Responses:**

$$\text{Total Valid} = \text{Number of Correct} + \text{Number of Incorrect}$$

- **Correct Percentage (%):**

$$\text{Correct } (\%) = \left(\frac{\text{Number of Correct}}{\text{Total Valid}}\right) \times 100\%$$

- **Error Percentage (%)**

$$\text{Error } (\%) = \left(\frac{\text{Number of Error-App}}{\text{Total Valid} + \text{Number of Error-App}}\right) \times 100\%$$

4.4.2 Speech-to-Text (STT) Game Classification

In the STT games, participants read aloud, and the app's AI model analyzed their speech. Responses were classified into four categories:

1. **Correct:** The participant read correctly based on both the STT AI model and a manual review by the researcher (a native Icelandic speaker).

2. **Incorrect:** The participant read incorrectly based on both the STT AI model and the manual review.

3. **Error-Human:** Issues caused by human factors, such as the participant speaking too early or too late, the facilitator talking over the participant, or the facilitator answering for the participant. These instances were excluded from performance calculations.

4. **Error-AI:** Instances where the participant read correctly (as determined by manual review), but the AI model failed to accurately recognize the speech.

Calculations for the STT game were performed as follows:

- **Total Valid Responses:**

- **STT Accuracy (%):**

$$\text{STT Acc. } (\%) = \left(\frac{\text{Number of Correct} + \text{Number of Incorrect}}{\text{Total Valid}}\right) \times 100\%$$

- **Read Accuracy (%):**

$$\text{Read Acc. } (\%) = \left(\frac{\text{Number of Correct} + \text{Number of Error-AI}}{\text{Total Valid}}\right) \times 100\%$$

4.4.3 Justification for Manual Review

Due to the limitations of the STT AI model in accurately recognizing children's speech in Icelandic, the researcher, a native Icelandic speaker, conducted a manual review. This step was necessary because the AI model sometimes failed to recognize correct pronunciations, especially given the nuances of children's speech. By cross-referencing the AI's assessments with a human evaluation, we aimed to obtain a more accurate measure of participants' performance. While this introduces an element of subjectivity, consistent criteria were applied to minimize potential bias.

4.4.4 Sentiment Analysis Calculation

As previously mentioned, the Sentiment Score for each participant was calculated using the formula:

$$\text{Sentiment Score} = \frac{\text{Positive Sentiment Count} - \text{Negative Sentiment Count}}{\text{Total Sentiment Indicators}}$$

This metric provided an overall measure of each participant's emotional engagement during the session.

4.5 Data Analysis Methods

4.5.1 Quantitative Analysis

Quantitative data were analyzed using descriptive statistics. The following metrics were calculated:

- Mean scores and accuracy rates for each task and difficulty level.
- STT and TTS performance metrics, as outlined in the previous section.
- Frequency of technical issues (Error-App and Error-AI) to assess app reliability.
- Number of positive and negative sentiment indicators to gauge emotional engagement.

4.5.2 Qualitative Analysis

Qualitative data from video recordings and comments were analyzed thematically to identify recurring themes and insights. The analysis involved:

- Identifying key themes such as user engagement, confusion, and satisfaction.
- Categorizing observed sentiments based on the predefined indicators.

- Examining patterns related to app usability, instructional effectiveness, and participant attitudes.

4.6 Limitations

4.6.1 Sample Selection Bias

Utilizing convenience sampling via personal connections may result in selection bias. Participants were not chosen randomly, and their familiarity with the researcher or facilitator could influence their behavior and responses during the study. This limits the generalizability of the findings to the broader population.

4.6.2 Observer Bias

The researcher conducted a manual review of participant responses in the STT games, potentially introducing observer bias. Although efforts were made to apply consistent criteria and maintain objectivity, the potential for subjective judgments could affect the accuracy of the assessments.

