Speech and Language Technologies Applied to the Therapy of Reading Disabilities

Matilde Huberty Nogueira Ramos

Thesis to obtain the Master of Science Degree in Information Systems and Computer Engineering

Supervisors: Prof. Alberto Abad Gareta
Prof. Isabel Maria Martins Trancoso

Examination Committee

Chairperson: Prof. Luís Manuel Antunes Veiga
Supervisor: Prof. Alberto Abad Gareta
Member of the Committee: Prof. Helena Gorete Silva Moniz

January 2021
Acknowledgments

First and foremost, a big-hearted thank you to my advisors for all the knowledge they generously shared with me. After a year filled with setbacks, they always kept encouraging me and helped me find ways around all the problems I encountered. This thesis would not be complete without their continued guidance and advice.

To doctor Luísa Teles, who spent countless hours helping me understand the intricate world of speech disorders and graciously made sure I had all the resources I needed for the development of this document. A special thanks to therapist Rosa Rodrigues for her kindness and availability whenever needed and to all of the other staff at the NCAB center for their contribution to this project.

To my friends, especially Hugo, Lucas, Francisco C., and Francisco T., for their companionship and support throughout these five years, which led me to where I am today. Each year was a challenge, and I couldn’t be luckier to have shared them with you.

Just as important was the support of my family; they are my inspiration and why I strive to be a better person every day. Thank you for helping me grow and for supporting me through all this time.
Abstract

Learning to read is a long-term process that starts at a very young age and usually results in acquired automatism and fluency while reading. However, some children never reach the level of skill expected for their age. Reading disabilities are the most common among children with learning disorders and, if not addressed, may lead to communication, emotional, and academic issues. Fortunately, speech therapy is a good opportunity to enhance a child’s reading skills. With the help of technology, children can have an accessible and reliable way to complement their therapy, whether it concerns speech assessment or practice. The purpose of this project is to develop a speech therapy tool aimed at European Portuguese children with reading disabilities that combines gamification and automatic speech recognition and validation.

Keywords

Speech Therapy; Mobile Application; Phonological Awareness; Gamification.
Resumo

Aprender a ler é um processo a longo prazo que começa muito cedo e que geralmente resulta na aquisição de automatismo e fluência durante a leitura. No entanto, algumas crianças nunca atingem o nível de habilidade esperado para a sua idade. As dificuldades de leitura são as mais comuns entre as crianças com distúrbios de aprendizagem e, se não forem corrigidas, podem levar a problemas de comunicação, emocionais e académicos. Felizmente, a terapia da fala é uma boa oportunidade para aprimorar as capacidades de leitura de uma criança. Com a ajuda de tecnologia, as crianças podem complementar a sua terapia de forma acessível e confiável, seja ela de intervenção ou de avaliação da fala. O objetivo deste projeto é desenvolver uma ferramenta de terapia da fala destinada a crianças portuguesas europeias com dificuldades de leitura que combina gamificação e o reconhecimento e validação automático da fala.

Palavras Chave

Terapia da Fala; Aplicação Móvel; Consciência Fonológica; Gamificação.
## Contents

1 Introduction .................................................. 1
   1.1 Problem ................................................ 4
   1.2 Goals ................................................... 4
   1.3 Organization of the Document ......................... 4

2 Background .................................................. 7
   2.1 Phonological Awareness ................................. 9
      2.1.1 Phonological Skills and Reading .................. 9
      2.1.2 Assessment in Early Ages ......................... 10
      2.1.3 Phonological Development in Portuguese Children . 11
      2.1.4 Exercising Phonological Awareness ............... 12
   2.2 Gamification and Education ............................ 13
      2.2.1 Phonological Awareness ............................ 14
   2.3 Automatic Speech Recognition ........................ 15
      2.3.1 Children’s Speech Recognition .................... 15

3 Related Work ................................................. 17
   3.1 DaisyQuest ............................................. 19
   3.2 Lexia Core5 Reading .................................... 20
   3.3 Apraxia World ........................................... 22
   3.4 VITHEA .................................................. 24
      3.4.1 VITHEA-Kids ....................................... 25
      3.4.2 VITHEA 2.0 ......................................... 26
   3.5 Discussion ............................................... 26

4 Implementation ............................................... 29
   4.1 Observations at NCAB Center ......................... 31
   4.2 Level Selection ......................................... 32
   4.3 Task Selection .......................................... 33
      4.3.1 Rhyming Tasks ..................................... 33
A  Study Protocol for Testing at NCAB Center  71
B  Form and Approval of the Ethical Committee at IST  77
C  Informed Consent  85
List of Figures

2.1 Example of a rhyming exercise. João (John) rhymes with balão (balloon). Extracted from a set of exercises available at NCAB center. 12

2.2 Example of a syllable identification exercise. Words that start with the same syllable are painted (papagaio (parrot), patins (skates) and panela (pan), but not pinheiro (pine tree)). Extracted from a set of exercises available at NCAB center. 13

2.3 Example of a syllable manipulation exercise. Removing a syllable creates a new word: sapato (shoe) becomes pato (duck) and orelha (ear) becomes olha (look). Extracted from a set of exercises available at NCAB center. 13

3.1 Capture of the Daisy Quest application. 20

3.2 Capture of a phonological awareness exercise in Lexia Core5 Reading. Extracted from [1]. 21

3.3 (a) Clothing store. (b) Capture of the game. (c) Speech exercise popup in the during-game condition. (d) Speech exercise popup in the after-game condition. Images extracted from [2]. 23

3.4 Capture of the VITHEA application. Extracted from [3]. 25

4.1 Examples of exercises used at the NCAB center. 31

4.2 General system architecture. 36

4.3 Architecture and resources used in a TTS request. 37

4.4 Architecture and resources used in an ASR request. 38

5.1 Title screen and settings. 41

5.2 Player screens. 42

5.3 Levels and tasks screens. 42

5.4 Rhyming identification exercise. 43

5.5 Examples of rhyming tasks. 44

5.6 Example of an exercise to identify the number of syllables with no displayed options. 45

5.7 Example of a phoneme omission exercise. 45
List of Tables

3.1 Summary of related work features. .............................................. 27
6.1 Summary of the participant's demographic information. ................... 53
6.2 List of tasks for the effectiveness evaluation. ................................ 55
6.3 List of tasks for the efficiency evaluation. .................................... 56
6.4 Satisfaction results. ................................................................. 57

Listings

4.1 Example of a rhyming exercise in XML form. ................................. 37
### Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR</td>
<td>Automatic Speech Recognition</td>
</tr>
<tr>
<td>GRXML</td>
<td>Speech Recognition Grammar Specification in XML form</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>IPA</td>
<td>International Phonetic Alphabet</td>
</tr>
<tr>
<td>NCAB</td>
<td>Neurodevelopment and Child and Adolescent Behavior</td>
</tr>
<tr>
<td>SSML</td>
<td>Speech Synthesis Markup Language</td>
</tr>
<tr>
<td>TTS</td>
<td>Text-to-Speech</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
<tr>
<td>X-SAMPA</td>
<td>Extended Speech Assessment Methods Phonetic Alphabet</td>
</tr>
</tbody>
</table>
Introduction

Contents

1.1 Problem ......................................................... 4
1.2 Goals .......................................................... 4
1.3 Organization of the Document ............................... 4
The acquisition of language in early childhood is essential to the development of communication skills, academic achievement and emotional regulation, in the way that it gives children the means to communicate their own needs and understand other people’s emotions [5].

The ability to read comes from a long-term learning process that usually ends with being capable to read with a considerable level of automatism. However, some children are unable to achieve this ability despite having a normal intellectual capacity and going through the same learning process. Children in this situation may simply require more practice than their counterparts to reach their expected skill level. Nevertheless, an underlying condition could be responsible for the problem. The most common disorder entailing reading disabilities is dyslexia, but there could also be articulation or phonological disorders. Incidentally, many other disorders can coexist with reading difficulties, even if reading is not the most prominent problem. Reading disabilities tend to emerge in early years and can have a significant impact on social interactions, communication, or educational achievements and can persist throughout the school years and into adulthood [6].

Adequate reading skills are critical in acquiring knowledge in all academic areas, whether in science or social studies fields [7]. Thus, a child with reading disabilities who does not seek intervention can develop into an adult with limited employment prospects, and may face greater difficulties in some aspects of adult life such as relationships and social situations [8]. Furthermore, students with poor reading skills are more prone to drop out of school, estimated at nearly 1.5 times the general population [7].

Reading problems have been estimated to account for at least 80% of children with learning disabilities. Furthermore, most children who are poor readers in the primary years continue to read poorly in adolescence [9]. Consequently, the need to invest in early intervention is evident and imperative. Some measures have already been developed to identify at-risk children with reading difficulties [10], however many of them are not designed to provide sufficient intervention to children until they are already starting or have surpassed the early primary school years.

One of the most pertinent strategies for acquiring literacy skills is repetition, which means the student should repeat the same activities often in order to gain speed and accuracy [11]. With this in mind, it is apparent how digital games show promise for these types of approaches, not only providing the automation of different types of activities but also allowing the student to execute the exercise as many times as they need. Moreover, well-designed games are a good way to keep students engaged and motivate them to work outside of school, besides serving as a complementary tool to therapists and educators. It is also possible to integrate into the game useful functionalities such as Automatic Speech Recognition (ASR), which allows for the independent processing and validation of the student’s speech, granting some autonomy from the therapist or parent.
1.1 Problem

The interaction with Luz hospital's center of Neurodevelopment and Child and Adolescent Behavior (NCAB) was essential to understand the context of our problem. It allowed for a discussion with doctor Luisa Teles that helped to recognize the main problems in the assessment and intervention of reading disorders. The NCAB center specializes in child development and meets students struggling with reading every day. When children show some difficulties with reading and writing, they usually come to the center looking for an evaluation of their capacities. This includes a doctor's appointment and a meeting with a therapist, who assesses the child's level of skill through a series of exercises. If the therapist detects some issues, the child is then directed to appropriate therapy sessions that address them. Therapy sessions are executed on a one-to-one basis, which means that there needs to be one therapist for each person with a disability. This, of course, is highly impractical. Moreover, although this is not the only center addressed to helping behavioral and developmental issues in early ages, the majority of children are not able to benefit from services like these for a number of reasons that can range from monetary incapability to geographical motives. Many of them grow up without any assistance and with no idea of why they are struggling.

Ideally, the assessment of reading problems should be available to everyone and at the earliest age possible. This way, we would be able to identify children that could have or potentially develop a reading disability in the future.

1.2 Goals

To address this situation, we developed a mobile application that helps to practice basic reading skills in European Portuguese. The app is targeted at children who are entering primary school (about 5 years old) and have developed comprehensive speech capacities but, most likely, have not yet learned how to read. The app provides basic exercises and records the child’s performance over time, which can be monitored by a therapist. If the performance does not correspond to the expected level of the child's capacities, there may be an underlying condition that should be assessed by a doctor. It can also serve as a helpful tool for therapists working with children. This way, we guarantee that an early intervention in potential reading disabilities is accessible to all children.

1.3 Organization of the Document

This thesis is organized as follows: In Chapter 2 we present the background related with our work and in Chapter 3 we examine some developed related approaches. Chapter 4 describes the implementation
stage, which comprehends our observations at the NCAB center and the app’s implementation process. Chapter 5 describes the overall system overview, including the user interface. The evaluation of the app is presented in Chapter 6. Finally, Chapter 7 concludes this thesis.
Contents

2.1 Phonological Awareness ................................................. 9
2.2 Gamification and Education ............................................ 13
2.3 Automatic Speech Recognition .......................................... 15
Language acquisition requires the coordination of four systems [12]: Phonology, Morphosyntax, Semantics and Pragmatics. Phonology studies speech sounds (i.e. phonemes), including the rules for combining and using them. Morphosyntax considers words and their relations in sentences. Semantics consists of the meaning in language, while Pragmatics is concerned with communication, particularly how language is used in context.

During the pre-school years, children can already speak, though they do not seem to be consciously aware of the existence of words, syllables, or phonemes and have limited ability to manipulate these components. In the early stages of learning to read, children are prone to learning decoding skills, that is, matching a letter or a combination of letters to their sounds. This mapping between speech sounds and letters is called the alphabetic principle (or Phonics), and it is the foundation for learning a language that relies on an alphabetic system, like English or Portuguese. The development of more complex language skills, such as vocabulary, grammar, and pragmatics, takes place later when children begin to rely on word meanings to gain fluency in their reading. Thus, it is clear that to understand and improve reading development we need to focus first on the basic skills that start this process. One of the skills linked to the understanding and development of the alphabetic principle [13] is phonological awareness.

2.1 Phonological Awareness

Phonological awareness can be defined as the ability to identify, process and manipulate the sounds of a language. Phonological awareness tasks may include rhyme production, phoneme recognition and word segmentation, and will be further explored in Section 2.1.4. First, however, we need a better understanding of how closely phonological processing and reading are related.

