



Tecnologias Assistidas para Terapia da Fala

RICARDO MIGUEL BARROS RAMIRO

Outubro de 2021

Tecnologias Assistidas para Terapia da Fala

Ricardo Miguel Barros Ramiro

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Sistemas Gráficos e Multimédia**

Orientador: Professor Doutor António Vieira de Castro

Coorientadora: Professora Doutora Tânia Rocha

Supervisora: Professora Doutora Maria João Cunha

Porto, outubro de 2021

Aos meus pais pelo que me proporcionaram ao longo da vida.

Resumo

Na generalidade o ser humano necessita de se relacionar com outros. Sendo que do ponto de vista social as relações com outras pessoas são fundamentais e elas são possíveis devido à comunicação. Comunicar está presente no nosso dia a dia e permite dizer às pessoas o que pensamos e ouvir também delas as suas ideias. Para se poder comunicar a maioria das pessoas utiliza a linguagem, sendo com ela que conseguimos estabelecer relações e nos podemos entender. A comunicação faz parte das necessidades básicas de bem-estar do ser humano.

A Terapia da Fala surge por sua vez para ajudar os indivíduos a melhorarem a sua comunicação através da linguagem. Este trabalho pretendeu com a ajuda de profissionais da área da Terapia da Fala estudar a possibilidade de utilizar tecnologia existente em prol da Terapia da Fala de maneira que esta pudesse auxiliar os pacientes a melhorarem a sua situação comunicativa e de uma forma mais autónoma. Permitindo principalmente aos pacientes praticarem as indicações dadas pelo seu Terapeuta com maior frequência. Dada a evolução da tecnologia de reconhecimento de voz procurou-se utilizá-la para criar uma solução que apoiasse a área da Terapia da Fala.

Alguns aspetos importantes deste trabalho basearam-se no levantamento do estado da arte das tecnologias de reconhecimento de voz e dos materiais utilizados pelos Terapeutas da Fala, para tal foram realizadas reuniões para melhor entender como se poderia ajudar a área e realizar o levantamento de requisitos. Seguiu-se a análise e desenvolvimento do protótipo e respetiva fase de avaliação. No final apresenta-se a conclusão e crítica de resultados ao trabalho realizado. Através deste trabalho foi possível concluir que a solução apresentada pode vir a ser útil para as pessoas com dificuldades na fala, contudo ela não substitui o papel do Terapeuta da Fala. Podendo isso sim funcionar como complemento às sessões terapêuticas. É importante referir que a solução foi validada por pessoas da área, mas, todavia, ainda é um protótipo que pode vir a ser melhorado.

Palavras-chave: Comunicação, Reconhecimento de Fala, Terapia da fala, Acessibilidade

Abstract

Overall, the human being needs to relate with others, from a social point of view, relationships with other people are fundamental and they are possible due to communication. Communicating is present in our daily lives and allows us to tell people what we think and listen to their ideas. To communicate most people, use language, it is with it that we can establish relationships and that we can understand them. Communication is part of the basic well-being needs of human beings.

Speech Therapy, in turn, appears to help our people improve their communication through language. This work intended, with the help of professionals in the field of Speech Therapy, to study the possibility of using existing technology in favor of Speech Therapy so that it can help patients improve their communicative situation in a more autonomous way. Mainly allowing patients to practice the directions given by their therapist more often. Given the evolution of voice recognition technology, the main purpose was it to be used to create a solution that would support Speech Therapy area.

Some important aspects of this work were based on the survey of the state of the art of voice recognition technologies and the materials used by Speech Therapists, for that meeting were established to better understand how the area could be helped and gather requirements as well. Then it was followed by the analysis and development of the prototype and respective evaluation phase. At the end a final and critical conclusion of the results is presented.

Based on this work, it was possible to conclude that a solution presented may be useful for people of a variety of ages with speech difficulties, however it does not replace the role of the Speech Therapist. This can indeed work as a complement to therapies. It is important to mention that the solution was validated by people in the area, but, nevertheless, it is still a prototype that could be improved.

Keywords: Communication, Voice Recognition, Speech Therapy, Accessibility

Agradecimentos

Agradeço a todos os que direta ou indiretamente me apoiaram na realização desta dissertação, especialmente ao meu orientador, o Prof. Doutor António Vieira Castro do Instituto Superior de Engenharia do Porto (ISEP) pela sua dedicação, apoio e pelos conhecimentos transmitidos.

Agradeço à minha coorientadora Prof. Doutora Tânia Rocha da Universidade de Trás-os-Montes e Alto Douro (UTAD) por todas as sugestões adicionais fornecidas ao longo deste trabalho.

Agradeço ainda aos meus supervisores, da Escola Superior de Saúde do Politécnico do Porto, à Prof. Doutora Maria João Cunha e ao Prof. Doutor André Araújo.

Aos meus colegas de curso especialmente ao Ruben Magina pela troca de ideias.

Aos meus pais e avós por todo o carinho e apoio dado ao longo da minha vida e motivação dada.

Ao ISEP, à UTAD e à Escola Superior de Saúde (ESS) por terem proporcionado uma orientação de qualidade sem o qual este trabalho poderia ficar comprometido.

Ao SiiS por me permitir fazer parte do grupo de investigação e aprender mais sobre as minhas áreas científicas de interesse.

Índice

Capítulo 1 - Introdução	1
1.1 Contextualização	1
1.2 O problema	2
1.3 Objetivos e contributos esperados.....	3
1.4 Motivação.....	4
1.5 Estrutura do documento.....	5
Capítulo 2 - Estado da arte.....	7
2.1 A Terapia da Fala.....	7
2.2 Tecnologias de Apoio à Terapia da Fala	8
2.2.1 Materiais de apoio na Terapia da Fala	9
2.2.2 Aplicações tecnológicas de apoio à Terapia da Fala	9
2.3 O processo de reconhecimento de fala	12
2.3.1 Terminologia básica.....	13
2.3.2 Tipos de reconhecimento de voz.....	14
2.3.3 Usos e Aplicações.....	14
2.3.4 Funcionamento dos Sistemas Reconhecimento de Voz	20
2.4 Avaliação dos Sistemas de Reconhecimento de Fala.....	22
2.5 APIs de reconhecimento de fala	23
2.5.1 CMU Sphinx	24
2.5.2 Google Cloud Speech-to-text	24
2.5.3 Microsoft Speech to Text.....	25
2.5.4 IBM Watson Speech to text.....	26
2.5.5 Mozilla Webspeech HTML API.....	26
Capítulo 3 - Análise de valor.....	29
3.1 Fuzzy Front End.....	29
3.2 NCD - Novo Conceito de Desenvolvimento.....	29
3.2.1 Identificação da Oportunidade	30
3.2.2 Análise de Oportunidade	31
3.3 Proposta de Valor	32

3.4	Quality Function Deployment (QFD).....	33
3.5	AHP - Método de Análise Hierárquica	34
Capítulo 4 - Análise e Desenho		39
4.1	Análise de requisitos	39
4.1.1	Requisitos Funcionais	39
4.1.2	Diagramas de Caso de Uso	40
4.1.3	Requisitos de Instalação.....	42
4.2	A arquitetura.....	43
4.2.1	Linguagens Utilizadas.....	45
4.3	Gamificação	45
Capítulo 5 - Implementação		47
5.1	Base de Dados.....	47
5.2	Código CSS e HTML	48
5.3	Código de reconhecimento de fala	49
5.4	Código PHP	50
5.5	Front End	53
5.5.1	Perfil Paciente	53
5.5.2	Perfil Terapeuta	56
5.6	Planeamento dos testes	59
Capítulo 6 - Avaliação		61
Capítulo 7 - Conclusões e trabalho futuro.....		67
7.1	Conclusões e Critica de Resultados.....	67
7.2	Trabalho Futuro.....	68

Lista de Figuras

Figura 1 – Esboço da proposta inicial	4
Figura 2 – “Mind Map” da Estrutura do Documento	5
Figura 3 – Exemplo de ecrã de Jogo do EasyTefa	10
Figura 4 – Exemplo de ecrã da aplicação Happies	11
Figura 5 – Exemplo de ecrã da aplicação Terapia da Fala	12
Figura 6 – Exemplo de ecrã do programa Word da Microsoft	15
Figura 7 – Amazon Echo de quarta geração em preto (2020)	17
Figura 8 – Google Nest Hub de segunda geração (2016)	18
Figura 9 – Ecrã de utilização da Bixby.....	19
Figura 10 – Modelo NCD (P. A. Koen et al., 2015)	30
Figura 11 – Proposta de valor de Canvas.....	32
Figura 12 – Exemplo de House of Quality (HoQ) (Peng et al., 2018)	34
Figura 13 – Diagrama hierárquico AHP.....	35
Figura 14 – Diagrama caso de Uso Paciente.....	41
Figura 15 – Diagrama caso de Uso Terapeuta da Fala.....	42
Figura 16 – Arquitetura da aplicação web a desenvolver	43
Figura 17 – Estrutura do sistema de reconhecimento de voz	44
Figura 18 – Maquete inicial do protótipo.....	44
Figura 19 – Ecrã phpMyAdmin para base de dados demo	48
Figura 20 – Tabelas e campos da base de dados.....	48
Figura 21 – Parte do código CSS do protótipo.....	49
Figura 22 – Função Javascript utilizada para finalizar reconhecimento de fala	50
Figura 23 – Função javascript onresult.....	50
Figura 24 – Código PHP de conexão à Base de dados e consulta em variável	51
Figura 25 – Variável Javascript para guardar array em PHP	51
Figura 26 – Código PHP para gerar tabela HTML.....	52
Figura 27 – Ecrã do menu principal utilizador cliente	53
Figura 28 – Ecrã da opção “Analisar” do utilizador paciente.....	54
Figura 29 – Ecrã da opção “Treinar” do utilizador paciente	54
Figura 30 – Ecrã da opção “Resultados” do utilizador paciente.....	55
Figura 31 – Ecrã da opção “Notas” do utilizador paciente	55

Figura 32 – Ecrã do menu principal utilizador terapeuta	56
Figura 33 – Ecrã da opção “Frases” do utilizador Terapeuta	57
Figura 34 – Ecrã da opção “Palavras” do utilizador Terapeuta.....	57
Figura 35 – Ecrã da opção “Notas” do utilizador Terapeuta.....	58
Figura 36 – Ecrã da opção “Resultados” do utilizador Terapeuta	58
Figura 37 – Ecrã da opção “Banco” de palavras do utilizador Terapeuta	59
Figura 38 – Ecrã de reunião para validar protótipo realizada via Skype.....	59
Figura 39 – Um dos momentos de avaliação do TeraFA.....	62
Figura 39 – Gráfico da amostra de sexo dos inquiridos.....	62
Figura 40 – Gráfico da amostra de faixa etária dos inquiridos	63
Figura 41 – Gráfico do nível de escolaridade dos inquiridos	63
Figura 42 – Gráfico sobre avaliação do sistema de reconhecimento de fala	64
Figura 43 – Gráfico sobre facilidade de navegar pela aplicação e aprendizagem das funcionalidades.....	64
Figura 44 – Gráfico de opinião se a aplicação pode ajudar as pessoas com problemas de fala	65

Lista de Tabelas

Tabela 1: Áreas de atuação do Terapeuta da Fala (Batista, 2011)	8
Tabela 2: Comparação de APIS de reconhecimento de voz	27
Tabela 3: Escala Fundamental de Saaty (Saaty, 1990).....	35
Tabela 4: Tabela AHP de comparação de critérios	36
Tabela 5: Tabela AHP de comparação das alternativas para a Curva de Aprendizagem.....	36
Tabela 6: Tabela AHP de comparação das alternativas para a Documentação.....	37
Tabela 7: Tabela AHP de comparação das alternativas para a Funcionalidades	37
Tabela 8: Tabela AHP de comparação das alternativas para os Idiomas Suportados	38
Tabela 9: Tabela AHP para os resultados obtidos	38

Acrónimos e Símbolos

Lista de Acrónimos

AJAX	<i>Asynchronous Javascript and XML</i>
API	<i>Application Programming Interface</i>
APTF	Associação Portuguesa de Terapeutas da Fala
ASR	<i>Automatic Speech Recognition</i>
AVC	Acidente Vascular Cerebral
CMU	<i>Carnegie Mellon University</i>
CSS	<i>Cascading Style Sheets</i>
ESS	Escola Superior de Saúde
gPRC	open source Remote Procedure Call
HCI	<i>Human Computer Interaction</i>
HTML	<i>HyperText Markup Language</i>
IOT	<i>Internet of Things</i>
IPA	<i>Intelligent Personal Assistant</i>
ISEP	Instituto Superior de Engenharia do Porto
JSON	<i>Javascript Object Notation</i>
NIDCD	<i>National Institute on Deafness and Other Communication Disorders</i>
REST	<i>Representational State Transfer</i>
SRV	Sistema de Reconhecimento de Voz
TdF	Terapia da Fala
TF	Terapeuta da Fala
URI	<i>Uniform Resource Identifier</i>
UTAD	Universidade de Trás-os-Montes e Alto Douro
UX	<i>User Experience</i>
VA	<i>Value Analysis</i>
WER	<i>Word Error Rate</i>

Capítulo 1 - Introdução

A escrita desta dissertação faz parte da realização do Mestrado em Engenharia Informática ramo de Sistemas Gráficos e Multimédia pelo Instituto Superior de Engenharia do Porto (ISEP). Com este trabalho pretende-se contribuir com o desenvolvimento de uma solução tecnológica que permita auxiliar os Terapeutas da Fala (TF) na ajuda de pessoas com dificuldade na fala quanto á correta dicção das palavras. Isto é, após um estudo dos procedimentos dos terapeutas da fala aquando do apoio aos seus pacientes.

Neste capítulo de introdução encontra-se a contextualizado o problema em estudo, os objetivos e contributos esperados neste trabalho através do desenvolvimento de um protótipo, bem como a motivação e estrutura do documento.

1.1 Contextualização

De forma a integrar a nossa sociedade, cada individuo necessita obrigatoriamente de comunicar com outros indivíduos, principalmente através da fala. Sendo que a fala surge como facilitador no sentido de ajudar o ser humano a comunicar com as outras pessoas, assumindo assim um papel importante como um dos principais meios de comunicação. “A fala é um dos aspetos mais importantes da comunicação humana.” (Darejeh et al., 2019). Todos os dias a maioria das pessoas fala com outras e utiliza a tecnologia por exemplo para as ajudar a comunicar à distância. No passado nem sempre foi assim, mas com a ajuda da tecnologia as pessoas começaram a poder comunicar com outras pessoas à distância, recorrendo por exemplo ao computador.

Hoje em dia podemos utilizar o telefone ou a tecnologia de vídeo chamada. Estas invenções foram sem dúvida alguma uma mais-valia para as pessoas poderem comunicar mais facilmente umas com as outras. Além de que fazem com que possivelmente as pessoas possam melhorar a fala, pois passam a praticar mais.

É muito provável que a tecnologia continue a evoluir, e certamente irão surgir novas invenções, que permitam ao ser humano fazer uso dela e progredir.

Assim como a tecnologia permitiu comunicar à distância, porque não pode a tecnologia ser utilizada para ajudar-nos mais a melhorar a fala?

É importante falarmos bem para que os outros nos possam entender melhor. Se conseguirmos falar melhor é possível que aumentemos também a confiança em nós mesmos e também que se evitem erros de interpretação no que dizemos e o que os outros entendem de nós. Por exemplo na engenharia existem muitas vezes problemas de comunicação onde não existe uma compreensão correta do que se nos pede e no que se faz.

Investir mais na melhoria da fala pode ser uma mais-valia para a sociedade e para cada individuo. Um aspeto importante deste trabalho está em avaliar o impacto que a tecnologia pode ter na fala. Pode a tecnologia ajudar mais o ser humano a resolver problemas existentes na Terapia da Fala?

1.2 O problema

Aproximadamente quinze por cento da população global, cerca de mil milhões de pessoas, vive com alguma forma de incapacidade, dos quais dois a quatro por cento possui sérias dificuldades de comunicação. (National Institute on Deafness and Other Communication Disorders, 2020) Estes dados mostram que os problemas de comunicação/fala são uma realidade e estão bem presentes na nossa sociedade.

É importante mencionar que os problemas de comunicação afetam as emoções e a vida social das pessoas e pode comprometer possibilidades futuras.