4.6.3 Technical Limitations

Technical issues with the app, such as audio not playing (Error-App) or the AI model misrecognizing speech (Error-AI), could impact the participants' performance and the study's results. These factors may confound the interpretation of the data regarding the effectiveness of the STT and TTS technologies.

4.6.4 Generalizability

Given the small sample size and the study's specific context, the findings may not be generalizable to the broader population of Icelandic children or to other languages and cultural contexts. Future studies with larger, more diverse samples are necessary to validate and extend these findings.

5 Game-Focused Analysis

This chapter presents the study's results, focusing on the effectiveness of the Text-to-Speech (TTS) and Speech-to-Text (STT) features in the app, participants' emotional responses, and overall game performance. We also discuss the technical challenges encountered and their impact on the results.

5.1 Effectiveness of Text-to-Speech (TTS) in the Game

This section analyzes the participants' performance in the listening tasks utilizing the TTS feature. The tasks involved letter, word, and sentence comprehension, where participants selected the correct option after listening to the TTS output.

5.1.1 TTS Performance Overview

Table 5.1: TTS Performance by Participant

P	Age	Gender	Tasks	Total Valid	Correct (%)	Errors (%)
1	8	M	6 (2L+2W+2S)	62	98%	20%
2	7	M	6 (2L+2W+2S)	57	100%	5%
3	5	F	3 (2L+1W)	36	53%	3%
4	6	M	4 (2L+2W)	46	94%	0%
5	9	F	5 (2L+2W+1S)	49	100%	12.5%
6	6	M	6 (2L+2W+2S)	62	98%	3%
7	8	F	4 (2L+2W)	36	100%	3%
8	8	M	6 (2L+2W+2S)	55	98%	14%
9	6	M	4 (2L+2W)	40	98%	20%

Table 5.1 presents the main results for the TTS games. The table consists of the following columns:

- *P:* Participant identifier.
- *Age:* Participant's age.
- *Tasks:* Number and types of TTS tasks the participant engaged in. In this notation, L refers to letter games, W to word games, and S to sentence games.

- *Total Valid:* Total number of valid spoken items the participant attempted.

  Invalid items are connected to human errors that occur when the instructor provides the answer, hints at the answer with finger-pointing, or responds on behalf of the participant.

- *Correct (%):* Percentage of items where the participant correctly identified what the TTS system presented in cases where the sound played correctly. If the teacher had to answer or if the student had to ask the teacher for the correct answer, the student's input was not considered correctly received.

- *Errors (%):* Percentage of items with app-related errors, such as the sound not playing correctly.

5.1.2 Analysis of TTS Results

The data in Table 5.1 indicate that most participants demonstrated high accuracy in the TTS tasks. Participants 2, 5, and 7 achieved 100% accuracy in their completed tasks, suggesting that the TTS feature effectively conveyed the information.

Participants 1, 6, 8, and 9 also performed well, with correct response rates close to or above 98%. However, Participant 3 had a significantly lower accuracy rate of 53%. This lower performance may be attributed to the participant's younger age (5 years old), which could affect listening comprehension and familiarity with the task format.

The 'Errors (%)' column reflects the proposrtion of tasks affected by technical issues, such as the audio not playing correctly due to a bug in the app. Participants 1 and 9 experienced higher error rates (20%), likely impacting their ability to respond correctly. Unfortunately, some errors can occur randomly, primarily due to some sound issues that we could not resolve before the user study. If a participant had a long session, if someone else was using the app simultaneously, or if there were any caching issues, they were more likely to encounter additional errors. While we did not explore the reasons for the significant differences between students' scores further, these technical challenges highlight areas for improvement in the app's reliability.

5.1.3 Conclusion on TTS Effectiveness

Overall, the TTS feature effectively delivered audio content that participants could comprehend and respond to accurately. The high accuracy rates suggest that TTS can be a valuable tool in supporting listening comprehension tasks for Icelandic children. However, technical issues must be addressed to minimize their impact on user experience.