2.1.1 Phonological Skills and Reading

The relationship between phonological skills and later reading achievements has no shortage of documented studies. Bradley and Bryant [14] conducted a study on this topic, where 400 children between the ages of four and five were tested with a few phonological tasks. None of the children knew yet how to read. Four years later, they were given a standardized test on reading and spelling, together with a mathematical test. Intellectual differences (IQ), memory, and social aspects were also tested to exclude their effects. Nevertheless, a strong relationship was found between the children's initial phonological awareness skills and their reading proficiency years later. However, this pattern was not found in their mathematical performance, which shows that phonological awareness is not a general predictor of academic achievement, but a predictor of reading and spelling success. Muter et al. [15] also investigated the development of language and reading skills in the first two grades of school. Their work concluded that the foundation of word-level reading skills comes from phoneme awareness and letter knowledge,
while vocabulary knowledge and grammatical skills play additional significant roles in reading comprehension.

Several studies have shown that phonological awareness can be a strong indicator of poor future reading skills or even the root cause of reading problems ([16], [17]). Thankfully, phonological skills are trainable. Interventions targeting phonological processing have proved to be good approaches to improve them in the long-term. Lundberg, Frost, and Petersen [18] followed young children from kindergarten, first and second grades for a period of 8 months. During this time, the children were put in training programs to teach phonological processing skills. A different control group followed the regular preschool program. The results showed that the undergone training not only enhanced these skills comparing to the control group, but also significantly increased their reading and spelling abilities. Hence, phonological skills can be developed outside the context of formal reading instruction.

### 2.1.2 Assessment in Early Ages

The use of phonological skills assessment as a measure towards predicting the child’s success in reading and distinguishing good and bad readers seems apparent. However, learning to read is a continuous process that is developed throughout many years of school. Thus, it makes sense to ask if this predictor works for all children, whether they are just starting to read or they already have some competence reading. While researching this question, Hogan et al. [19] conducted a study intended to investigate the usefulness of phonological awareness assessments in kindergarten, second and fourth grades to a sample of 570 children. The study found that applying kindergarten measures of phonological awareness provided information on the prediction of second-grade reading, which meets the findings from previous studies. However, this was not the case from second to fourth grade. Here, a measure of phonological awareness offered little to no relevant information to the prediction of word reading. This research infers that training phonological awareness to improve reading in children should be considered with children who are not yet too familiar with reading, such as preschoolers, kindergarteners, and, at most, first graders.

So far, we saw that there is correlation evidence establishing a causal relationship between phonological awareness and reading skills. Performance on phonological processing can be predictive of students who will struggle with the development of basic reading skills. This supports the importance of monitoring the development of this skill in young children, particularly children finishing kindergarten and entering elementary school. We may now wonder how this monitoring can be carried out and what exercises may be executed to this end. In the next Section, we will focus on the phonological path Portuguese children experience so we can better understand the foundations of the exercises, which will be exemplified in Section 2.1.4.
2.1.3 Phonological Development in Portuguese Children

Generally, in most languages, phonological development is one of the first language components to be established, and Portuguese is no exception [20]. At the start of primary school, phonological units are already acquired in most children. However, some concepts may not be entirely apprehended, which could lead to reading, writing, and communication problems. Speech sounds are commonly classified as vowels, which are characterized by the free passage of air from the vocal tract, and consonants.

Portuguese children are usually aware of plosive consonants first. The sound of these consonants is produced by blocking the airflow in the vocal tract, as is the case for the initial sounds in the words “pai” (father), “bola” (ball), “tia” (aunt), “dado” (dice), “casa” (house), or “gato” (cat). Nasal consonants, such as the ones in “mão” (hand), “nada” (nothing), or “unha” (nail), are also acquired early on. The awareness of these consonants is succeeded by fricative consonants, which are pronounced by squeezing air through a small hole or gap in the mouth. They can be found at the beginning of the words “faca” (knife), “voz” (voice), “seda” (silk), “zebra” (zebra), “chá” (tea) and “giz” (chalk). Liquid consonants are some of the last sounds to be acquired, and they may only be perceived when the child is already attending the first grades of school. These correspond to the consonant sounds present in the beginning of “lua” (moon), and in the middle of “olho” (eye), “caro” (expensive), and “carro” (car). Failing to assimilate these sounds may induce problems when mapping graphemes to phonemes, leading to mispronunciation of words.

Another recurring difficulty is the identification of borders in words [21]. Children tend to combine determinants to the words that proceed them “os amigos” (the friends) becomes “zamigos”. They can also partition a word in two or more words (“umbigo” is usually separated as “um” and “bigo”, where the first syllable is considered a quantifier and causes the common mistake of saying “dois bigos”).

The evolution of phonological awareness in children ranges from sensitivity to larger segments of speech, such as words or syllables, to sensitivity to the phonemic components of words [22]. In fact, children can indicate the number of syllables in a word before they can indicate the number of phonemes in a syllable or word [23]. Hence, phonological sensitivity evolves towards the apprehension of successively smaller phonological segments. Consequently, studies analyzing phonological awareness tend to practice with exercises that contain rhyming, syllabic, and phonemic tasks ([15], [19], [24]). A child must first be aware that two words sound the same to be able to rhyme. Rhyming skills help early readers understand that several phonemes can function as a unit, and that by changing one phoneme, a new word can be created [25]. This concept enhances the idea that tasks requiring dividing words into syllables, along with rhyming, are developed prior to tasks that require the manipulation of phonemes [21].

11
2.1.4 Exercising Phonological Awareness

After understanding how phonological awareness is developed in children, we can put together a series of exercises that take this into account. Typically, students that need therapy have not acquired the most basic skills of phonological awareness, particularly, rhyming and syllable manipulation. Hence, it is important to start by practicing these first and then move on to working with phonemes.

One of the most important aspects when dealing with exercises targeted at children is to make sure the child understands the tasks and the words presented. Therefore, the exercises should always have visual support, specifically pictures of the words mentioned. Furthermore, the child should always be able to hear the pronunciation of the words and repeat them if necessary. Maintaining the child’s attention is also one of the biggest challenges, so it is customary that the exercises involve writing, drawing, or some other kind of handwork.

- **Rhyming** - For two words to rhyme, they need to end in a similar sound. Figure 2.1 shows a possible exercise for detecting rhymes. The child hears a sentence and has to choose the word that rhymes with it by circling the right answer between two options.

![Figure 2.1: Example of a rhyming exercise. João (John) rhymes with balão (balloon). Extracted from a set of exercises available at NCAB center.](image)

- **Syllable Identification** - These exercises require awareness on the location of the syllables in a word. In Figure 2.2, the goal is to paint the words that start with the same sound (first syllable). Because the child is rewarded with drawing, it serves as motivation to stay committed to the exercise. This activity can be repeated with the first, final, or any middle syllable of the word.

- **Syllable Manipulation** - In these tasks, syllable awareness is required as well. However, in addition to acknowledging the syllables, the child needs to handle deleting, adding, and reordering them. Figure 2.3 demonstrates one of these exercises, focused on syllable deletion, where all the initial (left side) words have three syllables. The child hears the word and removes one of the syllables (the one with a red circle). This makes for the production of a new word with two syllables (right side). The child has to correspond the initial word to the altered one with a pencil, again contributing to their engagement.
Figure 2.2: Example of a syllable identification exercise. Words that start with the same syllable are painted (papagaio (parrot), patins (skates) and panela (pan), but not pinheiro (pine tree)). Extracted from a set of exercises available at NCAB center.

- **Syllable Segmentation and Formation** - Typically, in syllable formation, the child listens to a sequence of syllables and has to answer with the word that is created when putting them all together. Contrarily, in syllable segmentation, a word is pronounced, and the child needs to recite all the syllables that compose that word.

Figure 2.3: Example of a syllable manipulation exercise. Removing a syllable creates a new word: sapato (shoe) becomes pato (duck) and orelha (ear) becomes olha (look). Extracted from a set of exercises available at NCAB center.

When the previous competences are well assimilated, we can proceed to phonemic awareness skills. These imply the same type of exercises, namely phonemic identification, manipulation (addition, deletion, and reordering), segmentation, and formation.

### 2.2 Gamification and Education

Gamification is the use of game-play mechanics for non-game applications [26]. Theoretically, any application, task, process, or context can be gamified. Gamification’s main goal is to raise the engagement of users by incorporating game-like techniques and personalized fast feedback, making people feel more ownership and purpose when engaging with tasks [27].
The use of game mechanics, dynamics, and frameworks to promote desired behaviors has found its way into domains like marketing, politics, health, and fitness [28]. However, one of the most promising uses of gamification is as learning tools, due to its ability to teach and reinforce not only knowledge but also essential skills such as problem-solving, collaboration, and communication. Gamification as learning tools consists of extracting game elements of deemed enjoyable games and adapt them so they can be used in the teaching process. Thus, students learn as if they were playing a game [29].

Lee and Hammer [28] define three major areas in which gamification can serve as an intervention.

- **Cognitive** - Games provide complex systems of rules for players to explore through active experimentation and discovery. Consequently, players need to handle all sorts of cognitive concepts, for example, maths or physics, strategy, and planning skills.

- **Emotional** - Games can invoke feelings of curiosity, joy, pride, or frustration. More importantly, they make failure be perceived as a necessary part of learning and create an environment in which effort, not mastery, is rewarded. By offering active feedback and a way for learners to evaluate their capabilities, students can learn to see failure as an opportunity instead of a dreadful experience.

- **Social** - While playing a game, players get to try different roles and identities. These are not necessarily fictional roles, as children can play characters that lead to the discovery of new sets of skills, such as leadership, communication, or team-playing abilities. Games offer a way for players to explore new sides of themselves and try unfamiliar identities in a safe space of play.

In brief, by incorporating game elements into work activities, students become motivated to engage in learning, and through extended practice develop personal qualities such as persistence, creativity, and resilience [28]. Therefore, and as we will see in the next section, this method shows promise regarding teaching phonological skills in an appropriate and practical fashion.

### 2.2.1 Phonological Awareness

Ever since the first appearances of computer software, researchers have wondered about the success of computer-assisted instruction applied to students struggling to learn how to read. In 1995, Barker and Torgesen [30] tried to answer this question by training 54 children, divided into 3 groups, with a game composed of a set of 3 types of computerized exercises. The first group received some phonological awareness training, while the second received the same amount of training with a program designed to train decoding skills. The third group worked as a control group and practiced basic math skills. At the end of 8 weeks of training, the children exposed to the phonological awareness training programs made significantly greater improvements on several measures of phonological awareness, when compared to children in the other two groups. The study did not provide results on the advantages of computer-
assisted instruction over therapist or teacher instruction, but it provided proof that computer software is reliable in the training of phonological awareness.

Other studies performed over the years also support the effectiveness of this type of training. For instance, in 2006, Macaruso et al. [31] examined the benefits of a computer program designed to supplement regular reading instruction over one school year. They found that the first graders who participated in the program made significant reading gains, and the improvement was even greater for children with low academic achievements. In 2009, Wild [32] followed two intervention groups that undertook the same phonological awareness program. While one group practiced exercises using a paper-based format, the other practiced using a computer. Children in the computer-based group showed a significant learning advantage compared to the first group, which could be explained by faster and continuous access to feedback while executing the exercises.

In general, computer software proved to be efficient in increasing phonological awareness skills and showed promise in helping at-risk children of reading failure.

2.3 Automatic Speech Recognition

In simple terms, automatic speech recognition is the process of interpreting human speech on a computer. Applications where ASR is used vary from simple tasks to more complex ones [33]. More recently, with the exponential growth of big data and computing power, ASR technology has advanced to the stage where more challenging applications are becoming a reality [34]. Examples are voice search and interactions with mobile devices, voice control in home automation systems, air traffic control, security and biometric identification, or gaming. It is also a big advantage to people who suffer from writing disorders or to evaluate the speech of people with speaking or reading disabilities.

However, most available speech recognition systems are targeted at adults, to which automatic speech recognition and speech technologies are significantly developed and already show promising accuracy results.

2.3.1 Children’s Speech Recognition

Although automatic speech recognition systems have good performance with adult speech, this is not the case with children. In fact, the word error rates are, on average, two to five times worse for children speech than for adult speech [35].

When children read aloud, they display unpredictable behavior such as false starts followed by self-corrections, mispronunciations of words, and reading words slowly or syllable by syllable [36]. Moreover, children’s speech not only differs a lot from adult speech but also between each child. The spectral and temporal characteristics of children’s speech are highly influenced by growth and other developmental
changes, and only reach adult levels by the age of 15 [37]. The wide range of possible problematic events make ASR for child speakers a challenging problem. Still, over the years, new approaches have been developed that succeed at improving ASR system’s performances [38]. However, they require a large amount of training data on children’s speech, which is not generally available. Some projects have tried to answer this problem, such as BioVisualSpeech.