Os problemas de fala consistem numa “[...]diminuição e na dificuldade ou incapacidade de reproduzir certos sons na fala, sendo que a perturbação pode ser mais severa ou mais leve. O problema pode também estar na incapacidade de articular, definido na omissão de palavras ou na distorção de certos sons na fala; podendo também estar presente na incapacidade de eloquência, bem como na utilização de fluxos não típicos, ritmos, e/ou repetição de sons, ou perturbação da voz, caracterizado pela anormal fixação da voz, volume, ressonância, qualidade vocal ou duração.”¹

Verificando-se que existem falta de soluções de apoio à terapia da fala no apoio a Terapeutas e seus pacientes pretendemos analisar o potencial das novas APIs de áudio e da tecnologia de reconhecimento de fala para verificar a viabilidade da utilização desta na criação de um protótipo tecnológico.

¹ Disponível em <https://www.washington.edu/doit/what-speech-impairment>

Que potencial tem a tecnologia para ajudar as pessoas a comunicarem melhor no contexto da Terapia da Fala? Na terapia da fala é possível que exista potencial na tecnologia para ser mais utilizada, por exemplo no desenvolvimento de soluções interativas utilizando a tecnologia de reconhecimento de fala para ajudar as pessoas com problemas de fala a melhorarem a sua situação. Como? Este trabalho pretende dar resposta a esta pergunta. Será que a tecnologia de reconhecimento de voz pode ajudar os Terapeutas da Fala e respetivos pacientes.

É importante referir que das pesquisas efetuadas até à data não foram encontradas outras abordagens existentes significativas para ajudar na terapia da fala numa perspetiva focalizada na melhora da fala.

1.3 Objetivos e contributos esperados

O objetivo principal deste trabalho passa por auxiliar os terapeutas da fala em ajudar indivíduos com dificuldades na fala através da criação de uma aplicação que capture o áudio dito pelo utilizador e reconheça as frases/palavras ditas e as classifique quanto à sua correção, ou seja avaliando se estas foram ditas de forma correta ou incorreta. Isto de maneira a melhorar a situação dos indivíduos com problemas de fala. Partindo-se do princípio de que quanto mais o individuo praticar o(s) diferente(s) som/sons em que tem dificuldade, este irá melhorar a sua situação de dicção.

De uma forma geral pretende-se viabilizar uma solução que passaria pelo desenvolvimento de um *software* para ser utilizado num dispositivo eletrónico pessoal, um computador, *tablet* ou *smartphone*, onde exista a possibilidade de ser utilizado através da captação da fala para depois corrigir o que foi dito pelo utilizador. Desta forma o utilizador teria uma ferramenta tecnológica capaz de o ajudar a falar melhor.

Para o efeito pretende-se realizar um estudo sobre o “estado da arte” existente analisando os processos atuais e identificando soluções que já se utilizam.

Pretendemos colaborar com profissionais da área da Terapia da Fala de modo a identificar necessidades e perceber se a proposta desenhada após algum diálogo inicial tem potencial para contribuir para o fim terapêutico a que se destina. Nesta fase foi também tido como objetivo a existência de reuniões preliminares para identificar formas de validação e teste da solução a desenvolver.

Espera-se também contribuir com a avaliação de *APIs* de reconhecimento de fala, para verificar a possibilidade de criação do protótipo.

Como ilustrado na figura 1 o objetivo passa por dar ao paciente um texto adequado à sua patologia, o qual ele deve ler em voz alta para o microfone do seu dispositivo eletrónico, sendo que o sistema deve identificar as palavras/sons que não foram ditos corretamente e apresentá-los ao Terapeuta para que ele e o paciente os possam trabalhar da forma mais adequada no sentido de poder corrigir o problema.

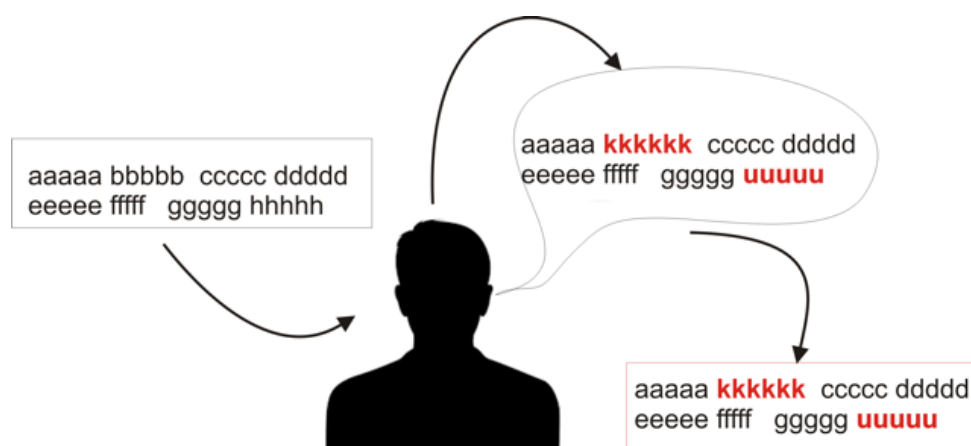


Figura 1 – Esboço da proposta inicial

Segundo (Furlong et al., 2017) a metodologia utilizada pelos Terapeutas da Fala (TF), concretamente a repetição de sons melhora o processo de aprendizagem do indivíduo, quando este denota dificuldades num dito som.

Baseando-nos na análise da evolução de tecnologias de reconhecimento de fala, e das APIs relacionadas com o reconhecimento de fala pretendemos trabalhar diretamente com profissionais de Terapia da Fala e investigadores da Escola Superior de Saúde (ESS) do Porto de modo a analisar, desenhar, desenvolver, implementar e testar o potencial de uma solução informática aplicada à área da Terapia da Fala sob a forma de um protótipo capaz de ajudar os indivíduos a melhorarem a fala através de uma interface adequada para o efeito e que focalize procedimentos adotados pelos terapeutas.

Dentro das possibilidades pretendemos escrever um ou mais artigos relacionados com o tema de modo a validar o presente estudo.

1.4 Motivação

O autor do presente estudo é licenciado em Engenharia Informática e desde sempre se sentiu altamente fascinado e motivado pela inovação tecnológica, pela sua capacidade e sobretudo pela sua aplicação em casos com potencial para contribuir para uma melhoria da sociedade.

Paralelamente, sempre apreciou de forma muito particular os mecanismos de interação humano computador (HCI) e o seu potencial para facilitar uma boa e eficaz experiência de utilizador (UX).

O desafio do desconhecido e da aplicação de um conceito a um caso concreto como o da terapia da fala é também altamente aliciante e motivador.

O principal motivo que me levou a considerar desenvolver este trabalho está relacionado com a constatação da falta de soluções tecnológicas utilizadas na terapia da fala e pelo interesse na área da saúde e no intuito de tentar ajudar as pessoas com problemas na fala para que possam estar em melhor situação e tenham mais uma opção à qual recorrer.

Pretendemos igualmente avaliar se a solução desenhada e desenvolvida tem potencial para contribuir para uma melhoria no processo de tratamento da terapia da fala.

1.5 Estrutura do documento

Este documento está dividido em sete capítulos, são eles a introdução, estado da arte, análise de valor, análise de requisitos e desenho, implementação da solução, avaliação, conclusões e trabalho futuro, como se apresenta no “mind map” da figura 2.

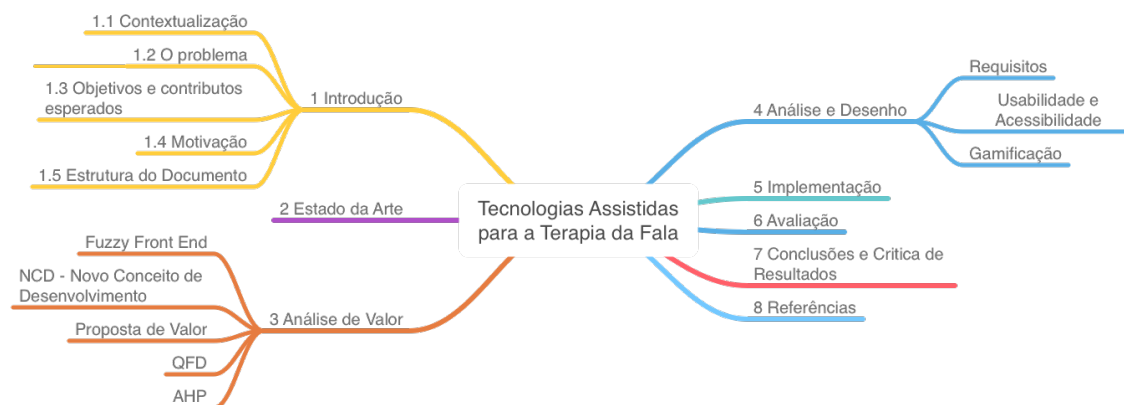


Figura 2 – “Mind Map” da Estrutura do Documento

O primeiro capítulo apresenta uma contextualização do problema, ao nível da comunicação e do impacto da utilização da tecnologia na Terapia da Fala. Em seguida são apresentados os objetivos deste trabalho e o contributo que se espera dar na matéria, bem como a motivação para escolha do tema e respetiva estrutura do documento.

No segundo capítulo apresenta-se o estado da arte, onde é feita uma revisão literária sobre a Terapia da Fala e em que consiste, os materiais e as tecnologias de apoio que a área utiliza. São descritas e avaliadas as tecnologias de reconhecimento de fala, aprofundando através da revisão de APIs para reconhecimento da fala. As APIs são avaliadas e apresentadas as vantagens e desvantagens das mesmas.

O terceiro capítulo consiste na análise de valor de maneira a verificar a viabilidade do projeto e ajudar na escolha das tecnologias a utilizar no desenvolvimento da ferramenta de apoio à Terapia da Fala.

O quarto capítulo tem como objetivo fazer a análise e desenho do protótipo, elementos necessários para o desenvolvimento do protótipo, bem como definir os diagramas de caso de uso para mostrar o comportamento do sistema e ajudar na definição dos requisitos, para desta forma identificar as interações entre o sistema e os seus atores. Outro dos pontos passa apresentar a arquitetura do sistema.

O quinto capítulo descreve o processo de implementação do protótipo. Aqui descreve-se parte do código utilizado, por exemplo na utilização da API de reconhecimento de voz. Também são apresentadas as diferentes interfaces da aplicação para os perfis de administrador e cliente, bem como dos testes.

No sexto capítulo é feita a avaliação da solução ao nível de usabilidade e funcionalidade do protótipo. A avaliação consistiu na realização de um conjunto de testes de usabilidade que foram dados aos utilizadores com o devido acesso ao software e posteriormente estes responderam a um questionário. No final são apresentadas algumas considerações das avaliações dos utilizadores.

O último capítulo, e sétimo, consistiu em analisar os resultados obtidos e na realização de uma crítica de resultados donde são retiradas conclusões ao trabalho desenvolvido.

Capítulo 2 – Estado da arte

O segundo capítulo pretende fazer o levantamento do estado atual de conhecimento sobre a Terapia da Fala (TdF) e das tecnologias de apoio existentes utilizadas. E realizar o levantamento do estado da arte sobre a tecnologia de Reconhecimento de Fala.

2.1 A Terapia da Fala

Antes demais é importante referir que as pessoas desenvolveram capacidades de comunicação verbal e não-verbal. Verbal, por exemplo ao nível oral através da fala, ou por escrito. E de uma forma não verbal, por exemplo através da linguagem gestual. É importante referir que há também quem considere a fala uma tecnologia.

Nesta secção pretendemos introduzir a TdF, fazendo o respetivo enquadramento no trabalho, descrevendo as áreas e grupos de atuação, assim como contextualizar os meios que utiliza. O estudo baseia-se no levantamento do estado da arte usando como referência as entidades representantes da profissão em Portugal, através de pesquisa e leitura científica bem como de reuniões com especialistas nas áreas de maneira a poder estar bem suportado.

O Terapeuta da Fala (TF) “[...] avalia e intervém em indivíduos de todas as idades, desde recém-nascidos a idosos, tendo por objetivo geral otimizar as capacidades de comunicação, linguagem, fala e/ou deglutição do indivíduo, melhorando, assim, a sua qualidade de vida. [refere igualmente que] as principais funções do TF são a de prevenção, avaliação, intervenção, investigação científica e educação contínua.” (Batista, 2011).

Para este trabalho decidiu-se que o grupo etário a que se destina o desenvolvimento do protótipo seria principalmente o dos adolescentes, pois estes são o público alvo mais indicado para a utilização desta ferramenta pela idade e conhecimentos de tecnologia pela ótica de utilizador e pela capacidade de ler. Se bem que pode ser utilizada por utilizadores de outras faixas etárias.

Segundo (Batista, 2011) as áreas de atuação que são alvo da prevenção, avaliação, intervenção e investigação científica do TF são as seguintes (ver tabela 1):

Tabela 1: Áreas de atuação do Terapeuta da Fala (Batista, 2011)

Fala	Articulação
	Fluência
	Ressonância
	Voz (incluindo as componentes aerodinâmicas da respiração)
Linguagem	Inclui a compreensão e expressão nas modalidades oral, escrita, gráfica e gestual bem como competências de pré-literacia e de literacia. As perturbações da linguagem podem afetar todos ou alguns componentes: (1) Forma da linguagem: Morfologia, Sintaxe e Fonologia. (2) Conteúdo da linguagem: Semântica. (3) Função da linguagem: Pragmática.
Comunicação não verbal	Inclui os traços suprasegmentais e mímica; bem como a utilização de métodos/meios de comunicação aumentativa e/ou alternativa
Aspetos cognitivos da comunicação	Inclui os traços suprasegmentais e mímica; bem como a utilização de métodos/meios de comunicação aumentativa e/ou alternativa
Deglutição ou funções relacionadas	
Funções sensoriais no âmbito da comunicação, da deglutição ou de outras áreas relacionadas (exemplo: défice de audição)	

As áreas de atuação que este trabalho considerou para utilização da tecnologia de apoio a desenvolver são a da fonologia, uma das áreas da linguagem relacionada com a forma como os sons da fala se organizam nas palavras, a articulação e a ressonância que pode resultar, por exemplo, de uma fenda palatina (Rua, 2015). Espera-se ainda que este trabalho possa ajudar também ao nível da comunicação não verbal, especificamente no que se refere às trações suprasegmentais (i.e., melodia da fala durante o discurso). (Rua, 2015)

Este trabalho pretende permitir que o utilizador pratique a sua dicção (sons da fala), produzindo os diferentes sons existentes na língua portuguesa – padrão europeu. Exclui-se assim, por agora, os demais idiomas, mas havendo a possibilidade futura de este trabalho ser extensível a outros.

2.2 Tecnologias de Apoio à Terapia da Fala

Quanto às Tecnologias de Apoio, elas podem ser consideradas produtos, serviços ou práticas com o objetivo de melhorar a qualidade de vida das pessoas. (Galvão Filho, 2009) Existem atualmente no mercado algumas soluções tecnológicas de apoio à TdF. São essas mesmas soluções que se pretende levantar e analisar. Em seguida são descritas tecnologias de apoio encontradas.

2.2.1 Materiais de apoio na Terapia da Fala

Num trabalho realizado por Carneiro e colegas (2014), esta refere que os TF utilizam diversos recursos como suporte ao desenvolvimento de competências nas crianças que acompanham. Utilizam sobretudo jogos, fichas em papel e jogos em computador ou tablet. Verificou-se que, com exceção das aplicações móveis, a maioria do material utilizado é construído pelo TF, aspeto este que acarreta um dispêndio de horas pós-laborais para este fim. Do mesmo modo, os jogos utilizados em computador são criados pelos TF (por exemplo recorrendo ao programa PowerPoint da Microsoft), objetivando o aumento da motivação, atenção e concentração das crianças na realização das tarefas, favorecendo assim a sua aprendizagem. (Carneiro et al., 2014) Relativamente aos meios de computador disponíveis existentes de apoio à TdF pode-se referir que existem alguns e que eles podem ser úteis.

Furlong num total de 14 estudos e avaliação de 11 programas de Terapia da Fala baseada em computador obteve “[...]resultados que mostram que a Terapia da Fala baseada em computador está associada a mudanças clínicas positivas para algumas crianças com distúrbios dos sons da fala. Embora a base de evidências para a TdF baseada em computador ainda esteja a emergir, conclusões preliminares [podem ser retiradas do estudo]. Este tipo de terapia pode ser um complemento útil para o tratamento de alguns distúrbios de sons da fala na infância, pois ela pode beneficiar crianças com disartria ou outros problemas de articulação, de fonologia ou outros, associados à deficiência auditiva, por exemplo (Furlong et al., 2017). Um dos benefícios encontrados tem que ver com o facto de poder ser utilizado sem a presença do TF, podendo reduzir os tempos de lista de espera verificados em alguns países para aceder a profissionais de terapia da fala e permitindo que as crianças passem mais tempo a praticar de forma independente.” Um desses recursos podem ser os Assistentes Pessoais Inteligentes (IPA) controlados por voz.