5.2 Effectiveness of Speech-to-Text (STT) in the Game

This section evaluates the STT feature's performance in voice-based tasks, where participants read aloud, and the app assessed their responses using the STT system.

5.2.1 Performance Overview

Table 5.2: STT Performance by Participant

P	Age	Gender	Tasks	Total Valid	STT Acc. (%)	Read Acc. (%)
1	8	M	3 (1L+2S)	25	68%	80%
2	7	M	3 (1L+1W+1S)	30	73%	83%
3	5	F	1 (1L)	2	100%	100%
4	6	M	1 (1L)	2	50%	100%
5	9	F	3 (1L+2S)	29	62%	93%
6	6	M	3 (1L+2S)	26	77%	55%
7	8	F	3 (1L+2S)	28	64%	78%
8	8	M	6 (1L+2W+2S)	45	80%	93%
9	6	M	3 (1L+2W)	27	74%	81%

Table 5.2 presents the main results for the STT games. The table consists of the following columns:

- *P:* Participant identifier
- *Age:* Participant age.
- *Tasks*: Number and types of STT tasks the participant engaged in. In this notation, L refers to letter games, W to word games, and S to sentence games.
- *Total Valid:* Number of times the participant spoke into the app for STT tasks.

  Invalid items are connected to human errors that occur when the instructor tells the answer, hints at the answer with finger-pointing, or responds on behalf of the participant.
- *STT Acc. (%):* Percentage of spoken inputs correctly recognized by the STT system. Valid recognition includes both accurately identified correct responses and correctly detected incorrect responses (i.e., when the system correctly identifies a mispronounced or incorrectly read word).
- *Read Acc. (%):* Percentage of tasks where the participant accurately read or pronounced the target content, regardless of whether the STT system recognized it correctly.

5.2.2 Analysis of STT Results

The results in Table 5.2 show variability in the STT recognition accuracy among participants. Participants 3 and 4 had limited data due to their engagement in fewer tasks but showed high ’Read Acc. (%)’. Participants 5 and 8 achieved high ’Read Acc. (%)’ (93%), indicating they read the content accurately. However, their ‘STT Acc. (%)’ was lower (62% and 80% respectively), suggesting that the STT system did not always recognize their speech correctly.

Factors contributing to these results may include:

1. **Age and Speech Clarity:** Younger children may have less clear pronunciation, affecting the STT system's ability to recognize their speech.

2. **STT Limitations:** The STT technology may not be fully optimized for children's speech in Icelandic.

3. **Environmental Factors:** Background noise or microphone sensitivity could have impacted recognition accuracy.

5.2.3 Conclusion

The STT feature showed potential but requires improvement to accurately recognize children's speech in Icelandic. Enhancing the STT system will improve the app's effectiveness in supporting reading-aloud activities.

5.3 Comparison of Reading and Listening Games

5.3.1 Performance Comparison Between Reading and Listening Games

This section, as shown in Table 5.3, provides a comparative analysis of participants' performance in the listening games (TTS) and reading games (STT).

Table 5.3: Performance Comparison Between Listening and Reading Games by Participant

	Listening Games (TTS)			Reading Games (STT)	P	
	Tasks	**Acc (%)**	**Err (%)**	**Tasks**	**STT Acc (%)**	**Read Acc (%)**
1	6	98	20	3	68	80
2	6	100	5	3	73	83
3	3	53	3	1	100	100
4	4	94	0	1	50	100
5	5	100	12.5	3	62	93
6	6	98	3	3	77	55
7	4	100	3	3	64	78
8	6	98	14	6	80	93
9	4	98	20	3	74	81

5.3.2 Analysis of Performance Comparison

Listening Games (TTS)

Participants showed high accuracy rates in listening games, with most achieving above 94%. Errors were minimal but varied among participants, possibly due to technical issues, as previously discussed.