BioVisualSpeech is a research project that explores multimodal human-computer interaction mechanisms for providing bio-feedback in speech and language therapy through the use of serious digital games [39]. In order to develop speech classifiers for a game concerning sibilants (sound of consonants generated by letting air flow through a very narrow channel towards the teeth), they built a children’s speech corpus. The corpus consists of 70 European Portuguese single words with sibilant consonants recorded from 365 children. The data annotations include information on the quality of the sound productions according to speech and language pathologists criteria. There have also been other efforts to collect European Portuguese speech recordings of children in a variety of reading tasks, as is the case for the LetsRead database [40] and the CNG database [41]. The LetsRead database contains reading aloud recordings of 284 European Portuguese children, adding up to about 5 hours and 30 minutes of sentences extracted from children fairy tales and grade-specific scholar books. The CNG database contains reading aloud recordings from 510 children. The prompts presented to the children for reading were chosen from a set of phonetically rich sentences, musical notes, isolated cardinals, and sequential cardinals.

Together with BioVisualSpeech, these resources are extremely valuable for the development of speech and language processing modules intended for children.
3

Related Work

Contents

3.1 DaisyQuest ............................................................ 19
3.2 Lexia Core5 Reading ............................................... 20
3.3 Apraxia World ..................................................... 22
3.4 VITHEA ............................................................... 24
3.5 Discussion ........................................................... 26
In this chapter, some approaches that are relevant to our project will be examined. First, two educational tools aimed at supplementing classroom instruction of phonological awareness will be reviewed. DaisyQuest uses a game with an engaging storyline for children, contrary to LexiaCore5 Reading, which offers independent exercise execution, but provides many tools for assessing and monitoring children's progress. In section 3.3, Apraxia World is presented as an approach highly focused on the gamification aspect of computer-assisted instruction. Finally, VITHEA shows promise on the way it interacts with patients, including the integrated automatic speech recognition and evaluation module.

3.1 DaisyQuest

DaisyQuest was released in 1992, making it one of the earliest developed computer-assisted instructions in phonological awareness. It is targeted at children learning English aged three to seven years (or preschoolers to second graders) [42]. The instructional activities are framed in a fairy tale world involving a search for Daisy, a shy and friendly dragon. As children master individual levels within the program, they gain clues to the location in which Daisy is hiding, and by completing all the instructional activities, they can eventually discover her hiding place [30]. Children have to execute tasks that include recognizing words that rhyme and recognizing words that have the same beginning, middle, and ending sounds.

Daisy’s Castle is a follow-up instructional program bundled together with DaisyQuest, in which the activities also take part in a fairy tale theme, this time involving a search for Daisy’s lost eggs. As children complete each level of instructional activities, they are given clues that take them to the location of the lost eggs. Daisy’s Castle teaches the additional skills of recognizing words that can be formed from a series of separately presented phonemes and counting the number of sounds in words.

DaisyQuest is an interactive program, and as such, children can practice these skills in a computer without the need to implement additional curricular materials. The programs use storylines, colorful graphics, and both digitized and synthesized speech to engage children in the learning process. The exercises are all presented orally with no written text. The programs keep track of children’s progress through the imaginary world in which Daisy is hiding, as well as offer children choices about the sequence of instructional activities. Each activity contains an instructional/practice module in which children are guided through a tutorial explaining each skill or concept briefly and then taught how to perform the activity. By this point, they are allowed to practice as much as they want [43]. Following the tutorial, the child practices with multiple-choice and yes/no items, and receives feedback from the programs’ speech capabilities [44]. Children navigate and enter responses by clicking with the mouse.

There are three skill levels for each activity, with response time decreasing at advanced levels. On the first level, there is no time limit. On the second and third levels, time limits are eight and four
seconds, respectively. These advanced levels provide more challenging practices and contribute to building fluency in each of the skill areas. If the child masters the skill, he/she is rewarded with a clue to the location of Daisy.

Several studies were conducted to determine the efficacy of DaisyQuest as an educational tool ([30], [44], [45]). In general, all studies found that exposure to both programs produced statistically significant improvements in children’s phonological sensitivity and word reading skills. These reports demonstrate that the programs are a reliable method for teaching phonological awareness.

![Daisy Quest Application](image)

Figure 3.1: Capture of the Daisy Quest application.

More recently, DaisyQuest was remastered as an application available for the iPhone. Figure 3.1 shows an example of a rhyming exercise, where the player is asked if the two words illustrated in the images (man and pan) rhyme. The player should answer by tapping the yes or no button. The application provides a simple yet engaging interface, while teaching children phonological awareness skills.

### 3.2 Lexia Core5 Reading

Lexia Core5 Reading [1] is one of the main programs marketed by Lexia Learning Systems. It is a computer-based approach, accessible via iPad or browser, aimed at improving English reading skills from 4-year-old children to adults. Although it is designed to supplement classroom instruction, it can be used as a stand-alone tool. The program provides learning in six areas of reading instruction: Phonological Awareness, Phonics, Structural Analysis, Automaticity, Vocabulary, and Comprehension.

The phonological awareness activities include identifying, segmenting, blending, and manipulating syllables and sounds in words. The program acknowledges that young children first develop an awareness of the phonological patterns that occur at the end of rhyming words, followed by an awareness
of syllables. Lastly, and most importantly for reading success, students should also develop phonemic awareness as they learn to analyze and manipulate individual sounds within words. Taking this into account, Lexia Core5 provides picture matching activities that emphasize recognition of rhyming words and the ability to blend syllables in spoken words. Figure 3.2 depicts one of these exercises, where the student listens to a sequence of syllables and has to click on the image of the word that is created by combining the syllables (in this case, "coffee"). Students also learn to segment spoken words by identifying the number of syllables they hear. Blending and segmenting activities begin with compound words and progress to three-syllable words. Once these skills have been established, students advance to phonemic awareness activities. In these exercises, they match pictures with the same beginning and ending sounds, and also blend and segment individual phonemes in words.

![Figure 3.2: Capture of a phonological awareness exercise in Lexia Core5 Reading. Extracted from [1].](image)

When students begin using Core5, they are automatically placed at the proper skill level. Each student moves at his or her own pace and can progress independently to higher levels. If the student struggles in a particular activity, the program removes some of the answer choices and stimuli on the screen. If the student continues to struggle, the program provides explicit instruction on the concepts and rules of the skill, allowing the student to demonstrate proficiency and then return to the standard-level activities. If the student continues to struggle within the explicit instruction, the teacher is notified and provided with the data and resources for direct instruction on that particular sub-skill.

Not only does Lexia Core5 present students with systematic and structured practice on the essential reading skills, but it also incorporates an assessment component. This process, referred to as Assessment Without Testing, provides performance data in real-time, without administering a test. Educators can monitor students’ usage and progress through data reports, collected as students work independently in the online activities, that provide insight into progress at the district, school, grade, class, and student level. These reports present educators with a performance predictor that informs them of
each student's percent chance of reaching end-of-year benchmarks. Class-level reports identify the
skill areas in which the student is currently working, the specific skills on which the student needs in-
struction, the student's progress towards weekly usage goals, and the completed levels with certificates
of achievement to celebrate success. Monthly changes in student reading gains are also identified,
allowing teachers to monitor progress for each student and for the entire class.

Lexia Core5 Reading provides a thorough and helpful tool for educators to monitor their students. However, it does not offer any type of gameplay, and the exercises are not connected in any way, which could fail to keep younger children engaged.

### 3.3 Apraxia World

Apraxia World [2] is a remote therapy tool for speech sound disorders that result from apraxia of speech, a motor speech disorder that makes it hard to say sounds and words. Apraxia World integrates speech exercises into an engaging platformer-style game.

The game was developed atop a full-featured game project available at the Unity Game Engine that comes with 48 levels divided into eight worlds, multiple characters, and an in-game store for clothing, weapons, and power-ups, shown in Figure 3.3(a). The gameplay is linear - the player controls a monkey avatar with virtual buttons and has to reach the goal line at the right side of each level by navigating platforms, caverns, and other obstacles while trying to collect assets and avoid or eliminate enemies. Figure 3.3(b) displays a capture of the game, where level and character information is shown at the top of the screen.

The game offers two types of assets to collect: coins and stars. Coins are dispersed throughout the levels and are used to purchase items in the store. A predetermined number of starts need to be collected by the player to complete a level. Each star requires the player to complete some speech exercises. By associating speech production with the stars, players can anticipate and control when speech exercises appear, which avoids detracting from the gameplay or interrupt the player while executing complicated moves. Two options were considered with the delivery of the exercises: during-game or after-game. In the during-game mode, an exercise popup (see Figure 3.3(c)) appears when the player attempts to collect a star, at which point the player must complete several exercises. If the player attempts to complete the level before collecting the required number of stars, a text banner prompts them to turn around and collect more stars. The after-game condition allows children to play the game normally, with the difference that collecting stars does not trigger speech exercises. The popups are all displayed when the player crosses the finish line, and are similar to the ones in the during-game condition, except that they have a Star Counter so that the player knows their exercise progress (Figure 3.3(d)). After a certain number of exercises, a star and some coins are awarded. Once all the required
exercises have been completed, the player can end the level or continue to practice to earn extra coins. Children’s answers were validated with a Wizard of Oz approach. It is possible to adjust the number of exercises for each level, as well as provide a customized list of words per level, according to each child’s therapy needs.

While testing the game with children from 4 to 12 years old, most of them preferred executing the exercises after completing each level, as it was less disruptive. However, this was not unanimous, which indicates that similar games should offer flexibility in how players can do their speech exercises. Children also liked having perceived control over the game, such as character appearance and exercise behavior.

Apraxia World shows that children are willing to complete speech exercises while playing a game they enjoy and is an excellent example of how gamification can be combined with therapy exercises. However, due to its complexity, the younger children (4-5 years) struggled with some aspects of the game. Particularly, they did not understand the relationship between completing exercises and asset collection. When facing harder sections in the level, they had to make multiple attempts to reach the
end, which resulted in a lot of gameplay time without speech exercises. Alternatives to this problem include starting exercises in a timed manner or incorporate an energy bar which the player can fill by executing more exercises. Some children also reported that they would prefer a game with a storyline.

3.4 VITHEA

VITHEA is an online platform with the intent of supporting the treatment of aphasia, a type of language disorder that occurs after brain injuries. The exercises in the platform are designed with the struggles aphasic patients have in mind, particularly, the difficulty to recall names or words. Although these are not typical exercises for patients with problems in phonological processing, the architecture of the system, including the usage of automatic speech recognition for European Portuguese speech, are extremely relevant.

VITHEA is comprised of two main modules [3], the clinician and the patient module. The clinician module is directed towards the therapist and is divided in three sub-modules. The user management module is where the information of the patients can be viewed and edited. The exercise editor allows to create, edit and delete new exercises and, finally, the patient tracking module allows to monitor the user’s performance, in terms of frequency of access and progress.

The patient module is used to manage the user’s information and to perform exercises, one of which is exemplified in Figure 3.4. The exercises are designed in a way that integrate an animated character that acts as a virtual therapist, text-to-speech synthesized voice, image and video displaying, audio recording, speech validation with automatic word recognition and feedback illustration. When starting a therapy session, the user is guided by a virtual therapist, modelled in 3D and with speech synthesis capabilities. The virtual therapist shows the exercise, which may include the presentation of images, videos, audios or textual information, to which the patient is required to respond verbally by naming the contents of the object or action represented. The answer is recorded, encoded and sent via network to the server side, where it is processed by the ASR system. After comparing the audio file with a set of predetermined textual answers for that question, feedback is sent back to the patient with the correctness of the answer provided.

It is expected that patients have hesitant speech with doubts and repetitions, so the ASR system uses word verification based on keyword spotting, which requires being able to detect if a claimed word is uttered in a given speech segment or not. In the simplest case, a true/false answer is provided, but a verification score might be also generated [46].

The ASR module, named AUDIMUS, was developed for European Portuguese by the HLT research group at INESC-ID. It is a hybrid speech recognizer that combines Hidden Markov Models (HMMs) and multilayer perceptrons (MLPs). A Markov process is used to model the basic temporal nature of the
speech signal. The MLP is used as the acoustic model to estimate posterior phone probabilities given the acoustic data at each frame. The phone probabilities are generated by several gender independent MLPs and later combined [47]. The decoder is based on a weighted finite-state transducer approach to large vocabulary speech recognition. The version of AUDIMUS integrated in VITHEA was trained with 57 hours of downsampled Broadcast News data and 58 hours of mixed fixed-telephone and mobile-telephone data in European Portuguese [46].

The platform serves as complement to conventional speech therapist session by allowing the patients to train at home, thus increasing the number of training hours, and by giving the therapists access to the evolution of their patients.

### 3.4.1 VITHEA-Kids

VITHEA-Kids is a platform based on the infrastructure of VITHEA designed for children with Autism Spectrum Disorder aimed at developing language and generalization skills. The platform allows caregivers to build customized exercises while taking into account specific needs/characteristics of each child. It also has a module for the automatic generation of exercises, aiming to ease the task of exercise creation for caregivers [48].