2.2.2 Aplicações tecnológicas de apoio à Terapia da Fala

Existem algumas aplicações de software gratuitas desenvolvidas de apoio à TdF. Para este trabalho foram consideradas para o estado da arte aplicações que tenham como referência o português da Europa, dado que o protótipo a desenvolver será desenvolvido para este idioma. Algumas das aplicações encontradas são exemplos de desenvolvimento académico como o caso

do EasyTefa e outras são aplicações que já são comercializadas em lojas de aplicativos móveis tais como, a App Store² da Apple ou a Play Store³ da Google como a aplicação Happies.

- **EasyTefa**

A aplicação Easytefa foi desenvolvida por Ruben Ribeiro (Ribeiro, 2019) para a dissertação de mestrado e desenvolvida com o objetivo de ajudar o Terapeuta da Fala na sua atividade e assim apoiar o paciente (principalmente crianças, mas também adaptável a adultos) no tratamento com a proposta de exercícios de comparação de imagens. Quando o paciente disser um dos nomes associados às imagens, o EasyTEFA reconhece e dá uma resposta visual ao utilizador, de modo a este poder ter uma melhor noção do erro e num cenário ideal, levá-lo a interessar-se a aprender mais sobre como superar o problema.

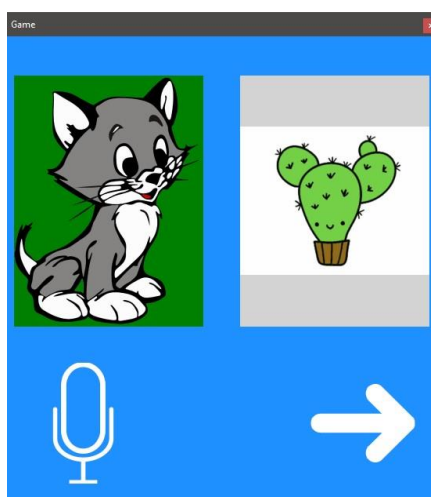


Figura 3 – Exemplo de ecrã de Jogo do EasyTefa⁴

- **Happies**

A Happies é uma aplicação paga desenvolvida pela empresa Happies, uma *startup* incubada no UPTEC – Parque de Ciência e Tecnologia da Universidade do Porto. A aplicação está orientada para crianças que está disponível para os sistemas operativos iOS e Android. Ela é constituída

² Disponível em <https://www.apple.com/app-store/>

³ Disponível em <https://play.google.com/>

⁴ Imagem disponível em <https://recipp.ipp.pt/handle/10400.22/15683>

por várias animações para apoiar o desenvolvimento da motricidade das bochechas, lábios e língua, e está disponível nos idiomas inglês, francês, espanhol e português.

Funcionando como um espelho, a criança pode ver ao mesmo tempo a animação e os movimentos que está fazendo com as bochechas, lábios ou língua.

Para avaliar melhor a evolução de cada criança, o utilizador pode registar cada exercício e avaliar o desempenho usando um sistema de pontuação.

Permite o registo de vários utilizadores onde depois posteriormente podem ser avaliados pelo desempenho de cada utilizador numa área de resultados.



Figura 4 – Exemplo de ecrã da aplicação Happies⁴

- **Terapia da Fala: Exercícios de Articulação Verbal**

Já em relação à App “Terapia da Fala: Exercícios de Articulação Verbal” ela foi desenvolvida pela PMQ Software para os sistemas operativos Android, iOS e iPadOS não sendo totalmente gratuita, ela foi pensada para crianças e também para adultos utilizando um sistema de pontuação baseado em 4 jogos em idioma português europeu. Aconselha a que as crianças sejam acompanhadas pelo TF, “promovendo o treino articulatório de 19 sons da fala em diferentes posições da palavra de uma forma divertida e dinâmica, através de jogos e exercícios com imagens reais e apelativas”.

⁴ Imagem disponível em <https://happies.pt/en/home2-2/>



Figura 5 – Exemplo de ecrã da aplicação Terapia da Fala⁵

- **Vantagens e Desvantagens das Tecnologias de Apoio em Terapia da Fala**

As principais vantagens encontradas estão relacionadas com o aumento da “motivação, atenção e concentração e a possibilidade de concretizar conceitos abstratos difíceis de explicar. Do mesmo modo, consideram que se a aplicação também for utilizada efetivamente fora da terapia, será uma mais-valia no que se refere à generalização de competências para os diferentes contextos em que a criança participa. As desvantagens mais relevantes dizem respeito à limitação de aplicações adequadas para grande parte das áreas de intervenção do TF, a dificuldade em adaptar as aplicações existentes e os preços elevados de algumas das já existentes que, mesmo assim, ainda são incompletas para as necessidades encontradas.” (Carneiro et al., 2014)

Da pesquisa realizada foi possível apurar que não existem muitas soluções tecnológicas disponíveis de apoio à TdF que utilizem o reconhecimento de fala.

2.3 O processo de reconhecimento de fala

A necessidade de realizar o levantamento do estado da arte sobre o processo de reconhecimento de voz passa assim por entender melhor como funciona a conversão da voz em texto e verificar se é viável a sua utilização no campo da TdF.

⁵ Imagem disponível em <http://www.terapia-da-fala-exercicios.pt/>

Quase desde o início da criação dos computadores que as pessoas queriam falar para os mesmos. Neste momento podemos afirmar que o reconhecimento de fala automático (ASR) ajudou a tornar isso uma realidade. Ou seja, já é possível interatuar com máquinas utilizando certas tecnologias, que iremos ver mais à frente.

Para Cook, (Cook, 2002) o processo de reconhecimento de voz é um processo pelo qual o computador (ou outro tipo de máquina) identifica as palavras ditas. Basicamente, significa falar para um dispositivo e ter um reconhecimento correto do que se está a dizer.

2.3.1 Terminologia básica

No seguinte ponto iremos descrever alguns termos básicos, tipos e utilizações feitas da tecnologia de reconhecimento de fala.

Alguns dos termos básicos/expressões utilizadas na tecnologia dos sistemas de reconhecimento de voz (SRV) são os seguintes:

- **Utterance:** é a vocalização (fala) de uma palavra ou algumas palavras, uma frase ou várias frases que representam um único significado para o computador;
- **Dependência do orador:** Os sistemas com dependência do orador são criados à volta do utilizador que fala;
- **Vocabulários** ou **dicionários:** são listas de palavras/*utterance* que podem ser reconhecidas pelo ASR;
- **Precisão:** A capacidade de um ASR pode ser medida quanto à sua precisão, na identificação correta das palavras/*utterance*;
- **Training/ Aprendizagem:** Alguns ASR são mais precisos para quem está previamente identificado do que para quem ainda não está. Ele pode aprender com a repetição de frases/palavras.

Esta análise pretendeu introduzir termos utilizados nos ASR e elucidar-nos sobre estes mesmos sistemas, para assim poder entender melhor o seu funcionamento do reconhecimento de fala.

2.3.2 Tipos de reconhecimento de voz

Consideramos ser também necessário identificar os tipos de reconhecimento de voz existentes, pois no passado os sistemas de reconhecimento eram divididos em classes para o caso concreto das *utterance* que eles eram capazes de reconhecer.

- **Palavras isoladas:** Os ASR de palavras isoladas normalmente requerem que cada *utterance* tenha um período de silêncio no início e outro no fim, ou seja falta de um sinal de áudio:
- **Palavras conectadas:** Sistemas de palavras conectadas (ou em inglês ‘connected utterance’) são semelhantes aos sistemas de palavras isoladas, mas permitem que falas separadas possam ser executados conjuntamente com uma pausa mínima entre elas:
- **Fala contínua:** No passado o reconhecedor de fala contínua eram os mais difíceis de conceber, o que na atualidade já não acontece. O reconhecedor de fala contínua permite que os utilizadores falem naturalmente, enquanto o sistema determina o conteúdo. Basicamente funciona como um ditado para o computador:
- **Fala espontânea:** Basicamente pode ser pensado como uma fala que soa natural e não ensaiada. Um sistema ASR com capacidade de detetar fala espontânea deve ser capaz de lidar com uma variedade de recursos naturais da fala, como palavras sendo executadas juntas, como por exemplo “Ummms” e “Ahhhhhs” e até gaguejos leves.

É importante mencionar que alguns sistemas ASR têm a capacidade de identificar utilizadores específicos quanto à sua voz, mas não é relevante por agora para o desenvolvimento deste trabalho. Neste caso o tipo de reconhecimento a utilizar tenderá a ser do tipo fala contínua.

2.3.3 Usos e Aplicações

Algumas das utilizações mais relevantes dadas aos sistemas ASR são as seguintes:

- **Ditado:** Ditado é um uso comum para os sistemas ASR atuais, um exemplo prático concreto é o do ditado no programa *Word* do *Microsoft Office*.

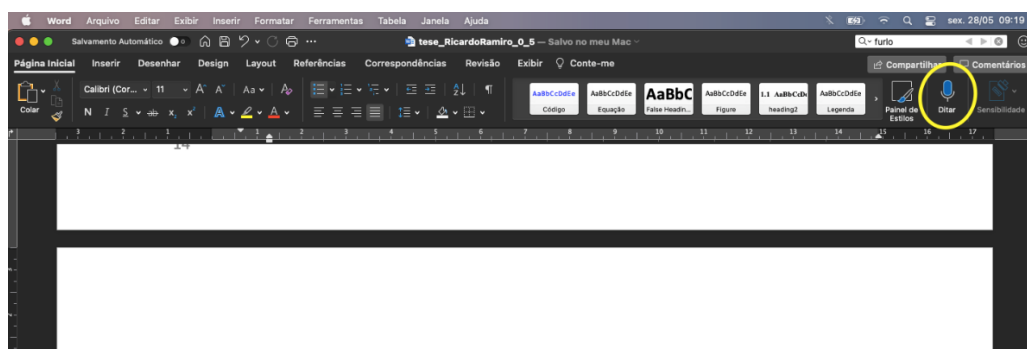


Figura 6 – Exemplo de ecrã do programa Word da Microsoft

- **Telecomunicações:** Alguns *voicemails* permitem a quem telefona dizer certos comandos em vez de premir as opções no ecrã ou teclado para escolher as opções;
- **Wearables:** Já se utiliza voz para executar certas ações nos dispositivos *wearables*, tais como nos *smartwatches*, exemplo Apple Watch;
- **Médicos/Deficiências:** Muitas pessoas têm dificuldades em digitar/datilografar e falar corretamente. As pessoas com problemas de fala podem utilizar o sistema para converter o discurso em texto e posteriormente irem corrigindo a sua fala. Este é o ponto em que este trabalho pretende contribuir;
- **Comando e controlo (Assistentes Pessoais Inteligentes controlados por voz):** São sistemas ASR criados para executar certas funções, exemplo os Assistentes Pessoais Inteligentes vulgo IPA. São designados por sistemas de comando e controlo.

Assistentes Pessoais Inteligentes controlados por voz

Nos últimos anos temos assistido a uma proliferação de soluções baseadas em comunicação por voz, normalmente designadas por assistentes pessoais inteligentes.

Mathew Hoy (Hoy, 2018) define os assistentes de voz como sendo a “[...] realização do sonho de ficção científica de interagir com os computadores, conversando com eles. A Siri da Apple, Microsoft Cortana, a Alexa da Amazon e Google Assistant são agentes de Software que funcionam em dispositivos de alto-falantes ou em certos *smartphones*. O software escuta constantemente uma palavra-chave para despertá-lo. Depois de ouvir essa palavra-chave, grava a voz do utilizador e envia-a para um servidor especializado, que processa e interpreta-o como um comando, o servidor por sua vez irá fornecer ao assistente de voz (IPA) as informações apropriadas para serem lidas de volta ao utilizador, reproduzindo o áudio solicitado pelo

utilizador ou concluindo as tarefas com os vários serviços e dispositivos conectados. O número de serviços que suportam voz estão a crescer rapidamente, e os fabricantes de dispositivos da internet das coisas (Internet of Things - IoT) também estão a incorporar o controlo de voz nos seus produtos. Como referido anteriormente nos *IPAs* podemos encontrar distintas soluções com alguma notoriedade entre as quais a Alexa desenvolvida pela Amazon, a *Siri* da *Apple*, a *Google Assistant* e a *Cortana* da *Microsoft*. Em seguida são analisados e comparados os distintos *IPAs* mencionados anteriormente.

- **A Siri**

A assistente Siri foi criado pela Apple e é o *IPA* que existe há mais tempo, ele foi lançado em 2010 e fazia parte do sistema operativo iOS em 2011.

Segundo a Apple a “Siri é um assistente inteligente que oferece uma maneira mais rápida e fácil de fazer as coisas. Permite fazer chamadas ou enviar mensagens, definir alarmes, temporizadores ou lembretes, obter direções, ver a agenda. Possibilita a criação de rotinas e atalhos para aplicações. É compatível com a aplicação de música *Apple Music*. Permite controlar alguns aparelhos inteligentes, tal como o homePod. A Siri responde a qualquer tipo de pergunta. Segundo a Apple a Siri está constantemente a aprender e em saber como ajudar mais. Ela mantém a informação privada e segura, e só podendo ser utilizada com dispositivos Apple.”(Apple, s.d.)

- **A Cortana**

A Microsoft lançou a Cortana⁶ em 2013 como sendo um assistente de produtividade pessoal da Microsoft com o objetivo de ajudar as pessoas a economizar tempo e concentrar a atenção no que é mais importante.

A Cortana permite gerir a agenda pessoal e manter atualizado o seu conteúdo. Criar e gerir listas de tarefas. Definir lembretes e alarmes. Encontrar factos, definições, informações e abrir aplicativos no computador.

⁶ Disponível em <https://support.microsoft.com/pt-pt/topic/what-is-cortana-953e648d-5668-e017-1341-7f26f7d0f825>

- **A Alexa**

A *Alexa* foi lançada conjuntamente com o dispositivo Echo em 2014. Ela é um serviço de voz baseado na nuvem propriedade da Amazon e que está disponível em dezenas de milhões de dispositivos da Amazon e de fabricantes de dispositivos terceiros. (Amazon Developers, 2020)



Figura 7 – Amazon Echo de quarta geração em preto (2020)⁷

- **O Google Assistant**

O Google Assistant por sua vez e tal como o nome indica desenvolvido pela Google, foi anunciado em 2016 juntamente com o dispositivo Google Home e é muito semelhante à Amazon Alexa. Ele está também disponível em vários dispositivos, tais como, *smartphones* mas com o sistema operativo *Android*, e outros alta-vozes inteligentes, ecrãs inteligentes “*smart displays*”, nos carros, nas televisões inteligentes, nos computadores, nos relógios inteligentes - *smartwatches* e em outros dispositivos. Segundo a Google, o Assistant tal como a Siri foi desenvolvido para manter a informação privada, segura e protegida.

⁷ Imagem disponível em <https://www.google.com/url?sa=i&url=https%3A%2F%2Freciclaiteiros.com.br%2Falexa-ler-livro%2F&psig=AOvVaw0v4fBvCPahkjIGUz2wXib8&ust=1634411701317000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCJCg0JqQzfMCFQAAAAAdAAAAABAD>



Figura 8 – Google Nest Hub de segunda geração (2016)⁸

- **A Bixby**

A Bixby é um assistente inteligente virtual sob a forma de aplicação móvel que está presente em diversos smartphones da Samsung e que tem diversas funções, como comandos de voz, recordatórios, reconhecimento de objetos (Bixby Vision) e controlo de dispositivos inteligentes de casa.

A Bixby foi pela primeira vez utilizada no Galaxy S8 e S8+ permitindo interatuar usando a voz, texto ou toques, estando integrado ao telefone, o que permite à Bixby realizar muitas das tarefas que se faz no smartphone. Para acioná-la existe um botão dedicado de Bixby à esquerda, o que é uma maneira prática de aceder. Um toque curto no botão Bixby posiciona na página principal da Bixby.

À data a Bixby está disponível no Galaxy S20, S20+, S20 Ultra, Note10, Note10+, S10e, S10, S10+, Fold, Note9, S9, S9+, Note9, S8 e S8+.