Reading Games (STT)

The accuracy rates in reading games were more variable and generally lower than in listening games. The STT accuracy (the app's recognition of speech) was lower compared to participants' actual reading accuracy (as determined by manual review), indicating limitations in the STT technology.

5.3.3 Comparative Insights

Performance Discrepancy

- The higher accuracy in listening games suggests that participants found these tasks easier or that the TTS technology was more effective.
- The lower STT accuracy indicates challenges with speech recognition, especially with children's speech in Icelandic.

Impact of Age

- Younger participants (e.g., Participant 3) had lower accuracy in listening games but high reading accuracy, albeit with limited data points.
- Older participants performed better in both game types but still faced STT recognition challenges.

Technical Factors

Technical issues, such as audio playback problems and speech recognition errors, impacted performance and user experience.

5.3.4 Conclusion on Game Type Performance

The comparison between reading and listening games highlights the effectiveness of the TTS feature and the need for improvement in the STT functionality. While participants could comprehend and respond accurately in listening tasks, they faced challenges in reading tasks due to the app's difficulty in accurately recognizing their speech.

5.4 Sentiment Analysis

Participants' emotional responses were analyzed through a systematic review of video recordings, verbal expressions, and documented interactions. Sentiments were classified as positive or negative based on direct verbal expressions, physical reactions and body language,

engagement indicators such as asking questions and showing curiosity, and responses to both success and failure.

Participant 1 showed strong positive engagement through animated reactions, upright posture when receiving feedback, and enthusiastic responses to correct answers. They explicitly praised the learning tool as "the best app." However, they also exhibited frustration through table-knocking behaviors, particularly when encountering system response issues or experiencing challenging tasks. Their engagement pattern showed initial frustration transitioning to increased positivity, though late-session impatience was noted.

Participant 2 demonstrated notable confidence, explicitly stating, "This is so easy", and showing satisfaction through seeking and receiving instructor approval. Their excitement over their achievement was evident in verbal expressions like "Hundred points" accompanied by smiles. Their negative reactions were limited to occasional frustration with system recognition and resistance to more challenging content. Participant 3's emotional journey was marked by evident happiness with success, shown through consistent smiling reactions to correct answers. However, they also faced moments of diminished confidence, physically manifested through withdrawn posture and decreased participation.

Participant 4 showed unique engagement with the system's feedback mechanisms, spontaneously mimicking system sounds and displaying visible pride in their practice efforts. Their main frustration centered on system timing issues, expressing impatience with delayed responses. Participant 5 exhibited strong positive engagement through success celebrations, engaged laughter during exercises, and an explicit desire to continue activities, though they occasionally expressed frustration with unexpected system responses.

Participant 6 demonstrated a high level of engagement with interactive elements and a strong connection to the learning material, particularly showing enthusiasm for familiar content. Participant 8 displayed a mix of positive reactions, including finding humor in mistakes while also experiencing technical frustrations and performance disappointment. Finally, Participant 9 showed strong initial confidence and sustained engagement throughout their session, though they encountered occasional task confusion and expressed concerns about content difficulty.

Participant 7 participated in the tasks but did not display clear positive or negative emotional indicators, such as smiling, frowning, or expressing frustration or excitement. Their reactions were neutral, making it difficult to assess their engagement or response to the app using the established sentiment indicators. As a result, Participant 7's data was not included in the sentiment analysis.

Throughout the sessions, participants generally maintained positive engagement despite technical challenges and moments of frustration. Their emotional responses indicated a strong commitment to the learning process, with satisfaction and pride in achievements often outweighing temporary setbacks. The documented reactions suggest that while technical issues and task difficulty could trigger negative responses, most participants demonstrated resilience and maintained overall positive engagement in the learning experience.

5.4.1 Sentiment Frequency Summary

Table 5.4 provides an overview of the sentiment analysis for each participant.