Unlike VITHEA, it uses multiple-choice exercises, which are often used in applications for children with ASD. These exercises are composed of a question, a stimulus (picture or text), and a set of possible answers (textual or pictures), in which only one of the answers is correct.

Similarly to VITHEA, VITHEA-Kids allows the caregiver to upload and manage image files, as well as create and manage users. Moreover, it allows the caregiver to customize several interaction aspects, such as the utterances performed by the animated character. On the child’s side, each exercise is presented by the animated character featured in VITHEA, which now also provides feedback using
the utterances defined by the caregiver. The exercise area is filled with the stimulus and the possible answers in a random order, and instead of orally naming the stimulus, the child should click the correct answer. Each exercise can contain from zero to four distractors, easing the task of creating several exercises with small variations in content and difficulty. When solving an exercise, if the child picks one of the distractors, it will disappear as a way to prompt the child to pick the correct answer.

The automatic generation of exercises is a key feature in VITHEA-Kids to facilitate the task of creating a new exercise for a specific child or about a specific theme. For instance, if the caregiver asks for a multiple-choice exercise about the topic “Objects”, in which a word has to be illustrated with the correct picture, the exercise generator should return the question, the stimulus, and the set of possible answers. The resulting exercise should be presented to the caregiver, who can choose to accept it as it is, discard it, or change it. The final exercise is then saved to the database and becomes available for the child to solve it.

VITHEA-Kids extends the features of VITHEA and includes exercises specific to Autism Spectrum Disorder with stimuli adapted for children. However, it loses the automatic speech recognition module of VITHEA and does not offer gameplay, which can discourage the child from continuing to use it.

### 3.4.2 VITHEA 2.0

VITHEA 2.0 provides a re-engineered Android mobile application to support the current functionalities of the VITHEA platform, as well as extend its features [49]. VITHEA 2.0 is available for phones and tablets, so it requires the capture of the voice signal through the device’s microphone and the communication with the server where the automatic speech recognition module is located.

Moreover, new features provide patients with a more personalized user experience, including selecting from three different virtual therapists, showing a statistics summary after the end of each round of exercises, and choosing the number of attempts they want per exercise. They can also play back the audio of their responses to each exercise.

On the other hand, clinicians can review their patient’s progress on their dashboard and refine the number of exercises to stimulate improvements in the area where patients might need extra training. To this end, clinicians can send a specific number and category of exercises for their patients to practice.

Altogether, the adaptation of a speech therapy platform with an automatic speech recognition module to a mobile application is interesting and could be applicable to our work.

### 3.5 Discussion

This chapter provided a brief survey of some related work and allowed us to recognize some important features they possess. A summary of the most prominent features of each discussed project is depicted
Table 3.1: Summary of related work features.

<table>
<thead>
<tr>
<th></th>
<th>PA Exercises</th>
<th>Gamification</th>
<th>Progress Monitoring</th>
<th>Exercise Editing</th>
<th>ASR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaisyQuest</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Lexia Core5</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apraxia World</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VITHEA</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

in Table 3.1. First, and something that is common to all approaches, exercises need to be clear and frequently repeated. The child should be able to focus on the exercise presented on the screen without distractions, and the exercises should be explicit and intelligible. Visual and auditory cues also contribute to this, whether they are pictures or spoken words. If the activities require the child to listen to something, the option to listen again should always be available. The exercises’ complexity should be adequate for the child’s capacities, and there should be a large variety of activities, from syllabic to phoneme tasks.

Automatic speech recognition proved to be a reliable way to provide active and personal feedback. One important feature in some of these tools was the ability to keep track of a child’s progression and accuracy in the exercises. This entails the development of a user management tool to integrate into the game, capable of monitoring long-term child progression.

When it comes to game design, children tend to enjoy story-driven games. They also enjoy the ability to personalize the game to their likings. Colorful graphics and visual stimuli are important to keep the child engaged. Moreover, the game’s mechanics should be simple to understand and easy to use, as younger children have some trouble with complex games. Regarding the integration of exercises in the game, this should be done in such a way that it does not disturb the game while the child is playing, but that also does not allow too much game time compared to the number of exercises executed. Thus, the ratio of game time and exercises should be balanced. Children should be able to expect when an exercise will interrupt the game and control when it will happen.
## 4 Implementation

<table>
<thead>
<tr>
<th>4.1 Observations at NCAB Center</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2 Level Selection</td>
<td>32</td>
</tr>
<tr>
<td>4.3 Task Selection</td>
<td>33</td>
</tr>
<tr>
<td>4.4 Game Concept</td>
<td>35</td>
</tr>
<tr>
<td>4.5 Architecture</td>
<td>36</td>
</tr>
</tbody>
</table>
In this chapter, we describe the project’s development process, from the first stages at the NCAB center to the selection and distribution of phonological awareness exercises and, ultimately, the final developed app’s concept and architecture.

4.1 Observations at NCAB Center

Initially, we met doctor Luísa Teles from the NCAB center to discuss possible ways to collaborate on a project that would help children in need of therapy. From there, we were able to partake in some of the therapy sessions as observers to better understand what we were working on.

In the following week, we observed four therapy sessions where only the therapist and the child were present. All but one child were in kindergarten, and all were girls.

One of the younger girls had autism spectrum disorder that originated speech and communication difficulties, so her exercises were more focused on general speech and text interpretation, such as telling a story. The session had plenty of rewards, allowing her to play with playdough or stamps after finishing the exercises.

The two other kindergarten girls had articulatory and phonological disorders and performed different exercises, each according to their struggles. For example, one of the girls struggled to pronounce words with the letter "L", so they played a domino game where each end of a tile had a picture of a word beginning with "L" (Figure 4.1(a)). Each time a tile was played the words depicted had to be pronounced. Concerning phonological problems, some syllable manipulation exercises were enforced, as well as letter and sound identification. Figure 4.1(b) shows an example of an exercise where the child had to point out which one of the given words did not start with the same sound as the others.

(a) Domino with the letter "L".  
(b) Identify the word not starting with the given sound.

Figure 4.1: Examples of exercises used at the NCAB center.
All previous observations were made in regular therapy sessions, where the therapist participated in all the activities with the child. The oldest girl, however, was already in second grade and came to the center after noticing some reading difficulties and possible dyslexia. After an appointment with a doctor, she was forwarded to a meeting with a therapist, where she undertook a thorough evaluation of her skills, including phonological awareness exercises, writing and reading words, pseudowords, and more lengthy texts. The observations were helpful to analyze therapist-patient interaction, and the session evaluating possible dyslexia allowed us to understand how the assessment of reading disorders is usually conducted.

Usually, at the NCAB center, children are encouraged to begin practicing reading exercises, particularly phonological awareness exercises, when they transition from kindergarten to primary school. As we have seen in Chapter 2, phonological awareness at this age is a reliable indicator of possible future reading problems, so this prescription is an effective measure to detect children with difficulties. This is a good place for us to intervene because phonological exercises are fairly simple to implement and execute, which allows for easier data gathering and is a more measurable approach. Our next step comprised of working together with the therapists to decide what exercises we would include and how to adapt them to a mobile format, using automatic speech recognition.

4.2 Level Selection

The first step towards adapting phonological awareness exercises to a game form was to decide which type of exercises to choose and how to categorize them into levels. For this, we relied on the NCAB center’s therapists, who gave us their thoughts on what exercises would be the best and how to vary the difficulty level between tasks. After some adaptations, we arrived at a structure of the following nine levels:

1. Rhyming Tasks
2. Syllable Identification Tasks
3. Syllable Manipulation (syllable deletion exercises)
4. Syllable Manipulation (syllable addition exercises)
5. Syllable Manipulation (syllable substitution exercises)
6. Phoneme Identification
7. Phoneme Manipulation (phoneme deletion exercises)
8. Phoneme Manipulation (phoneme addition exercises)
9. Phoneme Manipulation (phoneme substitution exercises)

Each one of these levels is composed of a few tasks. For example, syllable identification tasks can
require the player to name the number of syllables of a word or say words starting with the same first syllable.

Levels started out continuous, meaning that the player could not proceed to succeeding levels without finishing the first ones. However, a therapist pointed out that children have different interests and capacities, so a child that already mastered rhyming may lose interest by being required to do those tasks. On this account, we made levels 1, 2, and 6 accessible from the beginning so that the player can choose to focus on rhymes, syllables or phonemes. However, syllabic awareness levels (2, 3, 4, 5) and phonemic awareness levels (6, 7, 8, 9) are continuous, so the child still has to progress within different skill levels.

4.3 Task Selection

At each level, the player has to complete a series of succeeding tasks in order to finish it. Tasks are different on each level, both in number and content. In general, task exercises have a question and some answer choices, specifically correct answers and distractors. In every exercise, options have an audio stimulus - the player can hear the word being uttered -, and in most exercises, options are accompanied by a visual stimulus - a picture - as well. Variations of tasks without a visual stimulus (where only the audio is available) are provided to increase the challenge.

4.3.1 Rhyming Tasks

The first level of the game is dedicated to rhyming exercises. The level has six tasks, specified below:

- Identify rhyming word
- Identify rhyming word (no visual stimulus)
- Identify all rhyming words
- Identify all rhyming words (no visual stimulus)
- Identify all non-rhyming words
- Identify all non-rhyming words (no visual stimulus)

All tasks present a word and ask the player to identify another word that rhymes (or not) with the former. In the first two tasks, only one correct answer and one distractor are displayed as possible choices. The remaining tasks have two correct answers and two distractors.
4.3.2 Syllabic Awareness Tasks

4.3.2.A Syllable Identification

There are four levels assigned to exercises with syllables. The first level handles syllable identification within six tasks:

- Identify the number of syllables
- Identify the number of syllables (with no displayed choices)
- Identify words with the same first syllable (3 choices)
- Identify words with the same first syllable (5 choices)
- Identify words that start with a target syllable
- Identify words that start with a target syllable (no visual stimulus)

The first two tasks ask for the number of syllables of a word. However, while the first task displays two possible answer choices - one correct answer and one distractor-, the second task does not show any possible choice. The player must identify the number of syllables of the word using only its synthesized audio.

The third and fourth tasks ask the user to identify, from within a group of words, the ones starting with the same syllable, showing three and five possible choices to choose from, respectively.

Lastly, the final tasks present a specific syllable’s audio and ask the user to identify two words starting with the spoken syllable while avoiding the distractor.

4.3.2.B Syllable Manipulation

The other three levels are dedicated, respectively, to syllable deletion, addition, and substitution.

Syllable deletion tasks are distributed into the deletion of the last, first, and middle syllables. For each of them, we have two tasks, for a total of six tasks. The only difference between them is the number of possible choices - one task has three choices (one correct answer, two distractors), and the other has five choices (one correct answer, four distractors).

Both syllable addition and substitution exercises are divided into first and last syllable addition and substitution, respectively. They also have tasks with three or five choices, for a total of four tasks on each level.
4.3.3 Phonemic Awareness Tasks

Phonemic awareness tasks are similar to syllabic awareness tasks. The number of tasks is the same, and the only difference is that exercises pertain to phonemes instead of syllables. Thus, the first level is dedicated to phoneme identification, including tasks to identify the number of phonemes or words starting with the same phoneme.

The remaining levels are assigned to phoneme deletion, addition, and substitution. These also have the same structure as syllable manipulation tasks.

4.4 Game Concept

Given the phonological awareness exercises’ distribution into levels, we started to envision the game’s overall concept. The developed app is targeted at children who are finishing kindergarten and entering primary school (about 5 years old), although any child can use it as a means to train phonological awareness.

The app offers a series of phonological awareness exercises such that they become increasingly more difficult. At each level, the child has to complete a series of tasks to master it. After completing the exercises required in a task, the player is allowed to proceed to the next one and receives a puzzle piece as a reward. For each level in the game, there is one puzzle to be assembled. Each task rewards a different puzzle piece so that when every task is completed, all puzzle pieces for that level are unlocked. Once unlocked, the player can assemble all the pieces, unveiling a picture. As an additional reward for completing the level, the player can color in the revealed picture.

Exercises require the player to answer using their voice by recording it with a microphone. Hence, the app uses an automatic speech recognition system to validate the response. A text-to-speech system is also used to synthesize audio for the questions and words displayed. Children can hear the synthesized speech any time they want.

The app also includes a monitoring tool. While performing exercises, if the player gives a wrong answer, the exercise is recorded, and it is possible to playback the recorded audio. An overall user’s progress and performance reports are available, including the exercises that were failed more often. Although this does not give a final assessment of the user’s capabilities, it can be an indicator for therapists on the child’s struggles and overall skill level.
4.5 Architecture

The app was built for Android using Unity, which allowed access to extensive resources from the Asset Store and the integration of external tools like the ASR and TTS systems. Figure 4.2 shows the overall system architecture, including the resources used and the interaction between client and server. The specified architecture of the system’s TTS and ASR requests are described below, together with the app’s file system.

```plaintext
Figure 4.2: General system architecture.
```

4.5.1 Game files

Game files that include user data or general game data are stored locally, and binary encoded.