⁸ Imagem disponível em https://store.google.com/us/product/nest_hub_2nd_gen?hl=en-US

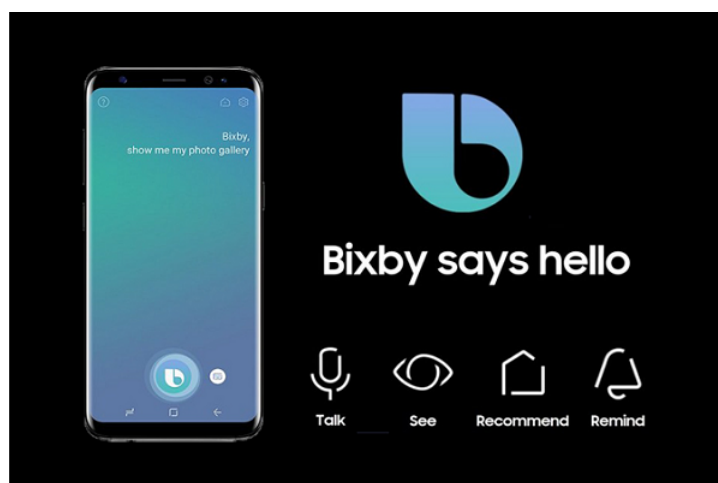


Figura 9 – Ecrã de utilização da Bixby⁹

Cada IPA tem características próprias exclusivas, mas o núcleo de funções são as mesmas, necessitando sempre de estar conectados à internet.

Em seguida são apresentados algumas das funções realizadas pelos IPAs:

- Enviar e ler mensagens de texto, fazer chamadas e enviar e ler emails;
- Responder a perguntas básicas de informação ("Que horas são? Qual a previsão para o tempo?");
- Definir temporizadores, alarmes e entradas de calendário;
- Definir lembretes, fazer listas e fazer cálculos matemáticos básicos;
- Controlar a reprodução de música e outros recursos multimédia, tais como, Google Play, iTunes, Pandora, Netflix e Spotify;
- Controlar outros dispositivos inteligentes, como termostatos, luzes, alarmes entre outros;
- Contar histórias.

Num estudo de Berdasco de 2019, foram apresentados "[...] resultados da avaliação dos quatro assistentes pessoais inteligentes [Alexa, Google Assistant, Siri e Cortana] em duas dimensões: a

⁹ Imagem disponível em <https://www.samsung.com/pt/apps/bixby/>

exatidão das respostas e quão natural sentem as respostas os utilizadores. Noventa e dois indivíduos participaram nesta avaliação. Os resultados mostraram que a Alexa e o Google Assistant são significativamente melhores que a Siri e a Cortana quanto aos pontos referidos anteriormente. Contudo, também não existe uma diferença relevante entre a Alexa e o Google Assistant.” (Berdasco et al., 2019)

Já para Hoy (Hoy, 2018) “a complexidade e precisão da tecnologia de reconhecimento de voz e dos IPAs cresceram exponencialmente nos últimos anos. Atualmente os assistentes de voz disponibilizados pela Apple, Amazon, Google e Microsoft permitem que os utilizadores façam perguntas e emitam comandos para computadores em linguagem natural. Existem muitos usos futuros possíveis desta tecnologia, desde a automatização residencial até ao suporte para companhia de idosos. Contudo [refere que e ao contrário do que mencionam a Google e a Apple] também pode haver problemas ao nível de privacidade e segurança.”

O mais relevante é que os IPAs podem ser uma tecnologia que poderá ajudar no futuro indivíduos com problemas de fala e outros tipos de incapacidades. Pradhan revela num estudo de 2018 que “[...] em termos de terapia da fala alguns comentários descreveram como a utilização de *IPAs* ajudou os utilizadores com problemas de fala a falar mais devagar, com maior clareza ou a falar mais alto. A natureza conversativa dos dispositivos foi vista como uma ajuda”. (Pradhan 2018). Noutro estudo ele refere também que os indivíduos expressaram entusiasmo à cerca de dispositivos inteligentes e que estes têm potencial para resolver problemas de acessibilidade do mundo real.

2.3.4 Funcionamento dos Sistemas Reconhecimento de Voz

O Sistema de reconhecimento de voz, sistema automático de reconhecimento de voz (ASR), reconhecimento de voz por computador, ou *speech-to-text* é a capacidade que permite a um programa processar a fala humana num formato escrito e está centrado na tradução da fala para um texto. Existe também o *text-to-speech* que converte texto em áudio, mas não iremos aprofundar neste trabalho pois ele não será utilizado na criação do protótipo. Mas é importante referir que esta tecnologia é utilizada pelos *IPAs* e que pode ajudar os indivíduos com dificuldades da fala, no caso concreto de existir uma solução de diálogo entre a máquina e o humano (HCI), sendo esta uma solução interativa e com potencial de ajuda. Aqui os *IPAs* poderiam ser de bastante ajuda, estivessem estes disponíveis para os programadores através de um SDK de desenvolvimento e seriam utilizados no desenvolvimento do protótipo.

Centrando-nos no funcionamento da tecnologia ASR a um alto nível, ela recorre ao áudio, mais concretamente à gravação de uma amostra digital convertendo o sinal analógico do microfone num sinal digital através do conversor de som analógico para digital da placa de som. Quando um microfone está a funcionar, as ondas sonoras fazem vibrar o elemento magnético do microfone, causando uma corrente elétrica na placa de som. Basicamente o conversor regista o valor da tensão elétrica em intervalos específicos. (Cook, 2002)

Para Cook, (Cook, 2002) os sistemas de reconhecimento de voz podem ser divididos em dois tipos principais. Os sistemas de reconhecimento de padrões que comparam padrões a padrões conhecidos com o intuito de encontrar uma correspondência e os sistemas fonéticos acústicos que utilizam o conhecimento do corpo humano (produção da fala e da audição) para comparar as características da fala (fonética, como sons de vogais). A maioria dos sistemas modernos concentra-se na abordagem de reconhecimento de padrões devido a combinar muito bem com as técnicas de computação atuais e a terem uma melhor precisão.

Segundo o mesmo os sistemas de reconhecimento podem ser divididos em cinco partes:

1. **Gravação de áudio e deteção da *utterance***: pode ser realizado de várias maneiras. Os pontos de partida podem ser encontrados comparando os níveis de áudio do ambiente (energia Acústica em alguns casos) com a amostra recém-gravada. A deteção do *Endpoint* é mais difícil porque as colunas tendem a deixar “artefactos” como por exemplo a respiração, suspiro, ecos.
2. **Pré-filtro** (pré-ênfase, normalização, *banding*): é realizado de várias maneiras dependendo de outros recursos do Sistema de reconhecimento. Os métodos mais comuns são o método “Bank-of-filters”, que utiliza uma série de filtros de áudio para preparar a amostra, e o método de Codificação Preditiva Linear, que utiliza uma função de previsão para calcular as diferenças (erros). Diferentes formas de análise espectral também são usadas.
3. **Enquadramento e Janelas** (dividir os dados num formato utilizável): envolve a separação dos dados da amostra em tamanhos específicos. Está geralmente incluído na etapa 2 ou etapa 4. Esta etapa também envolve a preparação dos limites da amostra para análise.
4. **Filtragem** (filtragem adicional de cada janela /quadro / banda de frequência): Filtragem adicional nem sempre está presente. É a preparação final para cada janela antes da

comparação e correspondência. Frequentemente, isso consiste em alinhamento e normalização do tempo.

5. **Comparação e correspondência** (reconhecimento da utterance): Há um grande número de técnicas para realizar a comparação e correspondência. A maioria envolve a comparação da janela atual com amostras conhecidas. Existem métodos que utilizam os Modelos Ocultos de Markov (HMM), análise de frequência, análise diferencial, técnicas / atalhos de álgebra linear, distorção espectral e métodos de distorção de tempo. Todos os métodos são utilizados para gerar uma correspondência de probabilidade e precisão.

2.4 Avaliação dos Sistemas de Reconhecimento de Fala

Existem maneiras de avaliar os sistemas de reconhecimento de voz, é importante referir que “Os sistemas de reconhecimento automático de fala (ASR) têm como objetivo converter um sinal de fala numa sequência de palavras para fins de comunicação baseada em texto ou para controlo de dispositivos. O objetivo de avaliar os sistemas ASR passa por simular o julgamento humano no desempenho destes sistemas, a fim de medir sua utilidade e avaliar as restantes dificuldades e também especialmente na comparação de estes sistemas. A métrica padrão utilizada de avaliação dos ASR é a taxa de erro de palavras ou WER, que é definido como a proporção de erros de palavras em relação às palavras processadas.”(Errattahi et al., 2018)

$$WER = \frac{S + D + I}{N1} = \frac{S + D + I}{H + S + D}$$

“Word Error Rate (WER) é a métrica mais popular para avaliação dos sistemas ASR, ele mede a percentagem de erros de palavras, onde são (substituições (S), inserções (I), exclusões (D)) em relação ao número total de palavras processadas. É definido como WER =

onde I = número total de inserções, D = número total de exclusões, S = número total de substituições, H = número total de acertos e N1 = número total de palavras de entrada.

As conclusões que foram retiradas por Errattahi é de que embora os resultados sejam promissores utilizando a fórmula WER, a maioria das pesquisas está limitada à deteção e sugere correção manual dos erros. Portanto, acreditamos que há necessidade de mais investigações sobre a correção automática de erros nos ASR e que deve ser dada atenção a

questões como a eficiência, a usabilidade e a robustez dos métodos desenvolvidos. (Errattahi et al., 2018)

2.5 APIs de reconhecimento de fala

Para este trabalho considerou-se importante fazer um levantamento do estado da arte de APIs de reconhecimento de fala. Pois pretende-se verificar se existia potencial para utilizá-la no protótipo.

Uma *API* significa Interface de Programação de Aplicativo e funciona como um conjunto de instruções de programação para aceder a uma ferramenta ou *software* que está disponível na *Internet*. Normalmente uma empresa de *software* disponibiliza a sua *API* ao público para que possa ser utilizado pela comunidade de programadores.

Como mencionado anteriormente este trabalho pretendeu estudar e utilizar as APIs de reconhecimento de voz para a criação do protótipo, surgindo assim a necessidade de encontrar e avaliar APIs (*Application Programming Interface*) capazes de auxiliar a fase de desenvolvimento. Neste caso o objetivo passa por comparar várias APIs que possam ser utilizadas e encontrar a que melhor serve os interesses deste projeto do ponto de vista de execução, tais como o custo, a facilidade de utilização da tecnologia e a possibilidade de ter suporte ao idioma português europeu.

Segundo foi possível pesquisar existe um conjunto de APIs de Reconhecimento de Voz disponíveis no mercado, entre elas podemos encontrar as *frameworks da CMU Sphinx, Google Cloud Speech-to-text, Microsoft Speech to text, Mozilla Webspeech API, IBM Watson Speech to Text* entre outras.

2.5.1 CMU Sphinx¹⁰

Quanto ao CMU Sphinx ele é um projeto *open source* com mais de 20 anos de existência, desenvolvido pela *Carnegie Mellon University* (CMU) e funciona *offline* com finalidade comercial ou de investigação. Ele é composto por um conjunto de bibliotecas e ferramentas de desenvolvimento de reconhecimento de voz que podem ser utilizados. Ela possui várias versões tais como o Sphinx2 que pode ser usado para reconhecimento baseado em comunicações telefônicas baseado em diálogo. Ou o Sphinx3 que é um pouco mais lento e mais preciso na decodificação do reconhecimento de voz. O Sphinx permite a criação de modelos linguísticos através de um ficheiro que contém palavras e frases que se pretende reconhecidas ao contrário de outras soluções.

2.5.2 Google Cloud Speech-to-text¹¹

Já o Google Speech-to-text é uma API que está disponível desde 2018 e permite a utilização de algoritmos de *deep learning*, para tal a Google adquiriu várias empresas de *deep learning*. Esta API suporta mais de 125 idiomas entre os quais o português europeu e é de livre utilização até um limite de sessenta minutos por mês. Ele permite a implementação na nuvem ou num servidor local e funciona em tempo real. Basicamente o *Google Cloud Speech to text* é um serviço que tal como o nome indica permite aos programadores converter áudio em texto, aplicando modelos de redes neuronais numa API que é fácil de utilizar, transcrevendo o que é dito pelos utilizadores para um microfone (previamente ativado) ou transcrever ficheiros de áudio. O *Speech-to-text* tem três métodos principais para realizar o reconhecimento de fala. São eles o reconhecimento síncrono (REST e gRPC) envia dados de áudio para a API Speech-to-Text, realizando o reconhecimento desses dados e devolvendo os resultados depois que todo o áudio seja processado. As solicitações de reconhecimento síncrono são limitadas a dados de áudio de 1 minuto ou menos de duração.

Outro é o do reconhecimento assíncrono (REST e gRPC) que envia dados de áudio para a API Speech-to-Text e inicia uma operação de longa duração. Usando esta operação, é possível

¹⁰ Disponível em <https://cmusphinx.github.io>

¹¹ Disponível em <https://cloud.google.com/speech-to-text>

pesquisar periodicamente os resultados do reconhecimento. É utilizado mais para dados de áudio de qualquer duração até 480 minutos.

O reconhecimento de streaming (apenas gRPC) realiza o reconhecimento de dados de áudio fornecidos num stream bidirecional gRPC. As solicitações de streaming são projetadas para fins de reconhecimento em tempo real, como por exemplo capturar o áudio ao vivo de um microfone. O reconhecimento de streaming fornece resultados provisórios enquanto o áudio está a ser capturado, permitindo que o resultado apareça, por exemplo, quando o utilizador ainda está a falar.

É importante referir que ele permite também configurar o número de alternativas, sendo uma ou mais. Se o Speech-to-Text determinar que uma alternativa tem um valor de confiança suficiente, essa alternativa será incluída na resposta. A primeira alternativa na resposta é sempre a melhor (mais provável) alternativa.

Se se definir `maxAlternatives` com um valor superior a 1 não implica nem garante que várias alternativas serão retornadas. Em geral, mais do que uma alternativa é mais apropriada para fornecer opções em tempo real aos usuários que obtêm resultados por meio de uma solicitação de reconhecimento de streaming.

Existe também o valor de confiança que é uma estimativa entre 0,0 e 1,0, sendo a palavra com o valor de confiança maior a mais provável de ser a que foi dita pelo utilizador.

Também é possível passar o áudio em vez de ser pelo micro por um ficheiro de áudio através de uma URI.

Para o idioma existe a propriedade `languageCode` que pode suporta uma variedade de idiomas e dialetos.

2.5.3 Microsoft Speech to Text¹²

A *Microsoft Speech to Text* faz parte dos serviços cognitivos do *Azure* e converte voz em texto para mais de 85 línguas e dialetos.

A API permite converter áudio em texto a partir de várias fontes tais como microfones, ficheiros de áudio. Permite também adicionar palavras específicas ao vocabulário base, assim como implementar na *Cloud* ou localmente (*on premise*).

¹² Disponível em <https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/>

2.5.4 IBM Watson Speech to text¹³

IBM Watson não é gratuito, funciona offline e em tempo real. Ela tem três interfaces para programadores, são elas WebSocket, Http Rest e Interface Http assíncrona. Pode ser implementado também na nuvem, mas permite um número limitado de idiomas.

2.5.5 Mozilla Webspeech HTML API¹⁴

No caso da *API Mozilla Web Speech*, ela tem como objetivo disponibilizar entradas de voz para texto e conversão de texto para voz, ou seja, funciona tanto para *reconhecimento de voz* como para síntese. Ela oferece suporte a reconhecimento e síntese baseados num servidor ou no lado cliente já incorporado. A API da Mozilla tem como base a API da Google, uma vez que fazem parte do mesmo consórcio. A API Web Speech permite incorporar dados de voz em aplicações web. A Web Speech API tem duas partes: *SpeechSynthesis* (Text-to-Speech) e *SpeechRecognition* (Asynchronous Speech Recognition) e é uma tecnologia experimental. A API Web Speech torna possível que a aplicação web seja capaz de lidar com dados de voz. Existem dois componentes para a API, *SpeechRecognition* e *SpeechSynthesis*. Em relação ao *SpeechRecognition* que é o que importa para este trabalho, ele faz o acesso por meio da interface *SpeechRecognition*, que por sua vez fornece a capacidade de reconhecer o contexto de voz de uma entrada de áudio (normalmente por meio do serviço de reconhecimento de voz padrão do dispositivo) e responde adequadamente. Geralmente, será usado o construtor da interface para criar um novo objeto *SpeechRecognition*, que possui vários manipuladores de eventos disponíveis para detetar quando a fala é inserida por meio do microfone do dispositivo. A interface *SpeechGrammar* representa um *container* para um determinado conjunto de gramática que o aplicativo deve reconhecer. A gramática é definida utilizando *JSpeech Grammar Format* (JSGF).

Em seguida encontra-se uma tabela comparativa das APIs.