5.4.2 Analysis of Sentiment Results

The sentimental analysis tracked both positive and negative emotional indicators across participants, revealing several key patterns:

1. **Overall Sentiments:** Participants exhibited more positive (25 instances) than negative (17 instances) emotions.

Table 5.4: Emotion Frequency by Participant

P	Age	Gender	Pos. Sentiment Responses	Neg. Sentiment Responses
1	8	M	5	5
2	7	M	3	2
3	5	F	2	2
4	6	M	2	1
5	9	F	3	1
6	6	M	3	1
7	8	F	0	0
8	8	M	4	2
9	6	M	4	3

2. **Age-Related Patterns:** No significant correlation was observed between age and the number of positive or negative sentiments.

3. **Primary Triggers for Negative Responses:** Technical issues (e.g., audio not playing), unexpected feedback, task complexity, and initial confidence challenges were common triggers.

4. **Engagement Levels:** Positive sentiments such as success satisfaction and content enjoyment suggest that the app engaged participants effectively.

5.4.3 Conclusion on Sentiment Analysis

The sentiment analysis indicates that the app generally elicited positive emotional responses from participants, enhancing engagement. Addressing technical issues and adjusting task difficulty could further improve user experience and reduce negative sentiments.

6. Discussions

6.1 Introduction

This study aimed to assess the effectiveness of integrating Speech-to-Text (STT) and Text-to-Speech (TTS) technologies into a gamified reading app to enhance reading performance among Icelandic children aged 5 to 9 years. By analyzing participants' performance and emotional responses, we sought to understand the impact of these technologies on their engagement in reading tasks.

6.2 Interpretation of Findings

6.2.1 Effectiveness of TTS Technology

The high accuracy rates in TTS tasks suggest that TTS technology could be an effective tool to enhance reading performance in young learners. Previous research has indicated that auditory support can improve literacy skills by providing correct pronunciations and aiding in word recognition (Marin et al., 2015).

6.2.2 Challenges with STT Technology

The variability and generally lower accuracy rates in STT tasks highlight challenges in speech recognition for children's speech, particularly in minority languages like Icelandic. These findings are consistent with Bailey and Wolfson's (2020) research, emphasizing the need for tailoring STT systems to the phonetic nuances of children's speech.

6.2.3 Emotional Engagement and Motivation

The sentiment analysis indicates that the app elicited more positive than negative emotions, contributing to engagement and motivation. This finding supports theories by Pekrun (2006) and Ryan and Deci (2000), emphasizing the role of positive emotions and intrinsic motivation in learning.

6.3 Answering the Research Questions

Based on the findings, we can address the research questions posed in this study:

Are STT and TTS sufficiently mature technologies in Icelandic to be applied in a reading application for children?

The study found that Text-to-Speech (TTS) technology is sufficiently mature in Icelandic to be effectively integrated into a reading application for children. Participants demonstrated high accuracy rates in listening comprehension tasks utilizing TTS, with most achieving above 90% correctness. This indicates that TTS can reliably deliver audio content that children can comprehend and respond to accurately, supporting its viability as a tool for enhancing reading skills in young Icelandic learners.

In contrast, Speech-to-Text (STT) technology exhibited significant limitations in accurately recognizing children's speech in Icelandic. The variability and generally lower accuracy rates in voice-based tasks suggest that STT systems are not yet fully optimized for the unique speech patterns of Icelandic children. Technical challenges, such as misrecognition of speech, and environmental factors, like background noise, impacted performance. Therefore, while TTS is ready for practical application, STT requires further development to meet the needs of Icelandic reading education for children.

What are the immediate effects of a single session with the gamified reading app on children's engagement levels, particularly comparing non-voice tasks (listening) with voice-based tasks (reading aloud).)?

A single session with the gamified reading app had a positive immediate effect on children's engagement levels, particularly in non-voice tasks involving listening. The high accuracy and positive emotional responses in TTS tasks indicate that children were highly engaged and found the listening activities both enjoyable and accessible. The gamified elements, such as interactive feedback and rewards, likely enhanced motivation and sustained interest during these tasks.