All exercises are specified in Extensible Markup Language (XML) files. Listing 4.1 shows an example of a rhyming exercise represented in XML form. These files specify every word used in the exercise, both answers and distractors, together with the rhyming type (for example, words rhyming with “ão”). In syllable and phonemic awareness exercises, the same is applied, although more information is necessary to report the number of syllables/phonemes and the syllable/phoneme manipulation exercises (the syllable to remove/add/substitute and the corresponding result).

Each exercise can be used for multiple tasks. Following the Listing 4.1 example, the exercise can be selected for any of the six rhyming tasks by selecting specific information from the file. For example, if the chosen task has 2 distractors, both distractors specified in the XML file will be selected, albeit in a random order. However, if the task only requires one distractor (the first rhyming task, for example), then only one of the two options is randomly selected and displayed. This random selection of specific information allows for some different alternatives to the same type of exercise, promoting dynamic exercises to avoid repetition.
Listing 4.1: Example of a rhyming exercise in XML form.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<exerciseCollection category="rhyming">
  <rhEx number="1">
    <rhymetype>ol</rhymetype>
    <rhymes>
      <rhyme>Caracol</rhyme>
      <rhyme>Sol</rhyme>
      <rhyme>Farol</rhyme>
    </rhymes>
    <distractors>
      <distractor>Bolacha</distractor>
      <distractor>Bola</distractor>
    </distractors>
  </rhEx>
</exerciseCollection>
```

4.5.2 Text-to-Speech

When an exercise is loaded, it is necessary to synthesize a speech of the question and all the possible answer choices.

Initially, we used Google Cloud’s TTS because of its speech synthesis quality. However, some of our tasks require the synthesis of isolated syllables or phonemes. For example, phoneme substitution exercises require the child to hear the phoneme to substitute. To this end, we needed to use Speech Synthesis Markup Language (SSML), which gives much freedom to manipulate synthesized speech, including the option to specify phonemes. Although Google’s TTS uses SSML, it does not support this option, so we resorted to using Amazon’s TTS. Amazon Polly also uses SSML and supports phoneme specification, using the International Phonetic Alphabet (IPA) or the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA).

![Architecture and resources used in a TTS request.](image)

**Figure 4.3:** Architecture and resources used in a TTS request.

Figure 4.3 depicts a request to Amazon Polly, where an SSML message is built from the exercise
data, specifically any word, syllable, or phoneme that needs to be spoken, together with the user’s TTS settings. An HTTP request is sent to the server with the SSML message, and the synthesized speech is returned. The audio is then stored in the app’s resources so that when the user wants to hear the word again, there is no need to perform another request to the server. When an exercise is loaded, existing audio files are replaced by new ones.

4.5.3 Automatic Speech Recognition

Automatic speech recognition was used to validate the user’s answer. We used the AUDIMUS ASR system, which is described in detail in Section 3.4. AUDIMUS uses keyword spotting, so it detects whether a word was uttered in the provided audio. Figure 4.4 shows the system when performing an ASR request. The microphone records the player’s answer, and the audio clip is sent to the server, together with the GRXML that contains the expected word. If there are several correct answers, the GRXML contains all the possible answers and their combinations. The ASR system responds with the recorded audio’s assessment, stating whether it contains any of the expected words or not, which is then conveyed to the user.

Figure 4.4: Architecture and resources used in an ASR request.
5 System Overview

Contents

5.1 Character Creation and Personalization .................................. 41
5.2 Levels and Tasks ............................................................. 42
5.3 Exercises ................................................................. 43
5.4 Puzzles ................................................................. 46
5.5 Picture Draw .............................................................. 46
5.6 Progress Monitoring ....................................................... 47
This chapter describes all of the app’s functionalities and some of the system’s logic. Overall, the app comprises levels and tasks, exercises, puzzles, picture drawing, and progress monitoring. Figure 5.1(a) shows the app’s title screen. On startup, the game will always load the last selected user, so the screen greets the last player. Cheerful music starts playing that, together with the vibrant colors used throughout the game, serve as a way to draw the child’s attention.

![Title screen and settings.](image)

**Figure 5.1:** Title screen and settings.

### 5.1 Character Creation and Personalization

#### 5.1.1 Character Creation

The first thing to do when a new player launches the app is to create their character. It is not possible to play the game without at least one character created in the app. The character screen title displays all the available characters, where it is possible to click and select the one to play with (Figure 5.2(a)). To create a new character, the user can select the plus button, which will display a new screen. Here, the player inputs his/her name, age and selects his/her character’s avatar (Figure 5.2(b)). After creating the character, the game is available to play. It is also possible to edit the player’s information or delete it entirely.

#### 5.1.2 Settings

In order to increase the game’s personalization, a few settings were made available to the user, presented in Figure 5.1(b). It is possible to (a) toggle the automatic play of the question’s synthesized text-to-speech audio when loading a new exercise; (b) adjust the reading speed; (c) change the number of exercises per task; (d) change the number of tries per exercise; and (e) adjust the music volume.
5.2 Levels and Tasks

The game is composed of a series of levels, where each level works on a specific phonological awareness skill. The structure of the levels is explained in chapter 4.

Figure 5.3(a) shows the screen with all the levels. When a level is unlocked, the player can access the required tasks for that level, depicted in Figure 5.3(b). The tasks are outlined on the screen and are continuous. When players complete all exercises required in the task, they unlock the next one and receive a puzzle piece as a reward. The level is finished when all tasks have been unlocked and completed, thus obtaining all the puzzle pieces, which allows the player to complete the puzzle for that level.

Each level has a theme, expressed in the background of the tasks screen. For example, Figure 5.3(b) shows the tasks for the first level on the game, displayed in what looks like a classroom environment. The picture assembled by the puzzle pieces is also related to the theme; in this case, the image is a school bus. Other levels have a space theme, or a dinosaur theme, for example.
5.3 Exercises

Depending on the level, exercises can be of different types: rhyming, syllabic awareness, or phonemic awareness. However, all exercises have three aspects in common:

- **A question**, which may ask to identify words with the same first syllable or words starting with the same phoneme, for example. This is usually on the left of the screen.

- **Answers** to the question and distractors, in the middle of the screen.

- **A record button**, using a push-to-talk method to record the user’s answer, on the right of the screen.

All exercises require the user to vocalize the answer into the microphone. Figure 5.4 depicts an example of an exercise, with all the aforementioned elements present.

![Figure 5.4: Rhyming identification exercise.](image)

Answers and distractors are usually accompanied by a visual stimulus. Clicking on any of the pictures on the screen triggers a speech synthesis of the word. This way, children have a visual and audio representation of the word available. A written form of the word is never displayed, in order to compel children to derive the answer from the sounds heard, rather than the order of the letters or memorization, especially with children who have already learned to identify and read/write letters and words. It is also possible to hear a speech synthesis of the question. All speech synthesis can be replayed any time the user wants.

The displayed choices (the correct answers and the distractors) are always randomized, so they always appear in a different order. This was also pointed out by the therapists as a way to avoid children memorizing the exercises and the positions of the answers.
To answer a question, the users need to push the record button as they speak and release it when they are done. When the record button is pushed, a three dotted animation starts, indicating that the microphone is recording. If the button is not pushed for a long enough time (i.e., it is clicked instead of held), the animated dots become red to signal it to the user, and thus no request is sent to the ASR system.

There may be exercises with several correct answers. In this case, answers can be given in any order; for example, if there are three correct answers, the user can say one of them for each push of the button or say all three, at any order, to the microphone at the same time. If the speech recognition system does not recognize the spoken word, the background turns red for a few seconds. If the answer is correct and recognized by the speech recognition system, the images corresponding to the guessed words are highlighted in green. After a determined number of failed attempts, another exercise is loaded so as not to increase the child’s frustration.

As a first effort to populate the app with exercises, a total of 126 exercises were created. The list comprises rhyming exercises (20), syllabic awareness exercises (71), and phonemic awareness exercises (35). The visual stimuli for the words amount to 156 pictures.

### 5.3.1 Rhyming Exercises

As previously mentioned, some exercises do not have a visual stimulus, requiring the player to touch the screen to activate the speech synthesis of the word. Figure 5.5(a) depicts an example of the second rhyming task, where this is applied. Figure 5.5(b) shows an example of an exercise to identify all non-rhyming words.

![Figure 5.5: Examples of rhyming tasks.](image)
5.3.2 Syllabic Awareness Exercises

Figure 5.7 portrays an example of a syllable identification exercise. In this case, the child has no possible choices displayed on the screen, so they have to intuitively attain an answer by themselves.

The 71 syllabic awareness exercises created include 30 syllable omission exercises (10 for each start/middle/end syllable), 20 syllable substitution, and 20 syllable addition (10 for each start and end syllable).

![Figure 5.6: Example of an exercise to identify the number of syllables with no displayed options.](image)

5.3.3 Phonemic Awareness Exercises

In Figure 5.7(a), we see an example of a last phoneme deletion exercise and a correct guess by its side.

The 35 phonemic awareness exercises created include 15 phoneme omission exercises (5 for each start/middle/end phoneme), 10 phoneme addition (5 for each start and end phoneme), and 15 phoneme substitution (10 for start phoneme and 5 for end phoneme).

![Figure 5.7: Example of a phoneme omission exercise.](image)
5.4 Puzzles

Each level has a corresponding puzzle, so there are nine puzzles in total. Each puzzle has as many pieces as the number of tasks within that level, for example, the rhyming level has six tasks, thus the puzzle corresponding to the first level has six puzzle pieces. At the beginning of the game, all puzzle pieces are locked and can’t be moved. Whenever the player completes a task, a puzzle piece is unlocked. When all pieces are unlocked, the player can drag them to their rightful place in the puzzle and reveal its image. The puzzle image is a black and white coloring picture, so that after puzzle completion, players gain access to the picture drawing screen, where they can color in the picture.

![Figure 5.8: Puzzle with two puzzles pieces unlocked and rightfully assembled.](image)

5.5 Picture Draw

Along with completing the puzzle, players also get the chance to color in the picture assembled by the puzzle, shown in Figure 5.9. When clicking on the brushstroke, a color picker appears where the user can pick any color they want. It is also possible to control the brush-width, brush transparency, and to click on the rubber to delete any drawings performed. Users can simply drag their finger or use a special pen for mobile devices to draw on the picture.
Because coloring in the picture digitally may not be as fun as coloring by hand, a button (the middle one on the top left) triggers a pop-up asking for an email. Upon providing an email, the system sends the picture to the specified address so that the therapist or monitor can print it, and the child can color it by hand.

![Figure 5.9: Picture drawing screen.](image)

5.6 Progress Monitoring

The app provides a long-term track of the child’s progress. When players finish an exercise (either by getting the correct answer or by reaching the maximum number of attempts at answering), the system checks if there were any wrong attempts. If so, the exercise is reported, together with the audio answers said by the player. This will later help us determine if there is a certain type of exercise that is repeatedly failed. User progress is kept in terms of level, task, and exercise, as follows:

- **Level Progress** - The first information provided by the progress monitoring module is the percentage of each level completed, as shown in Figure 5.10(a).

- **Task Performance** - Within each task, it is possible to examine which type of exercises the user struggled with the most. For example, assume that the most failed exercises in a rhyming task were the ones with words rhyming with “ão”. Consequently, the type “ão” would show up as the type of exercise with the most mistakes for that task. This way, for each task, we know the top three types of exercises with the most mistakes, as depicted in Figure 5.10(c).

- **Exercises History** - For each exercise type within the top three, it is possible to access all completed exercises of that type. Following the previous example, it would be possible to see all the
exercises finished of the type “ã” and the number of wrong answers given, as well as hear their recorded answers. An example of this is shown in Figure 5.10(d).

(a) Levels progress.  
(b) Rhyming level’s tasks.  
(c) Task performance.  
(d) Exercises history.

Figure 5.10: Progress monitoring screens
6 Evaluation

Contents

6.1 Evaluation With Children .................................................. 51
6.2 Measuring Usability .......................................................... 52
6.3 Methods ........................................................................ 52
6.4 Results ........................................................................ 54
6.1 Evaluation With Children

It is a common practice to use tasks during the evaluation of a system, as it ensures the software’s critical features are addressed. However, requiring children to follow a set of tasks may influence their behavior. By providing them with a specific goal, children may try to accomplish the task even without being intrinsically motivated by the game or even stray from it because other goals provided by the game appear more interesting [50]. Furthermore, assigning tasks can give away information about the game; for example, if the task asks the user to find level three of the game, it can give away the fact that there is a level three in the game (and possibly levels one and two) that can be played.