¹³ Disponível em <https://www.ibm.com/cloud/watson-speech-to-text>

¹⁴ Disponível em <https://wicg.github.io/speech-api/#introduction>

Tabela 2: Comparação de APIS de reconhecimento de voz

	<i>CMU Sphinx</i>	<i>Google Cloud Speech</i>	<i>Microsoft Speech to Text API</i>	<i>IBM Watson Speech</i>	<i>Mozilla Webspeech HTML API</i>
Gratuita	Sim, Open source	Não, mas oferece até 60 minutos por mês gratuitamente	Não, mas oferece 5 horas gratuitas por mês	Não, mas tem 500 minutos grátis por mês	Sim, Open source
Speech Recognition	Sim	Sim	Sim	Sim	Sim
Synthesis (texto-to-speech)	Não	Sim	Sim	Sim	Sim
Suporta Idioma pt-PT	Não	Sim	Sim	Não, apenas pt-BR	Sim
Documentação	Sim	Sim	Sim	Pouca	Sim, extensa
Tecnologia	Java, Python	Rest API	Visual Basic	Qualquer, REST/JSON service	Javascript

Num paper de (Kěpuska & Bohouta, 2017), “[...] foi possível retirar conclusões [através de uma ferramenta] que construíram para testar o Sphinx-4, a Microsoft Speech API e Google Cloud Speech API do Google, utilizando algumas gravações de áudio que foram selecionadas em muitos lugares com as frases originais, mostrou que o Sphinx-4 atingiu uma percentagem de 37 % de taxa de erro por palavra (WER), enquanto que a Microsoft Speech to text API atingiu 18% de WER e a Google API atingiu 9% de WER. Portanto, pode-se afirmar que a modelagem acústica e o modelo de linguagem do Google são superiores.”

Capítulo 3 – Análise de valor

Neste capítulo é introduzida a análise de valor e realizada sobre o assunto desta dissertação.

A análise de valor (VA) realizada pretende focar primeiro no valor e só depois no custo, ela “[...] é considerada um processo, ao contrário de uma técnica simples, pois é uma abordagem organizada para melhorar o lucro das aplicações do produto ou serviço e utiliza muitas técnicas diferentes para atingir este objetivo. A abordagem VA é quase universal e pode ser usada para analisar produtos ou serviços existentes oferecidos por empresas de produção ou de serviços. A VA pode ser definida como um processo de revisão sistemática que é aplicada a projetos de produtos, com a finalidade de comparar a função do produto exigido por um cliente tendo em conta os requisitos ao menor custo possível sem pôr em causa o desempenho e confiabilidade necessária.” (Rich & Holweg, 2000)

“Qualquer tentativa de melhorar o valor de um produto deve considerar dois elementos, o primeiro diz respeito ao uso do produto conhecido como valor de uso (*Use value*) e a segunda fonte de valor vem da propriedade (*Esteem value*).” (Rich & Holweg, 2000)

3.1 Fuzzy Front End

Para (Herstatt, 2014) a rápida transformação de tecnologias em novos produtos ou processos é um dos desafios centrais para qualquer empresa baseada em tecnologia. Dentro do processo de inovação, acreditamos, as fases iniciais ("front end difuso") tem o maior impacto sobre o todo o processo e o resultado (Processo de entrada-saída), pois vai influenciar extremamente o design e custos totais da inovação. No entanto, o “Fuzzy Front End” é muito estruturado para o processo de inovação, tanto em teoria como na prática.

3.2 NCD - Novo Conceito de Desenvolvimento

O modelo NCD (P. A. Koen et al., 2015) “divide a extremidade dianteira em três áreas distintas: o motor, a roda e o aro (Figura 5). O motor, no centro do modelo, fornece energia para o *front end of innovation*. O motor consiste em dois segmentos - atributos organizacionais, equipas e

colaboração. A roda, a parte interna do modelo, contém os cinco elementos de atividade do front end: identificação de oportunidade, análise da oportunidade, geração de ideias, seleção de ideias, e definição de conceito. O terceiro elemento, a borda, inclui os fatores ambientais que influenciam o motor e molda os cinco elementos da atividade. Isso inclui a capacidade organizativa da empresa, ameaças da concorrência, cliente e tendências mundiais, mudanças regulatórias e a profundidade e força de aplicar ciência e tecnologia.

Em contraste a outros processos lineares, o modelo é circular para indicar que as ideias fluem, circulam e interagem entre os cinco elementos. As setas apontando para o modelo representando os pontos de partida para projetos e indicar que os projetos podem começar em qualquer oportunidade para identificação ou geração e enriquecimento de ideias. Os projetos saem para entrar no processo de desenvolvimento de novos produtos ou tecnologia.”

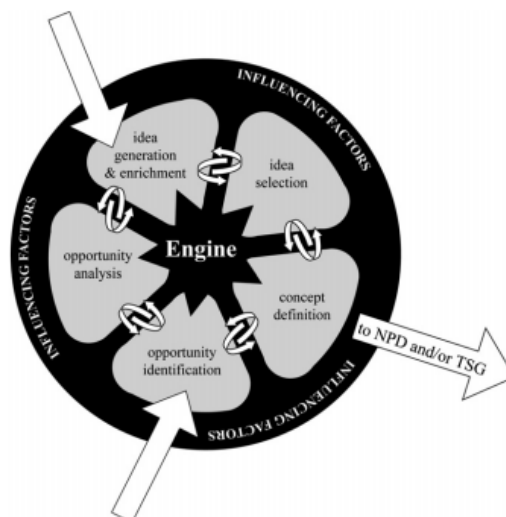


Figura 10 – Modelo NCD (P. A. Koen et al., 2015)

3.2.1 Identificação da Oportunidade

Neste tópico pretende-se descrever em que consiste a identificação de oportunidades para o problema em estudo.

Para Koen (P. Koen et al., 2001) a identificação da oportunidade é o ponto onde “ as organizações identificam as oportunidade que a empresa pode querer ou perseguir. As oportunidades de negócio ou tecnológicas são explicitamente consideradas para que os recursos sejam alocados em novas áreas de crescimento de mercado e/ou eficácia e eficiência operacional. Este elemento normalmente é impulsionado pelos objetivos de negócio. Por exemplo, a oportunidade pode ser uma resposta de curto prazo a uma ameaça competitiva,

uma possibilidade de avanço para obter vantagem competitiva ou um meio de simplificar/acelerar/ reduzir o custo das operações. A oportunidade pode ser uma direção inteiramente nova para o negócio ou uma pequena atualização de um produto existente. Também pode ser uma nova plataforma de produto, um novo processo de fabricação, uma nova oferta de serviço ou uma nova abordagem de marketing ou vendas.

As fontes e métodos que uma empresa utiliza para identificar oportunidades que deseja perseguir são a essência deste elemento. Pode haver um processo formal de identificação de oportunidade alinhado com todos os fatores de influência. Ferramentas e técnicas de criatividade (por exemplo, *brainstorming*, mapeio mental e pensamento lateral), bem como técnicas de resolução de problemas (por exemplo análise causal [...]) podem ser utilizados.”

O problema em estudo verifica se existe potencial tecnológico para ajudar/facilitar o trabalho dos terapeutas da fala em apoiar os seus pacientes com problemas de fala, pretendemos contribuir para a resolução deste problema através de um estudo suportado por um protótipo que se pretende desenvolver com profissionais da área da Terapia da Fala. Anteriormente foi mencionado que existe possivelmente potencial tecnológico para ajudar pessoas com problemas na fala, através da utilização de APIs de reconhecimento de voz, a análise de valor pretende ajudar neste mesmo ponto e elucidar qual a melhor API de reconhecimento de fala a utilizar no desenvolvimento do protótipo.

3.2.2 Análise de Oportunidade

Segundo Peter Koen (P. Koen et al., 2001) a “análise de oportunidade necessita de informações adicionais para traduzir a identificação da oportunidade em negócios específicos e oportunidades de tecnologia fazendo avaliações precoces e frequentemente incertas de tecnologia e mercado. [...] A quantidade de esforço despendida depende da atratividade da oportunidade, do tamanho do esforço de desenvolvimento futuro, da adequação à estratégia e cultura de negócios e da tolerância ao risco da tomada de decisão.

O processo de análise pode ser parte de um processo formal ou pode ocorrer iterativamente em reação às oportunidades identificadas, como cenários hipotéticos. [...] Tanto a inteligência competitiva quanto as análises de tendências são amplamente utilizadas neste elemento.”

Para análise da oportunidade será realizada uma análise SWOT para identificar as forças, fraquezas, oportunidades e ameaças para o desenvolvimento do protótipo de reconhecimento

da fala para os terapeutas da fala e respetivos pacientes. De maneira a realizar a análise SWOT este trabalho pretende avaliar a concorrência e conhecer melhor o mercado da terapia da fala.

3.3 Proposta de Valor

O criador do Modelo de Negócio de Canvas foi Alexander Osterwalder. Segundo o próprio (Osterwalder, 2013) “ [...] ao contrário de outros planos de negócio, o modelo de Canvas ajuda as organizações a conduzir conversações estruturadas, tangíveis e estratégicas sobre novos negócios ou negócios existentes. Empresas globais como General Electric, Proctor & Gamble e Nestlé utilizam o Canvas para gerir a estratégia ou novos motores de crescimento. As startups por sua vez utilizam para obter um modelo de negócio acertado. O principal objetivo do Canvas é ajudar as empresas a ir além do pensamento centrado no produto e em direção ao pensamento do modelo de negócios.

Nesta secção é mostrada, de forma simples, uma visualização do modelo de negócio de Canvas. Onde é apresentada a Proposta de Valor de Canvas (Canvas Value Proposition). Através do qual se demonstram várias dimensões de base do modelo de negócio. A proposta foi completada da seguinte forma e desenvolvida para terapeutas da fala e respetivos pacientes ou utilizadores que necessitem de melhorar a fala.

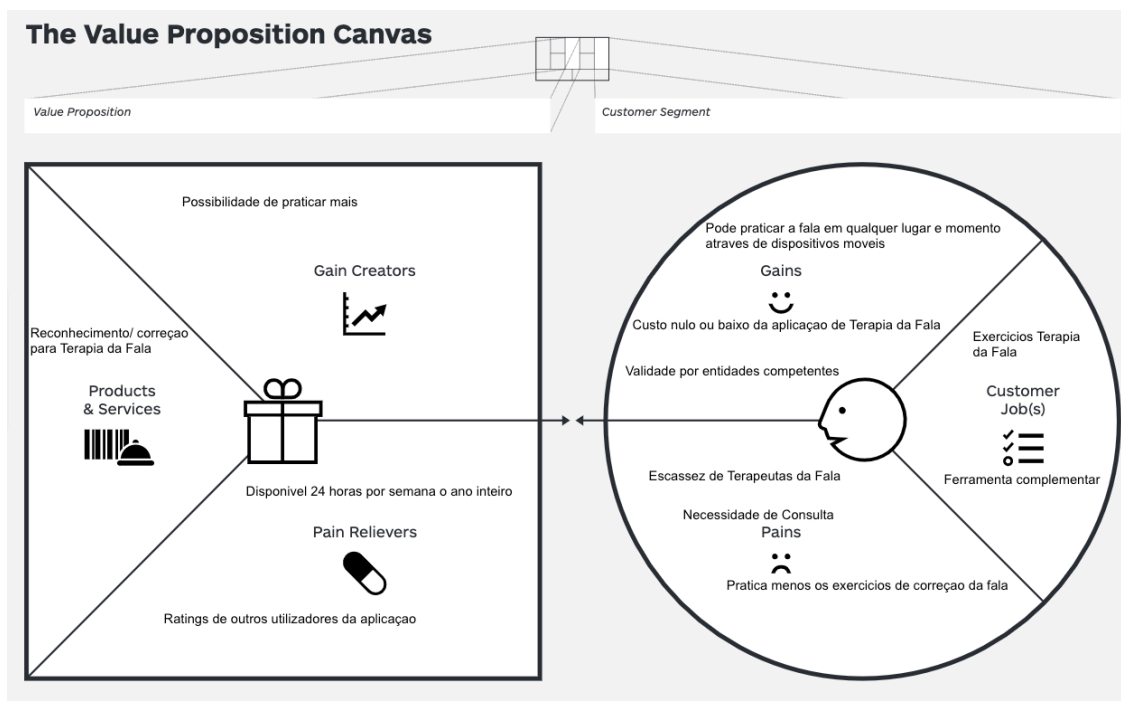


Figura 11 – Proposta de valor de Canvas

3.4 Quality Function Deployment (QFD)

O *Quality Function Deployment* (QFD) foi criado por Yoji Akao nos anos sessenta. Segundo Yoji, o QFD pode ser “[...]descrito como uma metodologia que torna todos os processos de desenvolvimento de produtos transparentes, desde a compreensão das qualidades exigidas pelos clientes até ao estabelecimento do planeamento da qualidade e determinação de qualidade do projeto.” (Akao, 2014)

“Um QFD típico pode ser dividido em quatro fases, denominadas por planeamento do produto, planeamento de implementação parcial, planeamento de processo e planeamento de produção. O QFD incorpora a voz do cliente “*customer’s voice*” em ‘requisitos de voz da empresa’ e é descrito geralmente por uma matriz de entradas (*WHATs*) e saídas (*HOWs*). [...] O QFD é um sistema virtual de planeamento multifuncional para ajudar na pesquisa e desenvolvimento de produtos ou serviços. Entre as quatro fases do QFD a *House of Quality* (HoQ) é a primeira fase. A HoQ estabelece a ponte dos *WHATs* para os *HOWs* e pode ser extensível às três fases subseqüentes.”(Peng et al., 2018)

Na figura 3 pode ser encontrada um exemplo para a estrutura de um HoQ.

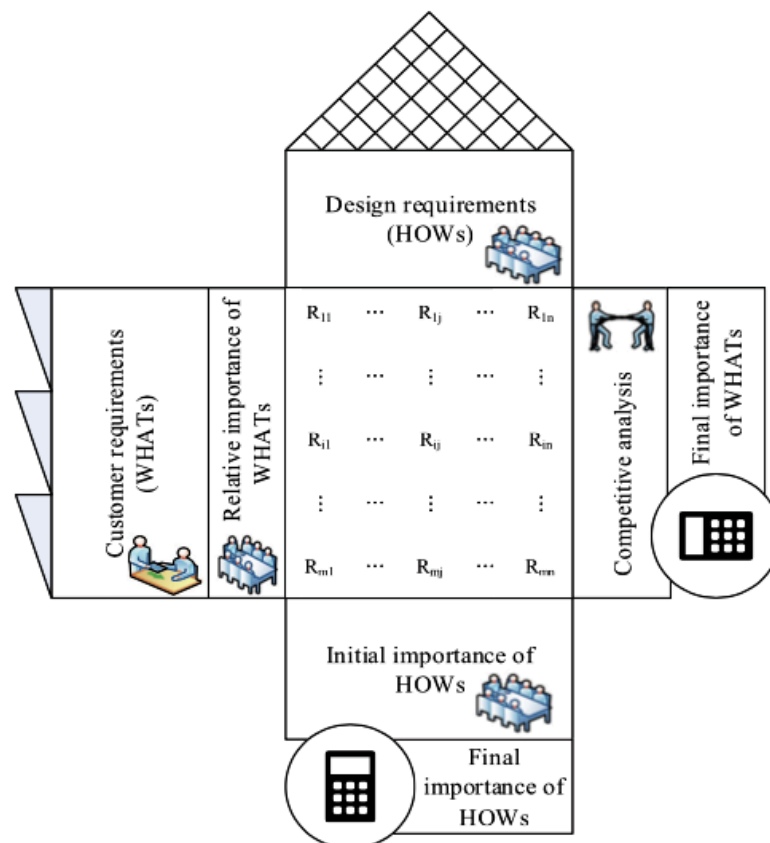


Figura 12 – Exemplo de House of Quality (HoQ) (Peng et al., 2018)

3.5 AHP – Método de Análise Hierárquica

Esta secção baseia-se no Método de Análise Hierárquica, de maneira a escolher a melhor API de reconhecimento de voz para ser utilizada neste trabalho.

O método AHP foi desenvolvido pelo matemático Thomas Saaty (Saaty, 1990) e “[...] é uma teoria de medição. Quando aplicado na tomada de decisão, ajuda a descrever a operação geral de decisão, decompondo um problema complexo numa estrutura hierárquica de vários níveis de objetivos, critérios, subcritérios e alternativas. O AHP fornece uma escala fundamental de magnitudes relativas expressas em unidades de dominância para representar julgamentos na forma de comparações emparelhadas.”