However, in voice-based tasks requiring reading aloud and utilizing STT, engagement levels were somewhat hindered by technical frustrations due to speech recognition inaccuracies. Children showed signs of frustration when the app failed to correctly recognize their spoken input, which occasionally led to negative emotional responses. Despite these challenges, the overall engagement remained positive; however, the comparison highlights that non-voice tasks currently offer a more seamless and engaging experience for children using this application.

6.4 Connection Findings to Literature

The findings from this study align with existing research emphasizing the importance of personalized and engaging educational technologies in enhancing reading performance among children. The user study demonstrated a degree of success in applying the TTS technology to support listening comprehension; however, additional research is necessary to ascertain its

actual impact on listening comprehension. Previous studies have indicated that auditory support can significantly aid young learners in developing literacy skills.

For instance, Marin et al. (2015) demonstrated that TTS technology enhances reading skills by providing correct pronunciations and aiding in word recognition. Similarly, Scholastic's "Ready4Reading" program integrates TTS features to provide immediate auditory feedback, enhancing phonemic awareness and pronunciation for young readers (Room, 2023). This program demonstrates that auditory support enhances children's ability to recognize words and understand their meanings, thereby improving overall reading comprehension.

The challenges encountered with STT technology highlight the necessity for language-specific and age-appropriate speech recognition models. Bailey and Wolfson (2020) emphasize that STT systems need to be tailored to the phonetic nuances of children's speech, especially in minority languages. SoapBox Labs, specializing in developing voice technology for children, recognizes that children's unique speech patterns require dedicated models for accurate recognition (Labs, 2023). Their research indicates that adult-centric STT systems often fail to accurately interpret children's speech, particularly in languages with limited resources like Icelandic. By creating child-specific STT solutions, SoapBox Labs has improved the efficacy of educational games that involve reading out loud, reducing frustration and enhancing learning outcomes.

Similarly, the Raddir project focuses on developing speech recognition tools for the Icelandic language to support children's reading development (Almannarómur, 2022). By collecting and analyzing children's speech data, Raddir creates more accurate STT models for Icelandic, thus improving interaction within educational apps and games. This underscores the importance of investing in language-specific resources to overcome technical challenges and improve user experiences in educational technologies. Some more recent models, such as a whisper-based model by Reykjavík University and the whisper-based model by Miðeind, have been trained on children's voices from the Talrómur dataset and could be better suited for STT in this setting than Google speech.

The positive emotional responses observed in the study are consistent with Pekrun's (2006) control-value theory of achievement emotions, which posits that emotions significantly influence students' motivation and academic performance. The engaging, gamified elements of the app likely contributed to increased motivation and enjoyment, aligning with findings that game-based learning environments can enhance educational outcomes by fostering positive emotions and intrinsic motivation. Ryan and Deci's (2000) self-determination theory further supports this, emphasizing that fulfilling the basic psychological needs of competence, autonomy, and relatedness enhances intrinsic motivation.

Educational programs like "Ready4Reading" have successfully utilized gamification to create immersive and enjoyable learning experiences for children, resulting in improved literacy skills (Room, 2023). These programs demonstrate that when educational content is delivered through engaging platforms that consider the emotional and motivational aspects of learning, children are more likely to engage and achieve better academic outcomes.

6.5 Implications for Educational Technology

6.5.1 Importance of Language-Specific STT and TTS Systems

Developing STT and TTS systems tailored to Icelandic is crucial for effective learning tools. Collaboration with projects like the Icelandic Language Technology Programme can improve recognition accuracy and support minority languages.

6.5.2 Enhancing User Engagement Through Gamification

The app's gamified elements contributed to positive emotional responses and sustained engagement. Incorporating game design elements can enhance motivation, aligning with the findings of Deterding et al. (2011).