In fact, Barendregt [50] conducted a study to compare two groups of eight and nine-year-old children, one group performing tasks and one group playing freely for fifteen minutes. The task’s influence was measured by the number of screens visited, the number of positive and negative verbal and non-verbal indications, the number of problems found, and the number of times the children asked for help. Although there were no significant differences in the number of verbal and non-verbal indications and the number of problems, there was a notable difference between the numbers of screens children visited in the task condition and the free-play condition. This could be because children in the task condition tend to browse a screen to see if they could fulfill the given task. If the screen did not seem to be the right one, they quickly tried another one. On the other hand, children in the free play condition could take more time to look for interesting things on each screen. The number of times children asked for help was also different. While most children asked for help between two and three times using the tasks method, none of the children asked for help in the free-play condition. This could be explained by the fact that tasks make children more aware that they are in a test situation and feel the need to perform well, even if they are not engaged.

By giving tasks, children receive additional external motivation, which may not represent the motivation provided by the game. Therefore, it is no longer possible to reliably detect real motivation problems. Conversely, the observational method is often used when testing software with children ([51], [52], [53]). In this case, children are observed while using the product for a brief period of time, and their reactions and comments are registered.

Another technique also commonly employed to discover problems while testing a product is the think-aloud method, where participants are asked to provide a running commentary of their thoughts and feelings while performing their tasks. However, applying this method with children can be a problem because they can have difficulty verbalizing their thoughts while talking to no one in particular [54]. Moreover, children tend to forget to think aloud and need to be reminded frequently to keep talking, which could result in children mentioning non-existing problems in order to please the observer. [52] To bypass this issue, we can instruct children to think-aloud in the beginning of the evaluation, but refrain to remind them throughout the test, together with observing the child’s behavior.
6.2 Measuring Usability

Citing ISO 9241-11 [55], usability is referred to as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". More precisely:

- **Effectiveness** - The accuracy and completeness with which customers achieve specified goals.

- **Efficiency** - The accuracy and completeness of goals achieved in relation to resources.

- **Satisfaction** - Freedom from discomfort and positive attitudes toward the use of the system.

Due to their objective nature, both effectiveness and efficiency can be measured using the user's success rate, defined as the percentage of tasks that users complete correctly [56]. In our case, user success can be derived from an observation checklist which is used by the author while performing the usability test. The user satisfaction was deduced from a post-session survey that evaluated fun, complexity, user affinity with the character and exercises, and interest in using the application again.

6.3 Methods

Initially, we intended to evaluate the app with children from the NCAB center. However, due to time, pandemic, and bureaucratic constraints, we decided to perform the tests with voluntary participants. Despite these constraints, we still devised a study protocol to submit to the center, which can be consulted in Appendix A. The study with voluntary participants was submitted and approved by the IST Ethical Committee with number 23-2020; the documentation for this can be consulted in Appendix B.

6.3.1 Subjects

We performed tests with 2 girls and 4 boys, totaling 6 children. Two of them were 6 years old (first-graders), and the remaining four were 4 years old, thus still in kindergarten. Of the demographic data collected from the parents, all considered their child to have a good or reasonable speech capacity, but a poor or very poor reading skill. Three of the children rarely or never used mobile devices. The remaining three used them sometimes, and either rarely, sometimes, or frequently used them to play games in particular. None of the children had a diagnosed speech disorder, although two of them attended speech therapy appointments to correct some speech production deficits. The main demographic characteristics of the 6 participants are summarized in Table 6.1.
Table 6.1: Summary of the participant’s demographic information.

<table>
<thead>
<tr>
<th>Demographic Element</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Sex</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Frequency using mobile devices</td>
<td>Never</td>
<td>Never</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Sometimes</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Frequency using mobile devices to play games</td>
<td>Never</td>
<td>Never</td>
<td>Never</td>
<td>Frequently</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Reading skill level</td>
<td>Poor</td>
<td>Good</td>
<td>Very Poor</td>
<td>Reasonable</td>
<td>Poor</td>
<td>Very Poor</td>
</tr>
<tr>
<td>Speaking skill level</td>
<td>Good</td>
<td>Good</td>
<td>Reasonable</td>
<td>Reasonable</td>
<td>Reasonable</td>
<td>Good</td>
</tr>
<tr>
<td>Diagnosed speech disorder</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

6.3.2 Procedure

Due to constraints caused by the pandemic, the tests were carried out on a location of the participant’s choice, provided that it was in a calm environment. In two cases, the test was conducted in a park, and the remaining tests were in a chosen indoor location. All the tests were applied by only one person (the author) who collected and processed all data and were performed individually, with the child’s parents or an adult responsible for the child present. Sessions lasted approximately 30 minutes and were administered on a tablet provided by the author.

The session started with collecting the informed consent (see Appendix C), together with some demographic information on the child. The test was then explained to the child, who was told that he/she would have 15 minutes to engage with the app in any way he/she wanted. To avoid pressuring children to play the game, care was taken to explain that they were not obliged to play and that they could stop playing the game at any time they desired. Participants were also told not to be afraid to speak their minds, although this was not emphasized throughout the session, as discussed in section Section 6.1. During the session, parents were asked to refrain from interfering except if it became indispensable. Due to the test’s time limit, participants were only able to play though the first level, which, as defined previously in Section 4.2, is focused on rhymes.

Although the app was developed considering young children’s handling, it was not thought out to be used completely independently, but with a parent or therapist, especially at first usage. For example, there are no tutorials for the exercises, so the push-to-talk button for the microphone needs to be explained, and some menu options have only text, making it hard for children who do not know how to read. Similarly, creating a character requires the child to write a name and age while also selecting buttons with some text. Because we do not intend to evaluate the app’s independence on first use, the author performed a brief demonstration of the app before the child started playing. Creating a character, which is a first and only time step, was the only instance in the session where help was provided freely. The demonstration showed the puzzles and explained how to answer exercises, never telling the child what they needed to do, instead only presenting the game’s possibilities.

While the children interacted with the app, the author observed and registered their behavior and actions. The observation checklist was created considering some goals that should be achieved to
evaluate effectiveness and efficiency (depicted in the tables alongside the results). For each element in the list, the author noted if the goal was achieved (Y) or not (N). Alternatively, goals can also be partially achieved (P); the criteria for this was left at the author’s discretion.

Once the 15 minutes ended, the child stopped playing with the app and was asked a brief questionnaire to assess user satisfaction (also depicted in the results) using a five-point Likert scale and the help of a Smileyometer [4] to facilitate children's expressions of thought. Finally, the session ended.

![Figure 6.1: Smileyometer as defined by Read et al. [4]](image)

### 6.4 Results

Overall, all children responded well to the game. Tests were harder with children who did not have experience with mobile devices, especially younger children. Most problems with these children were caused by difficulties using the touch screen, i.e., clicking and dragging. Nevertheless, as time went by, they started to learn and automatize the movements. The tests were also very dependent on the child’s personality - some children were shyer and needed more time to feel comfortable using the app. In contrast, others were very impulsive and wanted to click everywhere. Younger children had not yet learned rhymes, which made it harder when playing. A brief explanation was given to them before starting the test.

In the following sections, the effectiveness, efficiency and satisfaction results will be presented, followed by the author’s observations and discussion.

#### 6.4.1 Effectiveness

The observation checklist for effectiveness was established with general goals for children to achieve in situations that would naturally occur in the game. Table 6.2 shows the attained results for each child. A goal was considered achieved if the child was able to perform it. Sometimes, a goal was partially achieved, which in most cases meant that although the child struggled with the task, over time and with the author’s help, the performance improved. Tasks were deemed unsuccessful if the child could still
Table 6.2: List of tasks for the effectiveness evaluation.

<table>
<thead>
<tr>
<th>Evaluation Element</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Character Creation Screen</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child was able to pick their desired character’s avatar</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Exercise Screens (Levels, Tasks and Exercises)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child was able to navigate the exercise screens</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
</tr>
<tr>
<td>The child knew what to do during the game play</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was focused on the exercises</td>
<td>Y</td>
<td>P</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child showed positive reactions during the game play</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to choose a level</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to choose a task in the level</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child understood how to listen to the question and answer options</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child knew how to use the answer button</td>
<td>P</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
</tr>
<tr>
<td>The child knew how to answer the exercises</td>
<td>Y</td>
<td>P</td>
<td>N</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child understood if their answer was right or wrong</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to answer an exercise</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to complete a task</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Puzzles Screens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child was able to navigate the puzzles screens</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
</tr>
<tr>
<td>The child understood how to unlock a puzzle piece</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child understood how to complete the puzzle</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to assemble a puzzle piece</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><strong>Drawing Screens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child understood the relation between the puzzles and the pictures</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child understood how to draw on the picture</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to draw on the picture</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child was able to change the color, transparency or width of the brush</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
</tr>
</tbody>
</table>

not perform after several attempts and the author’s assistance. The participants’ behavior regarding the table’s evaluation elements will be further discussed in Section 6.4.4.

In Table 6.2, we have 21 tasks and 6 children, for a total of 126 tasks for all children. There were 106 successful tasks and 18 partially successful tasks. Considering the user’s success rate, we can calculate the effectiveness with the following equation:

\[
\text{Effectiveness} \% = \frac{(1 \times Y \text{ tasks}) + (0.5 \times P \text{ tasks}) + (0 \times N \text{ tasks})}{\text{Total tasks}} \times 100
\]

\[
= \frac{(106 + (0.5 \times 18))}{126} \times 100
\]

\[
= 91.27\%
\]

From the previous equation, we can estimate our game’s effectiveness to be approximately 91%.

### 6.4.2 Efficiency

Table 6.3 shows the conceived observation table for the efficiency evaluation. The criteria for this table is similar to the previous effectiveness table. This time, we have 12 tasks and 6 children, for a total of 72 tasks for all children. There were 50 successful tasks and 7 partially successful tasks. Invoking the
user’s success rate, we can calculate the following:

\[
\text{Efficiency} (\%) = \left( (1 \times Y \text{ tasks}) + (0.5 \times P \text{ tasks}) + (0 \times N \text{ tasks}) \right) \div \text{Total tasks} \times 100
\]

\[
= (50 + (0.5 \times 7)) \div 72 \times 100
\]

\[
= 74.31\%
\]

From the equation above, we can estimate the efficiency for our game to be approximately 74%.

Table 6.3: List of tasks for the efficiency evaluation.

<table>
<thead>
<tr>
<th>Evaluation Element</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character Creation Screen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child created his/her character on the first try</td>
<td>Y</td>
<td>P</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Exercise Screens (Levels, Tasks and Exercises)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child answered an exercise on the first try</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>The child completed a task on the first try</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>The child’s errors and mistakes were minimal</td>
<td>P</td>
<td>P</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
</tr>
<tr>
<td>The child knew how to recover from errors</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Help from the author was minimal</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
</tr>
<tr>
<td>Puzzles Screens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child managed to assemble a puzzle piece on the first try</td>
<td>Y</td>
<td>Y</td>