Segundo Saaty (Saaty, 2008) para realizar a tomada de uma decisão de forma organizada para gerar oportunidades é necessário decompor a decisão nos seguintes quatro passos:

1. Definir o problema e determinar o tipo de conhecimento procurado;
2. Estruturar a hierarquia de decisão do topo com o objetivo da decisão, seguido dos objetivos de uma perspectiva ampla, passando pelos níveis intermediários (critérios dos quais os elementos subsequentes dependem) até ao nível mais baixo (que geralmente é um conjunto de alternativas);
3. Construir um conjunto de matrizes de comparação entre pares, onde cada elemento num nível superior é usado para comparar os elementos no nível imediatamente abaixo em relação a ele;
4. Usar as prioridades obtidas nas comparações para pesar as prioridades do nível imediatamente abaixo. Fazer isso para cada elemento. Em seguida, para cada elemento no nível abaixo, adicionar os valores ponderados de maneira a obter a prioridade geral ou global. Continuar o processo de pesar e adicionar até às prioridades finais das alternativas.

Para a seleção da API de reconhecimento de voz foram escolhidos os seguintes **critérios**:

- **Idiomas PT Suportado:** idioma português europeu suportado pela API;
- **Curva de Aprendizagem:** a facilidade de aprendizagem na utilização da API;
- **Funcionalidades:** As funcionalidades que a API permite para sua utilização;

- **Documentação:** A quantidade e qualidade de recursos didáticos disponíveis para utilização das APIs.

As alternativas escolhidas para o método são as seguintes APIs e podem ser vistas na Figura 9:

- Web Speech API;
- CMU Sphinx;
- Google Cloud Speech to Text;
- Microsoft Speech to Text;
- IBM Watson Speech to Text.

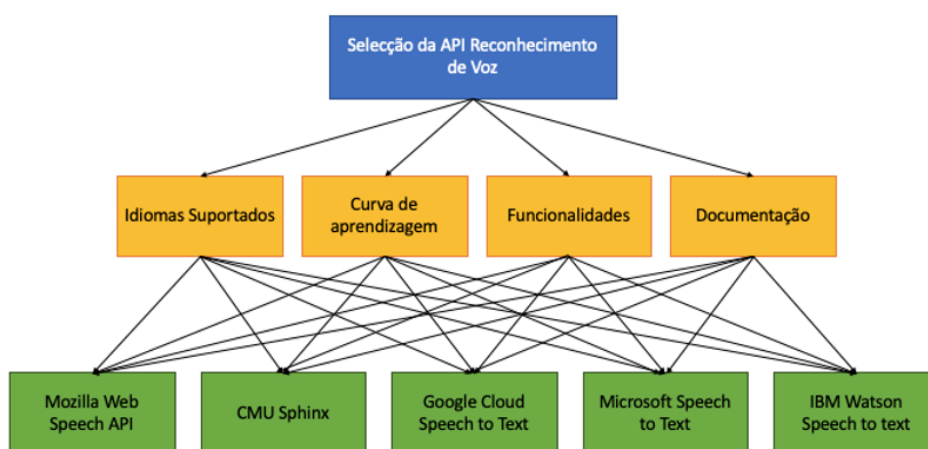


Figura 13 – Diagrama hierárquico AHP

Na figura 9 é possível visualizar a construção do diagrama hierárquico para escolha da API de reconhecimento de voz, assim como os critérios utilizados (a laranja) e as opções disponíveis de APIs (a verde).

Tabela 3: Escala Fundamental de Saaty (Saaty, 1990)

1	Igual importância	As duas atividades contribuem igualmente para o objetivo.
3	Importância pequena de uma sobre outra	A experiência e o julgamento favorecem levemente uma atividade em relação à outra.
5	Importância grande ou essencial	A experiência e o julgamento favorecem fortemente uma atividade em relação à outra.
7	Importância muito grande ou demonstrada	Uma atividade é muito fortemente favorecida em relação à outra, sua dominação de importância é demonstrada na prática.
9	Importância absoluta	A evidência favorece uma atividade em relação à outra com o mais alto grau de certeza.
2,4,6,8	Valores intermédios	Quando se procura uma condição de compromisso entre duas definições.

Por sua vez a escala fundamental de Saaty (ver tabela 2) é utilizada para definir a importância de cada critério sob a forma de peso.

Tabela 4: Tabela AHP de comparação de critérios

	Curva de aprendizagem	Documentação	Funcionalidades	Idiomas Suportados
Curva de aprendizagem	1	5	6	9
Documentação	1/5	1	6	9
Funcionalidades	1/6	1/6	1	7
Idiomas Suportados	1/9	1/9	1/7	1

λ_{max} : 4.233

RI: 0.89 CI: 0.078 CR: 0.088

Curva de Aprendizagem

Tabela 5: Tabela AHP de comparação das alternativas para a Curva de Aprendizagem

Curva de Aprendizagem	Mozilla Web Speech	CMU Sphinx	Google Cloud Speech to text	Microsoft speech to text	IBM Watson speech to text
Mozilla Web Speech	1	7	6	5	9
CMU Sphinx	1/7	1	1/5	1/5	8
Google Cloud Speech to text	1/6	5	1	5	9
Microsoft speech to text	1/5	5	1/5	1	7
IBM Watson speech to text	1/9	1/8	1/9	1/7	1

λ_{max} : 5.219

RI: 1.11 CI: 0.055 CR: 0.050

Documentação

A documentação é um critério de decisão importante na tomada de decisão da escolha da API de reconhecimento de voz. A tabela 5 apresenta a comparação de alternativas para a documentação.

Tabela 6: Tabela AHP de comparação das alternativas para a Documentação

Curva de Aprendizagem	Mozilla Web Speech	CMU Sphinx	Google Cloud Speech to text	Microsoft speech to text	IBM Watson speech to text
Mozilla Web Speech	1	7	5	7	9
CMU Sphinx	1/7	1	1/3	1/7	1/7
Google Cloud Speech to text	1/5	1/3	1	6	9
Microsoft speech to text	1/7	1/7	1/6	1	6
IBM Watson speech to text	1/9	1/7	1/9	1/6	1

λ_{max} : 5.402

RI: 1.11 CI: 0.101 CR: 0.091

Funcionalidades

Tabela 7: Tabela AHP de comparação das alternativas para a Funcionalidades

Documentação	Mozilla Web Speech	CMU Sphinx	Google Cloud Speech to text	Microsoft speech to text	IBM Watson speech to text
Mozilla Web Speech	1	5	4	5	6
CMU Sphinx	1/5	1	4	5	6
Google Cloud Speech to text	1/4	1/4	1	5	6
Microsoft speech to text	1/5	1/5	1/5	1	6
IBM Watson speech to text	1/6	1/6	1/6	1/6	1

λ_{max} : 5.426

RI: 1.11 CI: 0.107 CR: 0.096

Idiomas Suportados

Tabela 8: Tabela AHP de comparação das alternativas para os Idiomas Suportados

Idiomas Suportados	Mozilla Web Speech	CMU Sphinx	Google Cloud Speech to text	Microsoft speech to text	IBM Watson speech to text
Mozilla Web Speech	1	6	5	4	7
CMU Sphinx	1/6	1	5	4	5
Google Cloud Speech to text	1/5	1/5	1	5	7
Microsoft speech to text	1/4	1/4	1/5	1	7
IBM Watson speech to text	1/7	1/5	1/7	1/7	1

λ_{\max} : 5.429

RI: 1.11 CI: 0.107 CR: 0.096

Resultados

Para escolher a melhor API obtiveram-se os seguintes resultados (ver tabela 8) utilizando uma ferramenta AHP¹⁵ disponível na internet. De todas as alternativas disponíveis a melhor foi a W3C Mozilla Web Speech com 50,5%, seguidos pela API da Google Cloud Speech com 24,2%, da CMU Sphinx com 12%, da API da Microsoft com 11,1% e por último a da IBM Watson com 2,1%.

Tabela 9: Tabela AHP para os resultados obtidos

	Curva de aprendizagem	Documentação	Funcionalidades	Idiomas suportados	
Mozilla Web Speech	0.514	0.514	0.435	0.517	0.505
CMU Sphinx	0.069	0.118	0.253	0.200	0.120
Google Cloud Speech to text	0.282	0.256	0.173	0.111	0.242
Microsoft speech to text	0.135	0.078	0.089	0.127	0.111
IBM Watson speech to text	0.000	0.033	0.051	0.046	0.021

¹⁵ Disponível em <https://malik-aliyev-94.github.io/AHP-Calculator/>

Capítulo 4 – Análise e Desenho

O capítulo de Análise e Desenho pretende especificar sumariamente a estrutura base da solução a desenvolver. Aqui são descritas a arquitetura, a análise de requisitos, diagramas de caso de uso, entre outros elementos incluindo a apresentação de uma maquete da ideia inicial do protótipo.

4.1 Análise de requisitos

4.1.1 Requisitos Funcionais

Em seguida são descritos os elementos acordados para o desenvolvimento do protótipo baseadas nas entrevistas realizadas junto a *stakeholders*.

Definiu-se que o protótipo teria as seguintes características. O primeiro ponto base acordado foi o de que o protótipo seria desenvolvido para o idioma português europeu e que iria utilizar uma API de reconhecimento de voz que permitisse capturar o que foi dito pelo utilizador e o transcrevesse em texto. Definiu-se igualmente que o sistema seria destinado a TF e respetivos pacientes, pois embora as pessoas com problemas de fala pudessem praticar de forma autónoma, seria melhor que o TF orientasse as pessoas. Sendo assim seria necessário um perfil de administrador destinado a TF e outro perfil o de cliente destinado a pacientes. O perfil dos pacientes seria o de utilizador com privilégios básicos na aplicação. No perfil cliente existiria a possibilidade de aceder a opções de avaliação da fala tanto para frases como para palavras, recorrendo à leitura do que é mostrado no ecrã. Se a leitura for correta o utilizador iria ter um feedback quanto à resposta dada. Se estiver certa o ecrã mostra as palavras ditas a verde e aumenta a pontuação do utilizador, caso as palavras ditas não correspondessem ao que foi pedido, então iria aparecer uma mensagem a vermelho e a pontuação não aumentaria.

Em seguida podem ser consultados os requisitos funcionais em detalhe de todo o protótipo.

Requisitos funcionais:

- Efetuar Login: Autenticação de utilizadores (Dois perfis de utilizador. Administrador e Cliente)
- Existência de dois perfis/grupos de utilizadores, um para TF outro para pacientes.

- Desenvolvido para o idioma português europeu.
- Capturar o áudio do microfone do dispositivo do utilizador cliente num dado momento, especificamente na opção análise e treino.
- Registo de dados de utilizador, Nome, Idade, Género, *Username* e *Password*.
- Menu perfil cliente com as seguintes opções: instruções, análise de frases, treinar palavras, consultar resultados obtidos, ver notas terapêuticas e sair da aplicação.
- Menu perfil administrador: definir frases e palavras de cada utilizador, criar notas terapêuticas, ver resultados de todos os utilizadores, banco de sons palavras e sair da aplicação.
- Disponibilizar opção de instruções do sistema ao utilizador paciente.
- Na opção análise de frases, mostrar frase para o utilizador paciente ler e dar feedback se certo ou errado e mostrar resultado acumulado.
- Na opção análise de palavras, mostrar palavra para o utilizador paciente ler e dar feedback se certo ou errado e mostrar resultado acumulado.
- Na opção resultados do paciente, mostrar gráfico de evolução e tabela de acumulado de resultados.
- Na opção de “Sair”, realizar *logout* da aplicação tanto para o perfil paciente como para o perfil administrador.
- Disponibilizar opção “Notas Terapêuticas”, onde o perfil administrador possa deixar indicações ao utilizador cliente.
- Configurar palavras e frases por utilizador com perfil administrador.
- Configurar e consultar banco de palavras para os sons
- Consultar todos os resultados dos utilizadores com perfil cliente (pacientes).
- Criar notas terapêuticas para os utilizadores com perfil cliente (pacientes)

4.1.2 Diagramas de Caso de Uso

Em seguida são apresentados diagramas de caso de Uso para o protótipo TeraFA para o caso dos dois perfis de utilizador, o perfil paciente (cliente) e o perfil do Terapeuta da Fala (Administrador). Os diagramas pretendem ajudar a entender melhor o funcionamento da aplicação. No caso de ambos os utilizadores, é necessário estarem previamente autenticados de maneira a poderem aceder ao sistema. No caso do utilizador paciente, depois de realizada a autenticação, ele poderá aceder às funcionalidades de “Analisar”, “Treinar”, consultar os seus

“Resultados”, ver “Notas” criadas pelo Terapeuta da Fala, e “Sair” da Aplicação. Por exemplo, quando o paciente está na aplicação e seleciona a opção “Analisar”, ele terá de dizer a frase que lhe aparece no ecrã corretamente, se o utilizador paciente disser corretamente a frase, a sua pontuação aumentará, caso contrário o utilizador não verá a sua pontuação aumentar. Já a opção “Resultados” mostra o progresso realizado pelo paciente, em forma de uma tabela e de um gráfico de progresso. As “Notas” funcionam como um espaço em que o paciente pode consultar as indicações dadas pelo TF.

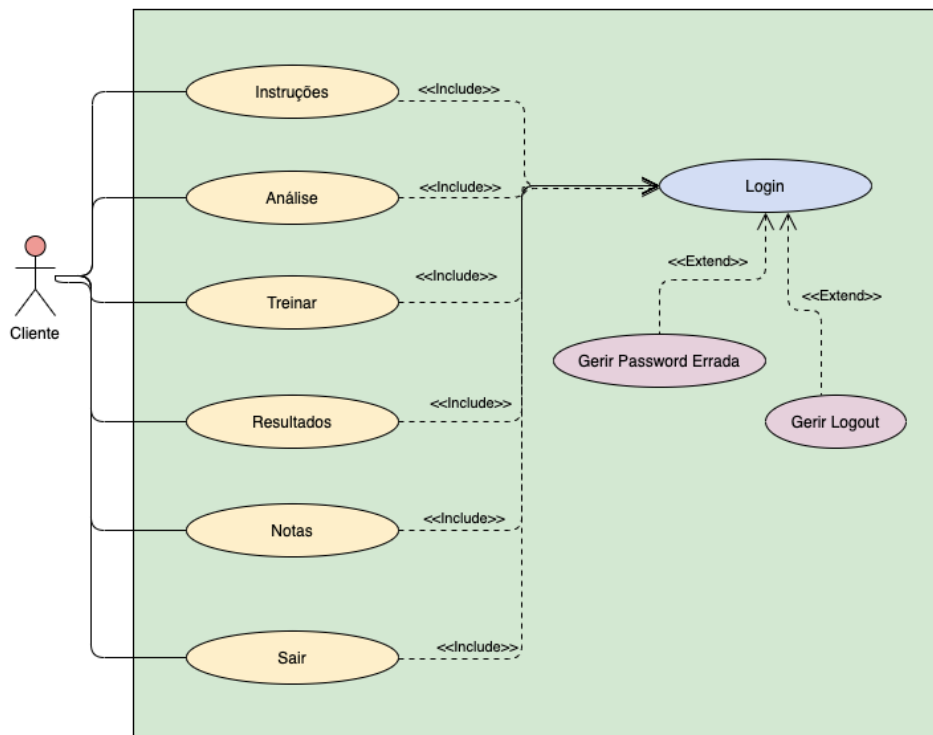


Figura 14 – Diagrama caso de Uso Paciente

O utilizador Terapeuta da Fala por sua vez tem a possibilidade de configurar o jogo, adicionando novas Frases/Palavras aos vários níveis do jogo. Basicamente o TF tem privilégios de administrador da aplicação. Este utilizador também terá a possibilidade de consultar o progresso realizado pelos pacientes na utilização da aplicação e configurar um banco de palavras que posteriormente pode adicionar às palavras praticadas pelos respetivos pacientes.