6.5.3 Addressing Technical Challenges

Technical issues need to be addressed to improve the app's usability. This includes refining audio playback mechanisms and enhancing STT algorithms to better recognize children's speech.

7 Conclusions

This research demonstrated both the potential benefits and challenges of integrating speech technologies into educational tools for minority languages. While the Text-to-Speech (TTS) features showed strong effectiveness, with high accuracy rates above 90% for most participants, the Speech-to-Text (STT) functionality revealed significant limitations in accurately recognizing children's speech in Icelandic, highlighting critical areas for technical improvement. Nevertheless, the overall positive emotional engagement from participants, as evidenced by the sentiment analysis showing more positive than negative responses, suggests that such technologies can effectively enhance the learning experience when thoughtfully implemented. These findings underscore the importance of developing language-specific speech recognition models while maintaining engaging, gamified elements that motivate young learners. Addressing the technical challenges in STT technology while preserving the successful aspects of the TTS and gamification features could make this application a valuable tool for supporting literacy development in minority language communities.

Bibliography

AI, G. (2023). *Read along by google: A fun reading companion* [Accessed: 2024-08-30]. https://ai.googleblog.com/2023/04/read-along-by-google.html

Almannarómur. (2022). *Raddir: Icelandic speech and language technology* [Accessed: 202408-30]. https://almannaromur.is/raddir

Amazon Web Services. (2023). *Amazon polly* [Accessed: 2024-08-30]. https://aws.amazon.com/polly/

Bailey, B., & Wolfson, L. (2020). The impact of speech-to-text technology on reading proficiency: A systematic review. *Journal of Educational Technology*, *15*(2), 123– 137. https://doi.org/10.1016/j.jedt.2020.01.003

Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014a). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, *56*, 85–100. https://doi.org/10.1016/j.specom.2013.07.008

Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014b). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, *56*, 85–100. https://doi.org/10.1016/j.specom.2013.07.008

Birgisdóttir, F. (2016). Orðaforði og lestrarfærni-tengsl við gengi nemenda í lesskilningshluta pisa. *Netla*.

Castek, J., Coiro, J., Guzniczak, L., & Henry, L. A. (2020). Digital reading and engagement: Examining the impact of digital reading applications on young readers. *Reading Research Quarterly*, *55*(S1), 123–138. https://doi.org/10.1002/rrq.298

Dalton, B., & Proctor, C. P. (2021). Advances in digital reading assessment: Innovations and future directions. *Educational Assessment*, *26*(1), 56–74. https://doi.org/10.1080/10627197.2020.1864823

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". *Proceedings of the 15th International Academic MindTrek Conference*, 9–15. https://doi.org/10.1145/2181037.2181040

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Google Cloud. (2023). *Cloud speech-to-text api* [Accessed: 2024-08-30]. https://cloud.google.com/speech-to-text

Guðmundsdóttir, B. (2014). Rannsóknir sem undirstaða nsköpunar í námsefnisgerð fyrir börn. *Í Árdís Ármannsdóttir (ritstjóri): Þekkingin beisluð–nsköpun og frumkvæði*, 187–209.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 403–422, Vol. 3). Lawrence Erlbaum Associates. https://psycnet.apa.org/record/2000-16751-016

Bibliography

Hamari, J., Koivisto, J., & Sarsa, H. (2016). Does gamification work? a literature review of empirical studies on gamification. *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)*, 3025–3034. https://doi.org/10.1109/HICSS.2016.375

Hedström, S., Fong, J. Y., Þórhallsdóttir, R., Mollberg, D. E., Mestrou, T., Guðmundsson, S. F., Jónsson, Ó. H., Þorsteinsdóttir, S., Magnúsdóttir, E. H., Richter, C. L., Pálsson, R., & Gudnason, J. (2022). Samromur l2 22.09 [CLARIN-IS]. http://hdl.handle.net/20.500.12537/263