<td>P</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child’s errors and mistakes were minimal</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child knew how to recover from errors</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Drawing Screens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The child managed to draw on the image on the first try</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>The child’s errors and mistakes were minimal</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>The child knew how to recover from errors</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

6.4.3 Satisfaction

Table 6.4 depicts the participant’s ratings obtained from the post-session questionnaires. Questions were answered following a 5-point Likert scale, where 1 was the lowest rating and 5 the highest rating possible. We can calculate the satisfaction following a similar approach to the previous calculations. Considering the total points of the questionnaire as the highest rating possible for all participants in all questions we have:

\[
\text{Satisfaction} (\%) = \frac{\text{Number of points rated}}{\text{Total points}} \times 100
\]

\[
= 191 \div 210 \times 100
\]

\[
= 90.95\%
\]
From the previous equation, we can estimate the satisfaction of our game to be approximately 91%.

Table 6.4: Satisfaction results.

<table>
<thead>
<tr>
<th>Question</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>The game was fun</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>The game was easy</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>I liked being able to choose my character</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>I liked being able to answer with my voice</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>I liked the puzzles</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>I would like to play again</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>I would rather perform the exercises on the app than on paper</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

6.4.4 Observations and Discussion

Children were generally focused throughout the duration of the test and showed positive reactions. Particularly, one of the participants always celebrated when knowing the answer ("This one rhymes!"); another rejoiced when unlocking a task ("I won number 2!" when unlocking the second task). All of them enjoyed unlocking the puzzle pieces ("There is only 3 to go!").

Screen navigation was very instinctive for children familiar with mobile devices, but those who did not have the same experience did not find it intuitive. Examples of this are the arrows to choose the character’s avatar, the exit buttons, and failed attempts to click on the levels for lack of practice with a touch screen. Nonetheless, by the end of the session, they started to get better at it, and screen navigation became more automatic.

As expected, understanding the goals of the rhyming exercises was easier for children who already knew how to rhyme. The biggest problem with the exercises, prevalent in almost every test, was the use of the answer button. Children made several different mistakes when trying to answer: only clicking the button or pressing it for too little time, speaking and then pressing the button, releasing the button while talking, or even forgetting to let go of pressing the button. Some children needed more help than others and constant reminders throughout the test on how to use the button, but in general, everyone improved over time. In the future, it would be of interest to create a tutorial just to practice using the answer button.

Some more impulsive children tried to answer the question without hearing the options audio, taking only the images into account. This brought up some situations where the child would assume the option to be something different (for example, assume that an option was "snake" when it was "serpent"). As recommended by a therapist, some levels did not have images and it was necessary to hear the option’s audio to know what it was ("I can only see hands", said one child when confronted with this type of exercise). This proved to be a good method to teach the children to hear the possible choices first. It might be interesting to start levels with a task that does not have pictures to teach them to hear the options and only then introduce levels with images. A tutorial can also be developed to deal with this
One task required users to find all the words that rhymed with the target word. Previously, participants had been executing tasks that only asked for one word, so when this task came along, they did not understand why the exercise was not over after saying the first word (many saying "I'm done" after discovering the first answer). To moderate this issue, it might be good to try to rewrite the question, emphasizing the existence of multiple answers, or somehow insinuate this on the user interface.

Children who had a harder time understanding how the exercises worked were usually more impulsive and behaved by pressing the answer button as soon as they heard the question and saying the question back aloud, naming all the options, or saying random words. One child in particular seemed to only care for the puzzles ("There are so many puzzles!", "I wanna do the yellow one now") and needed reminders that he was required to complete the levels first. This behavior might have been encouraged by a seeming lack of understanding of the exercises, reinforced by the young age, lack of familiarity with mobile devices, and shortage of rhyming knowledge. However, another impulsive child that had experience with mobile devices seemed to understand the game well and even took initiative when not sure of the answer ("I think these two rhyme. Let's see, let's see", followed by attempting to answer one of the options). On the other hand, shyer children tended to look at the author to ask if the answer was right and needed to be encouraged to answer to the game. The older participants (P1 and P5), however, hardly needed assistance and, in fact, showed the best results and gave the best rating to the game. Although some of the 4-year-old children were almost 5-years old, it would be best to proceed with testing amidst 5-year-old children to get a better understanding of the age and rhyming knowledge's influence.

The puzzles were very much enjoyed by all children and seemed to serve as good incentives. All children were interested in finishing them and were excited to assemble every new piece ("We have to find all others!"). Interestingly, all children preferred to assemble a piece as soon as it was unlocked, rather than keep playing and assembling all pieces at the end of the level. The picture drawing screen was not showed during the author's initial brief demonstration, so it came as a surprise to them. Most were confused, as they expected to keep playing to unlock more puzzle pieces. To some, trying to draw was intuitive ("It's to draw, I knew it!", "Now it's telling me to draw") but others (again, usually children unfamiliar with mobile devices) needed some incentive to try touching on the picture. Changing the color of the brush was also not intuitive to most. Even after understanding that they could draw, some children preferred to keep playing ("Why do I have to draw on the puzzle now?"), which implies that puzzles were a stronger motivation to play. The drawings, however, are optional and some children seemed to enjoy them.

Ultimately, all children were successful using the app, which is reflected in the effectiveness results, but efficiency can still be improved, starting with addressing the previously described issues. Neverthe-
less, the satisfaction results were very good, and all children manifested the desire to keep playing when
the time was up.

With every test, we observed a learning curve: even though some children did not do very well
initially, they would gradually learn the game’s mechanics and, by the end, were already a lot more
proficient than at the start. This makes us believe that effectiveness and efficiency would improve over
the limited time given for the child to be acquainted with the app. In the future, it would be interesting to
perform tests twice with the same children: a first one for the child to get acquainted and learn the game
with some help, and another one some time later to evaluate the game, when children are more familiar
with it.

From the above discussion, one can notice that some factors proved very relevant to justify some
of the observed differences between participants, namely, experience with screen navigation in mobile
devices, and familiarity with rhymes, the latter being mostly present in the older children. Probably due
to the very limited number of participants, no marked differences in behaviour were observed between
male and female children.
Contents

7.1 Conclusions ................................................. 63
7.2 Future Work ................................................ 64
7.1 Conclusions

Early childhood years are critical to the development of language. Reading problems are present in most children with learning disabilities; however, they often remain undetected. Although many centers such as the NCAB center manage reading disabilities, intervention is still needed in late kindergarten or early primary years. To this end, we developed an app, together with therapists of the NCAB center, that provides phonological awareness exercises integrated with automatic speech recognition technologies.

Our project was initiated at the NCAB center, where we were able to observe some therapy sessions and understand our research's current situation and objectives. Subsequently, we studied the background and the existing work related to our project, presented in Chapter 2 and Chapter 3.

Along with the therapists, we defined adequate levels and tasks for our game. Our app provides several tasks to practice phonological awareness, including rhyming, syllables, and phonemes exercises. Text-to-speech was used to generate synthesized speech for the exercises. Exercises are answered using the child's voice, through the device's microphone, and sent to an automatic speech recognition system to evaluate it.

The app also allows players to unlock puzzle pieces, assemble puzzles, and draw over the revealed pictures. Finally, there's also a progress monitoring feature, where therapists can review children's performance, including the exercises they struggle with most.

Our next step was to evaluate the app with young children concerning its efficacy, efficiency, and satisfaction. The game was well received with all of the children involved in the evaluation process. Although younger children and children more inexperienced with mobile devices need some time to learn the game's mechanics, all children enjoyed playing, which was reflected in the satisfaction results we obtained from them.

The development of this project did not come without challenges. Communication with the NCAB center's staff was not always easy, although their collaboration in the midst of their very busy schedule was very much appreciated. Our initial plan involved evaluations with children attending the center. The pandemic was a huge setback, initially involving long delays and later rendering this evaluation totally impossible, so alternative evaluation plans had to be made. Nonetheless, despite these challenges, the main objectives of this work were fulfilled and results are considered very satisfactory.
7.2 Future Work

The developed app can still be extended and enhanced; some possibilities are described below:

- **Populate the app** with more exercises;

- **Create new levels and new tasks**, for example, syllable segmentation exercises or order syllables to create a word;

- **Further personalization** to make the game more appealing and adaptable to each child, for example:
  - Allow children to pick the app's color scheme, music or font;
  - Create a module that allows therapists to select certain exercises for a child to perform. This way, children can practice the exercises they struggle with the most;

- **Make the app friendlier** to children using it for the first time, including, for example, TTS for all the buttons, create tutorial levels and help guides;

- **Develop a module to create new exercises** and edit existing ones supporting the upload of pictures;

- **Incorporate the app's architecture within a server**, which would mean all of the game, player, and exercise data would be stored in an online server. Although this could prompt slower response times caused by the need to communicate with the server, the information could be safely stored in the server and would save much space on the device;

- **Further extend the app to run on other operating systems**, particularly iOS.

- **Further testing**:
  - Test with more children;
  - Test with therapists, including the progress monitoring module, which was not evaluated, and assess the exercises' adequacy.
Bibliography


Study Protocol for Testing at NCAB Center
1. Identificação do projeto

a) Título do projeto:
Tecnologias da fala aplicadas à terapia das perturbações do desenvolvimento da linguagem

b) Autores/Promotores:

a. Promotor (Indivíduo ou entidade responsável pela execução do estudo):
INESC-ID, Human Language Technology (HLT)

b. Investigadores:

Prof. Doutor Alberto Abad – Instituto Superior Técnico, INESC-ID
Prof.ª Doutora Isabel Trancoso – Instituto Superior Técnico, INESC-ID
Doutora Luísa Teles – Centro de Neurodesenvolvimento e Comportamento da Criança e do Adolescente
Matilde Ramos – Aluna do Mestrado Integrado em Engenharia Informática e de Computadores do IST, INESC-ID

c) Natureza do estudo:
Estudo observacional

d) Local onde decorre o estudo:
Centro de Neurodesenvolvimento e Comportamento da Criança e do Adolescente

e) Descrição sucinta dos objetivos da investigação
Esta investigação pretende avaliar a usabilidade e utilidade de uma aplicação desenvolvida para crianças, cujo foco é a resolução de exercícios de consciência fonológica através da fala. Para tal, será feita a gravação áudio das respostas aos exercícios através de um microfone, as quais serão avaliadas através de um sistema de reconhecimento de fala automático. Os participantes irão ser observados a utilizar a aplicação durante algum tempo e irão, de seguida, responder a um breve questionário.
f) **Confidencialidade nos registos**

As folhas de notação e os resultados das avaliações ficarão armazenados num local com acesso limitado à equipa de investigação. Caso a informação seja guardada num servidor público, ficará num documento com acesso restrito protegido por palavra-chave que será de conhecimento exclusivo dos investigadores. Toda a informação será tratada de forma confidencial, de forma a poderem ser efetuadas análises estatísticas anonimizadas.

2. **Justificação científica da investigação**

As dificuldades de leitura tendem a surgir nos primeiros anos de escola das crianças e têm um impacto significativo nas suas vidas futuras, quer a nível escolar como de capacidade social e emocional. Os distúrbios de leitura são estimados em pelo menos 80% das crianças com dificuldades de aprendizagem, sendo o mais comum a dislexia. A consciência fonológica é reconhecida como um indicador de possíveis problemas de leitura em crianças mais novas.

As aplicações e jogos digitais mostram uma grande promessa na deteção e intervenção destes problemas, fornecendo automação de diferentes tipos de atividades e promovendo a execução dos exercícios tantas vezes quanto for necessário. Mais ainda, jogos educativos são uma boa maneira de manter as crianças empenhadas e motivadas a trabalhar tanto dentro como fora da escola, para além de servirem como boas ferramentas complementares para o trabalho de terapeutas e educadores.

A criação de ferramentas para o auxílio do diagnóstico de distúrbios da fala constitui um campo em que é necessária a investigação e desenvolvimento.

3. **Sujeitos**

   a) **Número de indivíduos previstos inclui:**

      Idealmente 50, mas poderão ser menos devido ao contexto atual da pandemia.

   b) **Critérios de inclusão:**

      - Crianças dos 4-7 anos que, preferencialmente, ainda não tenham aprendido a ler fluentemente, mas que já tenham capacidades de discurso compreensíveis.