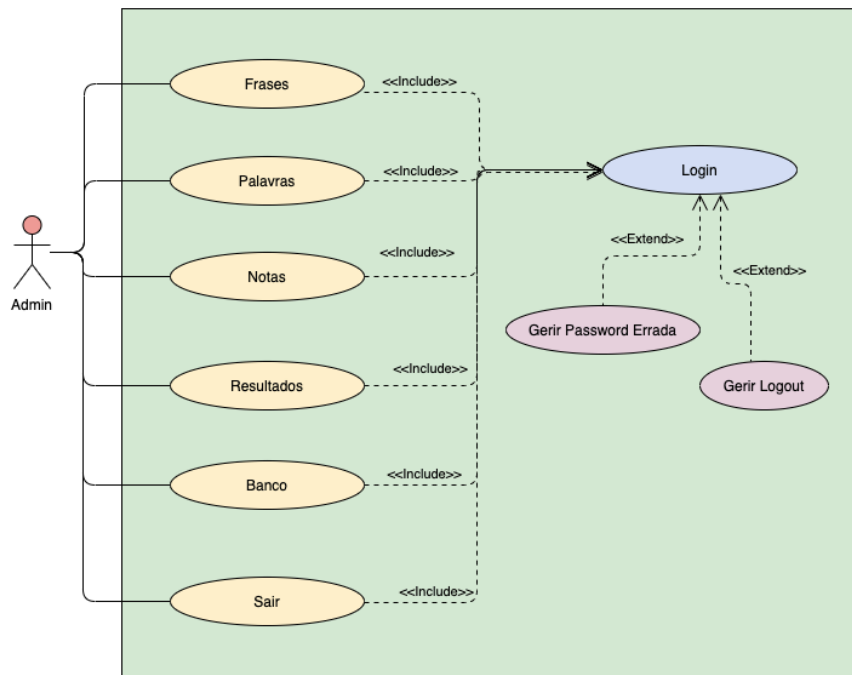


Figura 15 – Diagrama caso de Uso Terapeuta da Fala

4.1.3 Requisitos de Instalação

Para o desenvolvimento do protótipo, é necessário um servidor web para normal funcionamento do website. Em vez de se instalarem software individualmente foram considerados pacotes como por exemplo o WAMP (Windows Apache MySQL e PHP - Perl), XAMPP, ou MAMP (MacOS Apache MySQL e PHP-Perl). Estes pacotes são geralmente utilizados para instalar o Apache, MySQL, PHP, Perl. O XAMPP é um servidor multiplataforma, open source, que contém a base de dados MySQL, o servidor Web Apache e os interpretadores para as linguagens de programação PHP e Perl ao contrário do WAMP e do MAMP. A primeira letra X, deve-se ao facto de estar disponível para qualquer sistema operativo, Microsoft Windows, Linux e MacOS. A opção por um destes Web Servers varia dependendo do sistema operativo que se utilize.

Para o projeto decidiu-se optar pelo MAMP dado que o computador disponível para desenvolvimento possui o sistema operativo MacOS e assim não é necessário instalar cada uma das tecnologias por separada e seria utilizado PHP. Se necessário no futuro poder-se ia utilizar as demais opções.

4.2 A arquitetura

Neste ponto pretende-se descrever a estrutura base do projeto e identificar os componentes utilizados no sistema, como por exemplo a utilização da API de reconhecimento de voz.

A arquitetura da aplicação divide-se em três camadas, a camada cliente, a camada web e a camada de dados ver figura 15. Assim, o aplicativo da web é dividido numa camada de cliente, camada da web (ou aplicativo) e camada de dados. A camada de cliente consiste basicamente no navegador da web, vulgo *browser*, no caso deste projeto é utilizado o Chrome que se executa nos dispositivos do cliente (por exemplo, um computador desktop ou laptop) onde as páginas do aplicativo web são solicitadas, sendo em seguida exibidas no ecrã do cliente que as solicita. Já a camada web consiste no servidor da web Apache e no motor PHP que estão em execução do lado do servidor, bem como os ficheiros (por exemplo, HTML, imagens, JavaScript, CSS) e também ficheiros PHP que são igualmente armazenados no servidor. Por último temos a camada de dados que é basicamente onde se encontram as tabelas da base de dados nas quais os dados do sistema são armazenados.

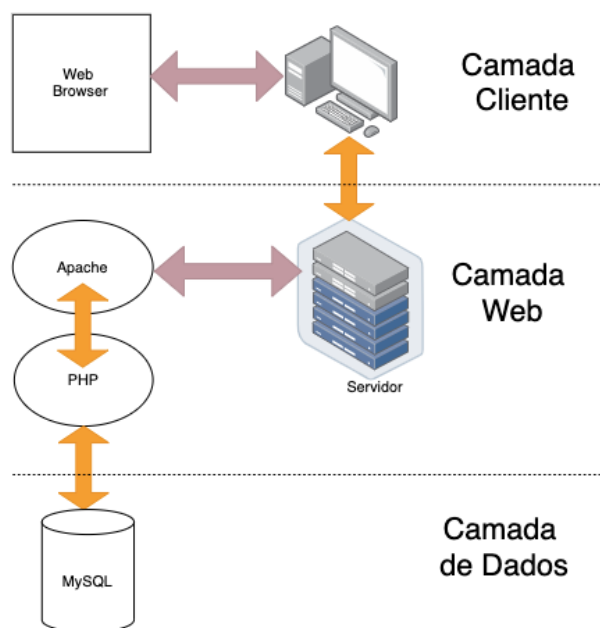


Figura 16 – Arquitetura da aplicação web a desenvolver

Um dos componentes a utilizar no projeto é o da API de reconhecimento de voz, para tal foram previamente analisadas diversas APIs. Mas para este projeto foi selecionada a API Web Speech da Mozilla desenvolvida conjuntamente com a Google e que é compatível com o navegador de internet Chrome em detrimento das demais dada a sua facilidade de utilização no projeto, pela existência de muita documentação e pela possibilidade de o protótipo ser utilizado quer em dispositivos móveis, quer em computadores desde que tenham a possibilidade de utilizar como navegador de internet o

Google Chrome¹⁶ e tenham um microfone para captar o áudio. Como referido anteriormente a Web Speech API possui quer a funcionalidade de síntese, quer a de reconhecimento de áudio, mas para o protótipo só será utilizado o serviço de reconhecimento de fala.

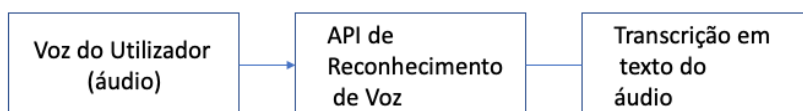


Figura 17 – Estrutura do sistema de reconhecimento de voz

Na figura 16 é possível visualizar o processo do sistema de reconhecimento de voz. Num primeiro momento o utilizador fala para um microfone, aí o áudio é captado e posteriormente processado pela API de reconhecimento de voz. O funcionamento da API é basicamente o de devolver a transcrição em texto do áudio de entrada recolhido do utilizador.

Já a figura 17 por sua vez disponibiliza uma primeira *maquete* da aplicação base. Do ponto de vista de elementos gráficos, ele possui um botão “Novo Teste” que quando pressionado pede autorização para utilizar o microfone do dispositivo e inicia a captação do áudio, transcrevendo-o depois em texto. Os outros elementos gráficos são campos de texto, uns estáticos (exemplo: o título “Acerta Palavras”) e outros dinâmicos como é o caso do campo de texto a verde com a palavra “Certo”, que indica se o utilizador disse a palavra pedida pela interface corretamente. Sucintamente o que o navegador Chrome faz com recurso à API é captar o áudio do utilizador e enviá-lo para os servidores de maneira a transcrever o áudio para texto.

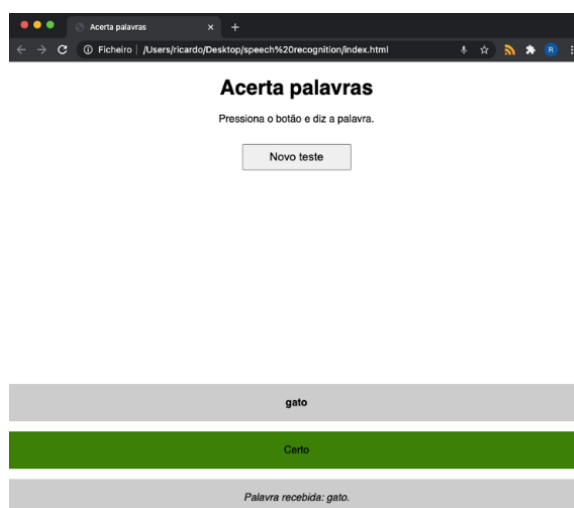


Figura 18 – Maquete inicial do protótipo

¹⁶ Disponível em <https://www.google.pt/intl/pt-PT/chrome>

4.2.1 Linguagens Utilizadas

Ao nível de linguagens de programação a utilizar do lado do servidor ela seria o PHP, pois é uma linguagem de script de código aberto (open source) muito utilizada no desenvolvimento web e que pode ser incorporada no HTML. O código PHP insere-se em instruções especiais de início e término de processamento com as instruções <? Php e ?> que permite sair e entrar do HTML. O código PHP é executado no lado servidor que depois cria o código HTML que se envia para o cliente.

O projeto utiliza SQL como linguagem de base de dados relacional de código aberto para gerir a base de dados. O serviço MySQL por sua vez utiliza o SQL (Structure Query Language), para inserção, acesso e gestão dos dados guardados na aplicação.

Do lado cliente as linguagens a utilizar seriam:

- CSS;
- Javascript;
- HTML.

O CSS significa Folha de Estilo em Cascatas em inglês (*Cascading Style Sheets*). Ele vai ser utilizado para definir a apresentação das páginas Web e utilizado conjuntamente com a linguagem HTML. No CSS vai permitir definir o layout, o tamanho das letras, tipo de letras, cores e outros elementos. Permitindo adaptar a apresentação a diferentes tipos de dispositivos. Mas neste caso o CSS vai ser utilizado para ser compatível com o browser Chrome de um computador. É importante ter em consideração que o CSS funciona à parte do HTML.

Já o Javascript é uma linguagem de programação que permite criar conteúdo que se atualizam dinamicamente e pode ser utilizado tanto no lado cliente como no lado servidor. Para este projeto ele será utilizado para o lado cliente.

Por sua vez o HTML significa Hypertext Markup Language e não é uma linguagem de programação pois não permite criar funcionalidades dinâmicas. Ele vai ser utilizado principalmente para organizar e estruturar seções, parágrafos, cabeçalhos e links para páginas do protótipo.

4.3 Gamificação

Para o desenvolvimento final pensou-se em utilizar os princípios utilizados nos jogos, com o objetivo de melhorar a experiência de utilizador e aumentar o interesse na aplicação,

funcionando assim como um mecanismo enriquecedor da aplicação e de certa forma complementar o projeto. Em seguida são apresentados alguns dos princípios utilizados na mecânica de jogos para assim poder entender em que consiste e como podem ser aplicados.

Segundo Robson a gamificação tem aplicações em contextos como o da saúde, sustentabilidade, administração pública, transporte, educação entre outras. O objetivo passa principalmente por fazer com que os utilizadores se sintam mais envolvidos. (Robson et al., 2015)

A nível histórico com o crescimento ocorrido na indústria dos videojogos, os criadores de jogos e investigadores tiveram a possibilidade de realizar estudos para melhor compreender o que faz do videojogo mais envolvente e de sucesso. Isso fez com que se chegasse a um número de teorias e lições sobre o desenho e gestão das experiências de jogos, e de frameworks que incentivam e motivam os indivíduos a jogar. Para o desenvolvimento do protótipo idealizou-se que o utilizador através da leitura acertada das palavras no ecrã em sequência correta tivesse a possibilidade de passar níveis, por exemplo subindo de nível de dificuldade.

Três princípios importantes de gamificação são os de **Mechanics**, **Dynamics** e **Aesthetics/Emotions**, segundo (Hunicke et al., 2004). **Mechanics** descrevem os componentes de um jogo, ao nível da representação de dados e dos algoritmos. **Dynamics** descreve o comportamento em tempo de execução de **Mechanics** baseado nos *inputs* do jogador. **Aesthetics/Emotions** descreve as respostas emocionais desejadas evocadas pelo jogador, quando interage com o sistema de jogo.

O objetivo de criar a aplicação com princípios de jogo, passa primeiramente por manter o utilizador motivado, para que sinta que a aplicação é divertida e assim possa continuar a utilizá-la, sentindo por exemplo que o tempo despendido passa rápido. Para isso é importante que o utilizador se sinta autónomo na utilização da aplicação e sinta que está a construir competências, ou seja evoluir. Um caso prático pode ser quando o utilizador está a praticar a dicção de palavras para melhorar a pronúncia dos sons das mesmas. Se o utilizador sentir que está a melhorar, ele vai querer continuar a praticar mais. Outro dos métodos pensados para motivar o utilizador passa possivelmente por recompensá-lo com prémios/badges.

Capítulo 5 – Implementação

Este capítulo pretende descrever o processo de desenvolvimento do protótipo bem como as ferramentas utilizadas.

Para o desenvolvimento foram utilizados principalmente um editor de texto Visual Studio Code, a ferramenta phpMYAdmin para gerir a base de dados e o navegador Google Chrome para execução do desenvolvimento e testes.

Como referido no capítulo anterior as linguagens utilizadas foram PHP, Javascript, *HTML*, *CSS*, *recorrendo-se também a AJAX, jQuery e Bootstrap*. Quanto à API de reconhecimento de voz a utilizar ela seria a Web Speech API. A fase de implementação pretende ir de encontro aos requisitos funcionais estabelecidos no capítulo anterior, análise e desenho.

5.1 Base de Dados

No desenvolvimento do protótipo optou-se por utilizar uma base de dados SQL para guardar informação relevante ao projeto, como por exemplo as frases e palavras definidas pelo TF a serem lidas pelos pacientes, os resultados obtidos no uso do projeto pelos utilizadores, um banco de palavras para consulta dos TF, as notas terapêuticas dos TF para os pacientes e uma tabela para registo de dados dos utilizadores. Para administração da base de dados é utilizada também a ferramenta open source phpMyAdmin. Ela foi desenvolvida em PHP para administração do MySQL pela Internet, para ser possível criar a base de dados, e administrar tabelas.

À base de dados criada deu-se o nome de demo, correspondente a demonstração, e foram criadas as seguintes tabelas, **bank** para funcionar como banco de palavras que guarda palavras e sons que utilizam os TF; **notes**, para guardar notas terapêuticas dadas aos utilizadores pelos TF; **scores**, tabela para guardar os resultados dos pacientes nas opções de Análise e Treino do protótipo, tabela **sentences** para guardar frases por utilizador, **words** para guardar palavras por utilizador; e **users** para guardar dados dos utilizadores.

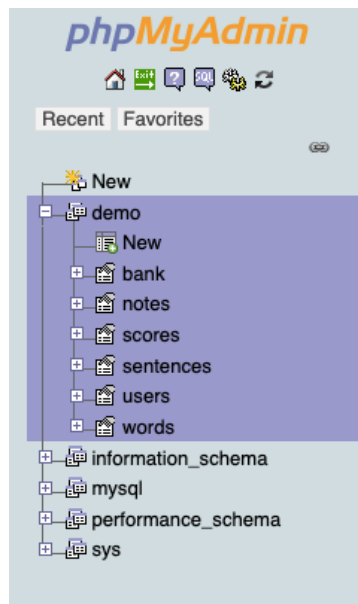


Figura 19 – Ecrã phpMyAdmin para base de dados demo

Showing rows 0 - 24 (25 total, Query took 0.0054 seconds.)

```
select * from information_schema.columns where table_schema = 'demo' order by table_name,ordinal_position
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Show all | Number of rows: 25 | Filter rows: Search this table

Options

TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH
def	demo	bank	id	1	NULL	NO	int	
def	demo	bank	letter	2	NULL	NO	varchar	
def	demo	bank	sound	3	NULL	NO	varchar	
def	demo	bank	position	4	NULL	NO	enum	
def	demo	bank	word	5	NULL	NO	varchar	
def	demo	notes	id	1	NULL	NO	int	
def	demo	notes	username	2	NULL	NO	varchar	
def	demo	notes	note	3	NULL	NO	varchar	
def	demo	scores	Username	1	NULL	NO	varchar	
def	demo	scores	Score	2	NULL	NO	int	
def	demo	scores	Date	3	NULL	YES	datetime	
def	demo	sentences	id	1	NULL	NO	int	
def	demo	sentences	username	2	NULL	NO	varchar	
def	demo	sentences	sentence	3	NULL	NO	varchar	
def	demo	users	id	1	NULL	NO	int	
def	demo	users	username	2	NULL	NO	varchar	
def	demo	users	password	3	NULL	NO	varchar	
def	demo	users	name	4	NULL	NO	varchar	
def	demo	users	age	5	NULL	NO	int	
def	demo	users	gender	6	NULL	NO	varchar	
def	demo	users	created_at	7	CURRENT_TIMESTAMP	YES	datetime	
def	demo	users	role	8	NULL	NO	enum	
def	demo	words	id	1	NULL	NO	int	
def	demo	words	username	2	NULL	NO	varchar	
def	demo	words	word	3	NULL	NO	varchar	

Figura 20 – Tabelas e campos da base de dados

5.2 Código CSS e HTML

Em relação ao código HTML e CSS ele é relativamente simples em todo o protótipo. Por exemplo existe um ficheiro style.css onde se pode encontrar todo o código das folhas de estilo. Nele podemos encontrar parte da formatação aplicada ao protótipo, ver figura 18.