Hilmarsdóttir, Þ. (2020). *Þjálfunarapp í lestri fyrir börn*.
Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3–34. https://doi.org/10.1177/1529100615569721
Immordino-Yang, M. H., & Gotlieb, R. (2017). Embodied brains, social minds, cultural meaning: Integrating neuroscientific and educational research. *Mind, Brain, and Education*, *11*(4), 183–186. https://doi.org/10.1111/mbe.12158
Karam, R., Patton, K., Choudhury, M., & Olvera, A. (2018). Personalized learning in the classroom: Transforming education with technology. *International Journal of Education and Development using ICT*, *14*(2). https://ijedict.dec.uwi.edu/viewarticle.php?id=2452
Ke, F. (2019). Designing and integrating digital games in learning and assessment. *International Journal of Learning, Teaching and Educational Research*, *18*(3), 17–30. https://doi.org/10.26803/ijlter.18.3.2
Kim, Y. J., & Frick, T. (2020). Innovative digital tools for reading engagement and motivation: A study of elementary school students. *Journal of Educational Computing Research*, *58*(5), 1189–1214. https://doi.org/10.1177/0735633120906286
Kucirkova, N. (2020). Personalized learning with digital technologies: A scoping review of the literature. *British Journal of Educational Technology*, *51*(5), 1465–1481. https://doi.org/10.1111/bjet.12968
Labs, S. (2023). *Children's speech recognition & voice technology solutions* [Accessed: 2024-08-30]. https://www.soapboxlabs.com
Marin, L. F., de Oliveira, J. P., Marcolino, M. M., & Pereira, R. G. (2015). Text-to-speech technology for enhancing reading skills: A case study with elementary school students. *Computers in Human Behavior*, *51*, 35–45. https://doi.org/10.1016/j.chb. 2015.04.013
Miller, S., & Warschauer, M. (2014). Young children's internet use at home and school: Patterns and profiles. *Journal of Early Childhood Literacy*, *14*(2), 147–174. https://doi.org/10.1177/1468798412454130
Mozilla. (2023). *Deepspeech* [Accessed: 2024-08-30]. https://github.com/mozilla/DeepSpeech
OECD. (2019). *Pisa 2018 results: Country note for iceland* (tech. rep.). OECD. Retrieved July 10, 2024, from https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/pisa/publications/national-reports/pisa-2018/featuredcountry-specific-overviews/PISA2018_CN_ISL.pdf
OECD. (2023). *Pisa 2022 results (volume i): The state of learning and equity in education*. https://doi.org/10.1787/53f23881-en
Ólafsdóttir, S., Sigurðsson, B., et al. (2017). Hnignandi frammistaða íslenskra nemenda í lesskilningshluta pisa frá 2000 til 2015: Leiðir til að snúa þróuninni við.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9
Room, S. M. (2023). *Scholastics partnership with soapbox labs enhances phonics instruction* [Accessed: 2024-08-30]. https://www.scholastic.com/mediaroom/pressreleases/2023/soapbox-labs-partnership.html

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment: Advances in gamebased learning. In S. Tobias & J. D. Fletcher (Eds.), *Advances in game-based learning* (pp. 43–58). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780123884378000032

Sigurgeirsson, A., Bjarnadóttir, K., Jónsson, Ö., Rögnvaldsson, E., & Guðnason, J. (2021). Talrómur: A large icelandic tts corpus. *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, 440–444. https://aclanthology.org/2021.nodalida-main.50

Steingrímsson, S., Bjarnadóttir, K., Barkarson, S., Rögnvaldsson, E., Loftsson, H., Þorsteinsson, V., & Jónsson, H. (2020). Language technology programme for icelandic 2019–2023. *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*, 3421–3428. https://aclanthology.org/2020.lrec-1.418

Tiro. (2023). *Tiro text-to-speech services* [Accessed: 2024-08-30]. https://tiro.is/