      - Terapeutas do Centro de Neurodesenvolvimento e Comportamento da Criança e do Adolescente especializadas em distúrbios da fala.

4. **Descrição resumida do plano de investigação**

A recolha dos dados será feita entre Novembro e Dezembro de 2020.

Os testes serão realizados, de preferência, num tablet fornecido pelo responsável do teste. Os testes serão realizados individualmente, com a presença dos pais ou responsáveis, caso se
Aplique, e poderão ser gravados (apenas voz) para facilitar o tratamento de dados futuros. Depois de transcrita, esta gravação será descartada. Será recolhido junto do responsável pela criança alguns dados demográficos da criança, juntamente com o preenchimento do consentimento informado. O teste a realizar terá a duração de 15 minutos, em que a criança poderá usar livremente a aplicação como bem entender. O responsável pelo teste irá registar numa folha algumas observações referentes à performance da criança enquanto utiliza a aplicação, que irão depois ser usadas para avaliar o sistema. A aplicação requer a realização de alguns exercícios, que têm de ser respondidos através da voz. A resposta é então gravada pelo microfone do tablet e avaliada por um sistema de reconhecimento de voz automático.

No caso dos testes com terapeutas, serão pedidas a realização de algumas tarefas simples através da aplicação. A criança pode a qualquer momento desistir de usar a aplicação, se não se mostrar interessada, terminando o teste. Caso contrário, ao fim dos 15 minutos o teste será dado como terminado e irá fazer-se um breve questionário à criança que irá avaliar a sua satisfação face à aplicação utilizando uma escala de 1 a 5 (escala de Likert). No caso dos testes com terapeutas, a sessão será dada como terminada depois do participante completar as tarefas, colocando-se também um questionário no final.

A aplicação pede à criança o seu nome e idade, que serão usados apenas para efeitos de personalização do jogo e para identificação da criança aquando do tratamento dos dados. Estes dados serão depois tornados anónimos. Para efeitos de avaliação do sistema, serão recolhidos os áudios das respostas. Será também registado se a criança possuir alguma condição que provoque perturbações na fala.

5. Risco/Benefício para o participante

O participante envolvido não terá qualquer risco consequente da participação neste estudo, nem qualquer benefício direto.

6. Benefício para o investigador/instituição

Os investigadores envolvidos poderão tirar conclusões sobre a utilidade e usabilidade da aplicação que foi desenvolvida, o que irá beneficiar trabalhos futuros sobre o tema.
7. Termo de responsabilidade

Eu abaixo assinado, na qualidade de investigador principal, declaro por minha honra que as informações prestadas neste questionário são verdadeiras. Mais declaro que, durante o estudo, serão respeitadas as recomendações constantes das Declarações de Helsinquia a de Tóquio, da Organização Mundial de Saúde e da Comunidade Europeia, no que se refere à experimentação que envolva seres humanos, bem como o constante DL 43/04 de 19 de Agosto, DR I Série.

Lisboa, 11 de Novembro de 2020

[Signature]

Alberto Abad
Assistant Professor at Instituto Superior Técnico, University of Lisbon
Coordinator at Human Language Technologies Lab INESC-ID
Form and Approval of the Ethical Committee at IST
COMISSÃO DE ÉTICA DO IST (CE-IST)
Pedido de Parecer ¹

Preencha o formulário, digitalize-o e envie para comissaoetica@tecnico.ulisboa.pt.

O parecer deve ser emitido em ✗ Português  □ Inglês

1. IDENTIFICAÇÃO GERAL

Título

| Tecnologias da fala aplicadas à terapia das perturbações do desenvolvimento da linguagem |

Investigador responsável - Professor ou investigador doutorado do IST

| Alberto Abad, INESC-ID/IST, alberto.abad@tecnico.ulisboa.pt |

Equipa (se aplicável)

| Alberto Abad, INESC-ID/IST, alberto.abad@tecnico.ulisboa.pt |
| Isabel Trancoso, INESC-ID/IST, isabel.trancoso@tecnico.ulisboa.pt |
| Matilde Ramos, INESC-ID/IST, matilde.ramos@tecnico.ulisboa.pt |

Unidade de investigação do IST onde decorrerá o estudo

| INESC-ID |

Diferenciação do pedido:

- ☐ Projeto de investigação
- ✗ Dissertação de mestrado
- ☐ Dissertação de Doutoramento
- ☐ Outro (especificar) ________________________________

¹ A Comissão de Ética apenas se pronuncia sobre estudos cuja parte experimental ainda não tenha sido realizada. Excepcionalmente, em casos justificados, poderá pronunciar-se sobre estudos por finalizar, cuja parte experimental tenha sido concluída durante os 3 meses anteriores à data de submissão do Pedido de Parecer.
Duração:    Início 16/11/2020    Fim 20/12/2020
Fonte de financiamento (se aplicável):

a) A ser submetido a financiamento externo – indicar a entidade à qual será submetido

b) Com financiamento externo – indicar a entidade e a referência do projeto

c) Com financiamento do IST – indicar o centro de investigação que o financia

2. OBJETIVOS DO PROJETO (<3000 caracteres)

Esta investigação pretende avaliar a usabilidade de uma aplicação desenvolvida para crianças, cujo foco é a resolução de exercícios de consciência fonológica através da fala. Para tal, será feita a gravação áudio das respostas aos exercícios através de um microfone, as quais serão avaliadas através de um sistema de reconhecimento de fala automático. As crianças irão ser observadas a utilizar a aplicação durante algum tempo e irão, de seguida, responder a um breve questionário.

3. METODOLOGIA E PLANIFICAÇÃO (<5000 caracteres)

Neste estudo serão aceites crianças entre os 4-7 anos que, preferencialmente, ainda não tenham aprendido a ler fluentemente, mas que já tenham capacidades de discurso compreensíveis. Os testes serão realizados, de preferência, num tablet fornecido pelo responsável do teste. Os testes serão realizados individualmente, com a presença dos pais ou responsáveis, e poderão ser gravados (apenas voz) para facilitar o tratamento de dados futuros. Depois de transcrita, esta gravação será descartada. Aos pais, será explicado o teste e garantido que a anonimidade será preservada. Será então recolhido junto do responsável pela criança alguns dados demográficos da criança, juntamente com o preenchimento do consentimento informado. À criança, será explicado o teste e feito uma
demonstração da aplicação. De seguida, a criança poderá usar livremente a aplicação como bem entender, durante 15 minutos. A criança pode a qualquer momento desistir de usar a aplicação, se não se mostrar interessada, terminando o teste. Durante a sessão, o responsável pelo teste, que estará presente, irá registar algumas observações numa folha. Estas observações dizem respeito à performance da criança enquanto utiliza a aplicação, que irão depois ser usadas para avaliar o sistema. A aplicação requer a realização de alguns exercícios, que têm de ser respondidos através da voz. A resposta é então gravada pelo microfone do tablet e avaliada por um sistema de reconhecimento de voz automático. Ao fim dos 15 minutos o teste será dado como terminado e irá fazer-se um breve questionário à criança, que irá avaliar a sua satisfação face à aplicação utilizando uma escala de 1 a 5 (escala de Likert). Finalmente, após o preenchimento do mesmo, dar-se-á por encerrada a sessão.

Os testes serão realizados num sítio à escolha do participante, desde que seja num ambiente calmo, e serão aplicados por apenas uma pessoa, que estará presente na sala e irá recolher e processar todos os dados. As informações serão armazenadas num repositório privado, acessível apenas aos investigadores do estudo.

4. INTERAÇÃO DO PROJETO, À LUZ DA LEGISLAÇÃO APLICÁVEL, COM ALGUM DOS SEGUINTE

DOMÍNIOS:

- Ao envolvimento de seres humanos em projetos de investigação, de forma direta como objetos de pesquisa ou de forma indireta, mas também suscetível de os afetar;
- Ao respeito pelo consentimento informado (se aplicável, anexar formulário correspondente);
- À proteção da privacidade e dos dados pessoais²;
- Ao respeito pela integridade académica;
- À relação inerente entre os diferentes membros da comunidade académica;
- À proteção dos direitos de propriedade intelectual;
- À proteção de pessoas especialmente vulneráveis;
- À utilização de animais em projetos de investigação.

Em caso afirmativo, apresentar indicação sumária das medidas adotadas para lidar com o(s) tema(s) sensível(eis) em causa (<3000 caracteres).

---

² Se aplicável, preencher caixas relativas ao Encarregado de Proteção de Dados e ao Responsável pelo Tratamento de Dados, a indicar na declaração de consentimento informado dos participantes no estudo.
No início de cada sessão será entregue o documento de consentimento informado, no qual o responsável da criança aceita a recolha de dados demográficos, dados de voz, e dados de saúde (se o participante tem alguma perturbação da fala). A participação no estudo é voluntária e os participantes podem desistir da participação a qualquer altura. Será garantida a anonimidade do participante, sendo que qualquer informação recolhida será anônimizada e tratada apenas para fins de investigação. Os dados recolhidos serão guardados num repositório privado protegido com senha de acesso, acessível apenas aos investigadores do estudo.

**Encarregado de Proteção de Dados (Data Protection Officer, DPO) que acompanhará o estudo (se aplicável)**

Alberto Silva, INESC-ID, alberto.silva@inesc-id.pt

**Responsável pelo tratamento de dados que acompanhará o estudo (se aplicável)**

Matilde Ramos, INESC-ID, matilde.ramos@tecnico.ulisboa.pt

**5. QUESTÕES ÉTICAS SOBRE AS QUAIS SE PRETENDE O PARECER DA CE-IST (<3000 caracteres)**

- Recolha de dados demográficos.
- Recolha de dados de voz.
- Recolha de dados de saúde (se o participante tem alguma perturbação da fala).
- Recolha de inquéritos de satisfação.
- Armazenamento da informação recolhida num repositório privado, acessível apenas aos investigadores do estudo.
- Utilização dos dados recolhidos para estudos futuros relacionados.

**6. ESTUDOS SEMELHANTES QUANTO À INVESTIGAÇÃO NO ÂMBITO DAS IMPLICAÇÕES ÉTICAS (indicar referências bibliográficas, se aplicável) (<3000 caracteres)**


COMPROMISSO ÉTICO

No que é aplicável, considero-me obrigado a conhecer e a respeitar os direitos humanos (sobretudo, em matéria de conhecimento e consentimento livre, específico, informado, esclarecido e explícito relativos à investigação pelos visados) e dos animais, bem como os princípios éticos e deontológicos nacionais (nomeadamente, o Código de Conduta e Boas Práticas da Universidade de Lisboa e o Regulamento da Comissão de Ética do IST) e internacionais aplicáveis.

Caso o projeto envolva questões respeitantes à proteção da privacidade e dos dados pessoais, considero-me obrigado a dar conhecimento ao Encarregado de Proteção de Dados (Data Protection Officer, DPO) da Instituição em que se desenvolve o estudo\(^3\) e a respeitar as orientações deste recebidas.

Dou o meu consentimento para a publicação do Parecer no site da Comissão de Ética do IST (http://etica.tecnico.ulisboa.pt/).

☐ Sim ☐ Não

Lisboa, 11 de Novembro de 2020

[Signature]

Alberto Abad
Assistant Professor at Instituto Superior Técnico, University of Lisbon
Coordinator at Human Language Technologies Lab INESC-ID

---

\(^3\) No caso do estudo se desenvolver no IST ou numa unidade de investigação sem DPO próprio, o DPO a informar será o da Universidade de Lisboa (rgpd@ulisboa.pt).
STATEMENT

Ref. n.º 23/2020 (CE-IST)
Date: 02/12/2020

Nome IR: Alberto Abad, INESC-ID

Name of the projet: "Tecnologias da fala aplicadas à terapia das perturbações do desenvolvimento da linguagem".

Prof. Alberto Abad

A Comissão de Ética do Instituto Superior Técnico (CE-IST) apreciou os documentos submetidos com vista à obtenção de um parecer sobre os aspetos éticos do projeto mencionado.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Documentos</th>
<th>Versão e data</th>
</tr>
</thead>
</table>
| #938350 | Consentimento Informado.pdf  
Formulario_COMISSAO DE ETICA IST_11-11-2020.pdf  
Questionário Demográfico.pdf  
Questionário Final.pdf  
Tabelas de Observação.pdf | 20-11-2020 |

The following members of the EC-IST participated in the ethical assessment:

<table>
<thead>
<tr>
<th>Name</th>
<th>Role in Ethics Committee</th>
<th>Qualification</th>
<th>Gender</th>
<th>Affiliation to IST (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>António Pinheiro</td>
<td>Presidente</td>
<td>Professor</td>
<td>M</td>
<td>Y</td>
</tr>
<tr>
<td>Mário Gaspar da Silva</td>
<td>Member</td>
<td>Professor</td>
<td>M</td>
<td>Y</td>
</tr>
<tr>
<td>Isabel Sá Correia</td>
<td>Member</td>
<td>Professor</td>
<td>F</td>
<td>Y</td>
</tr>
<tr>
<td>Rui Medeiros</td>
<td>Member</td>
<td>Professor</td>
<td>M</td>
<td>N</td>
</tr>
</tbody>
</table>

Foram analisados os seguintes documentos:

A CE-IST desenvolve as suas apreciações em conformidade com o ICH-GCP, a Schedule Y, as linhas de orientação ICMR, o Regulamento da IC-IST e com outra regulamentação aplicável.

Nenhum dos investigadores que participam no projeto de investigação em apreciação participou no correspondente processo de decisão e de votação.

Com base na apreciação dos documentos atrás mencionados, a CE-IST decidiu (por unanimidade) emitir um parecer favorável sobre o modo como as questões éticas são tratadas no projeto em apreço.

A CE-IST entende dever ser informada acerca do desenvolvimento do projeto, de quaisquer acontecimentos adversos que venham a ocorrer e de eventuais revisões dos protocolos seguidos, da informação aos participantes ou do formato do consentimento informado, e entende dever receber uma cópia do relatório final do projeto de investigação.

Prof. António Pinheiro
Presidente da Comissão de Ética  
do Instituto Superior Técnico (CE-IST)
Informed Consent
Eu, ________________________________________________, na qualidade de participante ou encarregado do participante, declaro ter lido este documento e aceito a participação num estudo que tem por objetivo testar uma aplicação completando alguns exercícios de consciência fonológica. O estudo proposto foi-me explicado de forma clara e foi-me dada a oportunidade de colocar questões. Declaro que aceito a participação, voluntária, neste estudo.

As gravações áudio realizadas durante este estudo poderão ser usadas para investigação e desenvolvimento de sistemas automáticos de processamento de fala. Fui informado de que os investigadores ouvirão a gravação do participante e poderão disponibilizá-la para os mesmos fins a outros investigadores no contexto do mesmo estudo. Garantindo o anonimato, dou autorização para que sejam usadas no contexto de outros estudos. Dou o meu consentimento para usarem a minha fala ou a fala do participante pelo qual sou responsável, gravada para estes fins, desde que seja garantido o anonimato. Para além disso, dou o meu consentimento para a recolha de dados demográficos do participante e que fique registado se o mesmo tem alguma perturbação da fala, na qualidade de que se mantenha a sua anonimidade.

Foi-me explicado que:
• Sou livre para sair do estudo a qualquer momento sem necessidade de justificar a minha decisão.
• Os dados a respeito do participante serão estritamente confidenciais. Posso, a qualquer momento, exercer o meu direito de acesso, retificação e oposição.
• A publicação dos resultados da investigação respeitará o direito ao anonimato do participante.

Recebi uma cópia desta declaração de consentimento informado devidamente assinada e datada.

__________________________       __________________________     _____________ 
Nome do Participante      Assinatura do Participante         Data

Discuti este estudo de investigação com o participante, utilizando uma linguagem compreensível e apropriada. Informei adequadamente o participante sobre a natureza deste estudo e sobre os seus possíveis benefícios e riscos, considerando que o participante compreendeu a minha explicação.

__________________________       __________________________     _____________ 
Nome do Investigador      Assinatura do Investigador         Data