```

1  body, html {
2  |   margin: 0;
3  |   }
4  |
5  |   html {
6  |     height: 100%;
7  |     background-color: #white;
8  |   }
9  |
10 |   body {
11 |     height: inherit;
12 |     overflow: hidden;
13 |   }
14 |
15 |   h1, p {
16 |     font-family: sans-serif;
17 |     text-align: center;
18 |   }
19 |
20 |   div p {
21 |     padding: 20px;
22 |     background-color: rgba(0,0,0,0.2);
23 |   }
24 |
25 |   div {
26 |     overflow: auto;
27 |     position: absolute;
28 |     bottom: 0;
29 |     right: 0;
30 |     left: 0;
31 |   }
32 |
33 |   button {
34 |     margin: 0 auto;
35 |     display: block;
36 |     font-size: 1.1rem;
37 |     width: 170px;
38 |     line-height: 2;
39 |     margin-top: 30px;
40 |   }

```

Figura 21 – Parte do código CSS do protótipo

No código HTML utilizam-se um title, parágrafos - p, div, botões entre outros elementos.

5.3 Código de reconhecimento de fala

O código javascript utilizado consiste maioritariamente na definição de variáveis e funções. Existem variáveis para por exemplo guardar as frases ou palavras que vão ser utilizadas pelo sistema de reconhecimento de voz, ou uma variável para definir o idioma do reconhecimento de fala. Na figura 21 pode ser visualizado o procedimento que é utilizado para quando ocorre o fim do reconhecimento de voz. Aqui inicializam-se o texto do button testBtn a Iniciar e coloca-se o estado a enabled, ativo. Define-se também que pare o reconhecimento de fala e a gravação do áudio.

```

recognition.onspeechend = function() {
    stop(preview.srcObject);
    recognition.stop();
    testBtn.disabled = false;
    testBtn.textContent = 'Iniciar';
}

```

Figura 22 – Função Javascript utilizada para finalizar reconhecimento de fala

```

327 recognition.onresult = function(event) {
328     var speechResult = event.results[0][0].transcript.toLowerCase();
329     if(speechResult === phrase) {
330         resultPara.textContent = 'Recebido: ' + speechResult;
331         resultPara.style.background = 'green';
332         score ++;
333         diagnosticPara.textContent = 'Ponto(s): ' + score;
334         stop(preview.srcObject);//adicionada
335     } else {
336         resultPara.textContent = 'Recebido: ' + speechResult;
337         resultPara.style.background = 'red';
338         diagnosticPara.textContent = 'Ponto(s): ' + score;
339         stop(preview.srcObject);//adicionada
340     }
341 }

```

Figura 23 – Função javascript onresult

A propriedade onresult do objeto recognition (SpeechRecognitionEvent) retorna um objeto do tipo SpeechRecognitionResultList.

O objeto speechResult por sua vez tem um getter para que possa ser aceder como um array, sendo que o primeiro [0] retorna SpeechResult na posição 0 e o segundo [0] retorna o SpeechRecognitionAlternative na posição 0. Já o bloco condicional if else compara se o que a API captou corresponde à palavra da variável phrases, ver figura 23.

5.4 Código PHP

Utiliza-se o PHP de maneira a ler por exemplo as frases existentes na base de dados. Inicializa-se uma variável, neste caso *name* para guardar as frases que estão na base de dados. Para isso é estabelecida uma conexão com a base de dados *demo* e realizada uma consulta com a query "SELECT sentence FROM sentences WHERE username = ".\$q.""; onde a variável *q* corresponde ao username do utilizador que se encontra conectado. Se a conexão à base de dados foi

realizada com sucesso então, a variável do tipo array \$name irá conter as frases que se irão mostrar na opção “Analisar” para este utilizador.

```
<?php
$name = "";
$q = htmlspecialchars($_SESSION["username"]);
$con = mysqli_connect('localhost','root','root','demo');
if (!$con) {
    die('Could not connect: ' . mysqli_error($con));
}
mysqli_select_db($con,"sentences");
$sql="SELECT sentence FROM sentences WHERE username = '".$q."'";
$result = mysqli_query($con,$sql);
$name = array();

if ($result) {
    while ($row = mysqli_fetch_array($result)) {
        $name[] = $row['sentence'];
    }
}
mysqli_free_result($result);

mysqli_close($con);
?>
```

Figura 24 – Código PHP de conexão à Base de dados e consulta em variável

```
198     var phrases = <?php echo json_encode($name); ?>;
```

Figura 25 – Variável Javascript para guardar array em PHP

Na figura 25 linha 198, pode ver-se a devolução em uma string que contém a representação JSON do valor \$name. JSON significa Javascript Object Notation e é um formato utilizado para troca de dados simples e rápido entre sistemas.

```

114     <h2>Banco de Palavras</h2>
115
116     <table border="2" width=80%>
117         <tr>
118             <!--<td>Id</td>-->
119             <td>Fonema</td>
120             <td>Som</td>
121             <td>Posição</td>
122             <td>Palavra</td>
123             <td>Editar</td>
124             <td>Apagar</td>
125         </tr>
126
127         <?php
128             include "config.php"; // Using database connection file here
129
130             $records = mysqli_query($link,"select * from bank order by sound asc"); // fetch data
131
132             while($data = mysqli_fetch_array($records))
133             {
134
135
136                 <tr>
137
138                     <!--<td><?php //echo $data['id']; ?></td>-->
139                     <td><?php echo $data['letter']; ?></td>
140                     <td><?php echo $data['sound']; ?></td>
141                     <td><?php echo $data['position']; ?></td>
142                     <td><?php echo $data['word']; ?></td>
143                     <td><a href="edit_bank.php?id=<?php echo $data['id']; ?>">Editar</a></td>
144                     <td><a href="delete_bank.php?id=<?php echo $data['id']; ?>">Apagar</a></td>
145                 </tr>
146             <?php
147             }
148         <?>
149     </table>

```

Figura 26 – Código PHP para gerar tabela HTML

Na figura 26 pode-se ver o código utilizado para criar uma tabela em HTML para o banco de palavras, nele recorre-se ao PHP para gerar os dados de uma forma dinâmica recorrendo a uma conexão à base de dados, linha 128 e 132. E também se faz chamadas a outros ficheiros PHP, para editar a entrada do banco de palavras ou para eliminar a entrada, linhas 143 e 144 ou conectar através de outro ficheiro PHP, linha 128. Pode-se ver também a execução da consulta à base de dados, linha 138.

5.5 Front End

No Front End encontram-se explicadas as interfaces que podem ser vistas quando acedidas pelos dois tipos de utilizadores existentes na aplicação, o utilizador paciente e o utilizador Terapeuta da Fala/Administrador. Aqui encontra-se o design de tudo que os utilizadores podem ver no protótipo.

5.5.1 Perfil Paciente

Neste ponto podem ser vistas as interfaces resultantes do desenvolvimento do protótipo. Na figura 27 encontra-se o menu para os utilizadores pacientes. O paciente terá a opção de escolher entre Instruções, Analisar, Treinar, Resultados, Notas e Sair. Na opção “Instruções”, o utilizador terá a possibilidade de consultar as indicações relativas à utilização do protótipo e na opção “Sair” realizar o logout, saindo assim da aplicação. As demais opções serão explicadas mais à frente.

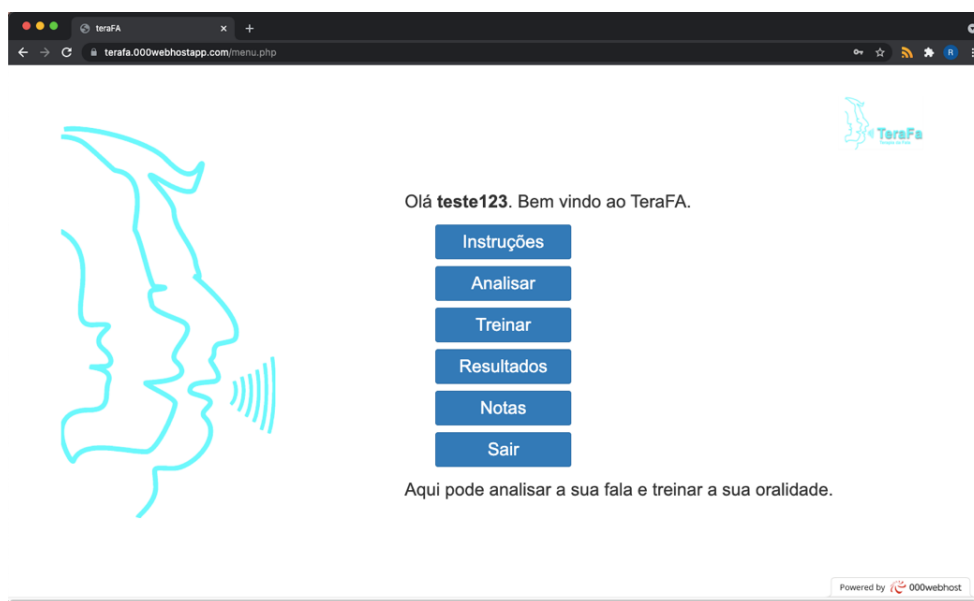


Figura 27 – Ecrã do menu principal utilizador cliente

Na figura 28 pode ser visto o ecrã da opção “Analisar”, aqui o utilizador tem a possibilidade de analisar a fala, recorrendo à leitura de uma frase quando pressionado o botão Iniciar.

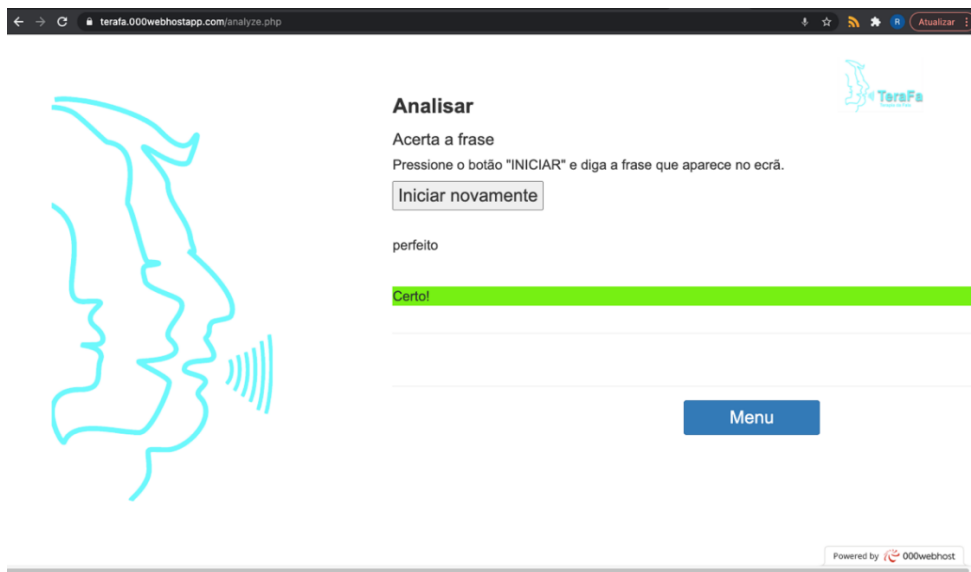


Figura 28 – Ecrã da opção “Analisar” do utilizador paciente

Já a opção “Treinar”, figura 29 não difere muito da opção Analisar. A interface é praticamente a mesma, diferenciando apenas nos elementos de texto, por exemplo no que é mostrado a ler pelo utilizador, em vez de aparecer uma frase, aparece apenas uma palavra de cada vez que é pressionado o botão “INICIAR”. Se o botão for pressionado ativa-se o reconhecimento de fala e inicia-se a gravação do áudio através do micro. Na parte de baixo existe um *media player* que permite ouvir o que foi dito pelo utilizador. Existem também mais dois botões que contêm a possibilidade de guardar o resultado “Guardar resultado” e de voltar ao menu principal “Menu”.

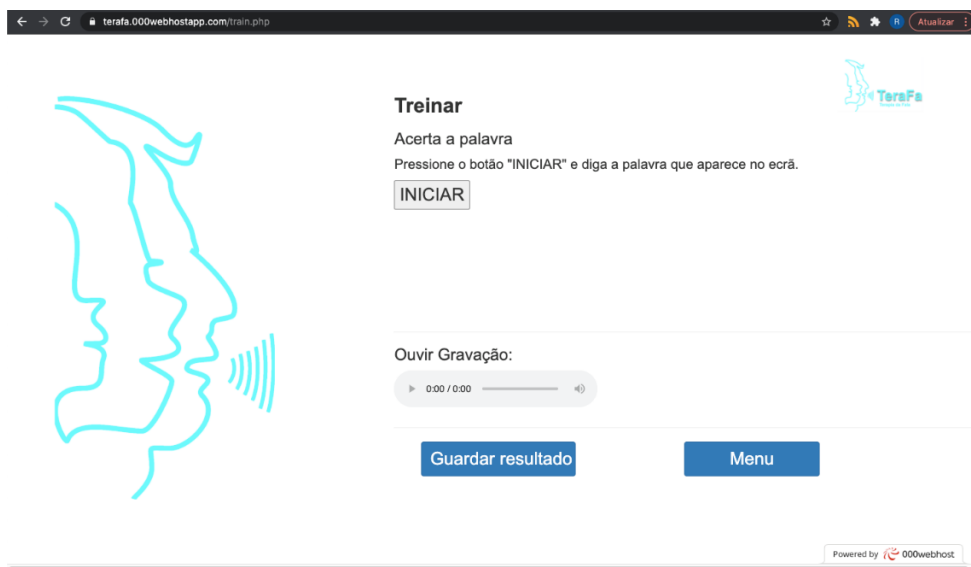


Figura 29 – Ecrã da opção “Treinar” do utilizador paciente



Figura 30 – Ecrã da opção “Resultados” do utilizador paciente

A opção “Resultados”, ver figura 30, por sua vez e como o nome indica, serve para mostrar os resultados obtidos pelo utilizador sob a forma de uma tabela com os pontos obtidos e com a data em que praticou a fala, bem como com um gráfico.

Por último temos a opção “Notas” onde o paciente pode verificar as indicações dadas pelo TF, ver figura 30 como exemplo.

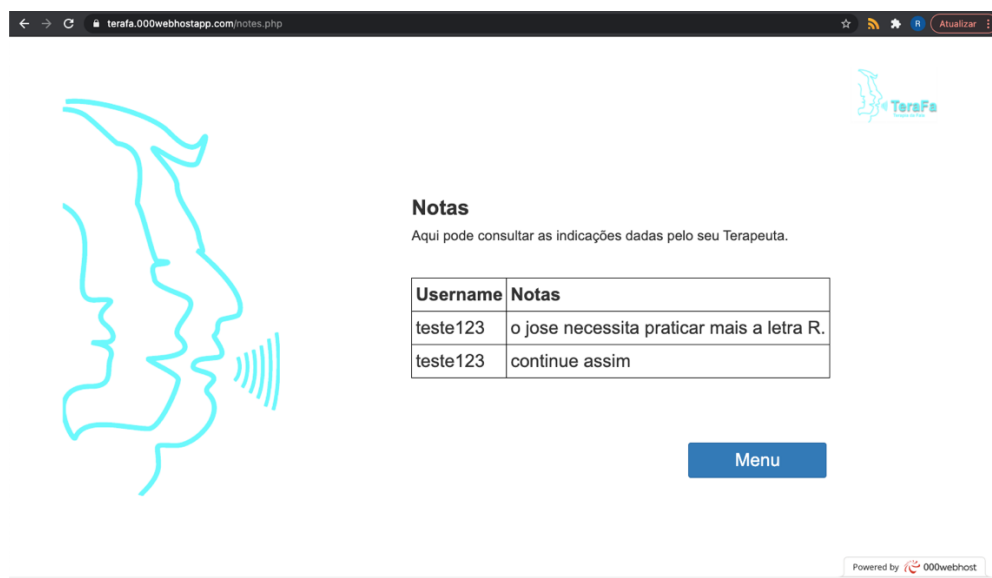


Figura 31 – Ecrã da opção “Notas” do utilizador paciente

5.5.2 Perfil Terapeuta

O ponto 5.5.2 pretende descrever as interfaces que estão disponíveis para o tipo de utilizador Terapeuta, administrador. Elas são basicamente opções de configuração para serem usadas pelos pacientes. Na figura 32 encontra-se o menu do utilizador Terapeuta, ele contém as opções “Frases”, “Palavras”, “Notas”, “Resultados”, “Banco” – banco de sons de palavras e a opção de realizar logout do sistema. Todas as opções mostradas tiveram como base o conjunto de requisitos elaborados na fase de Análise do ponto 4.

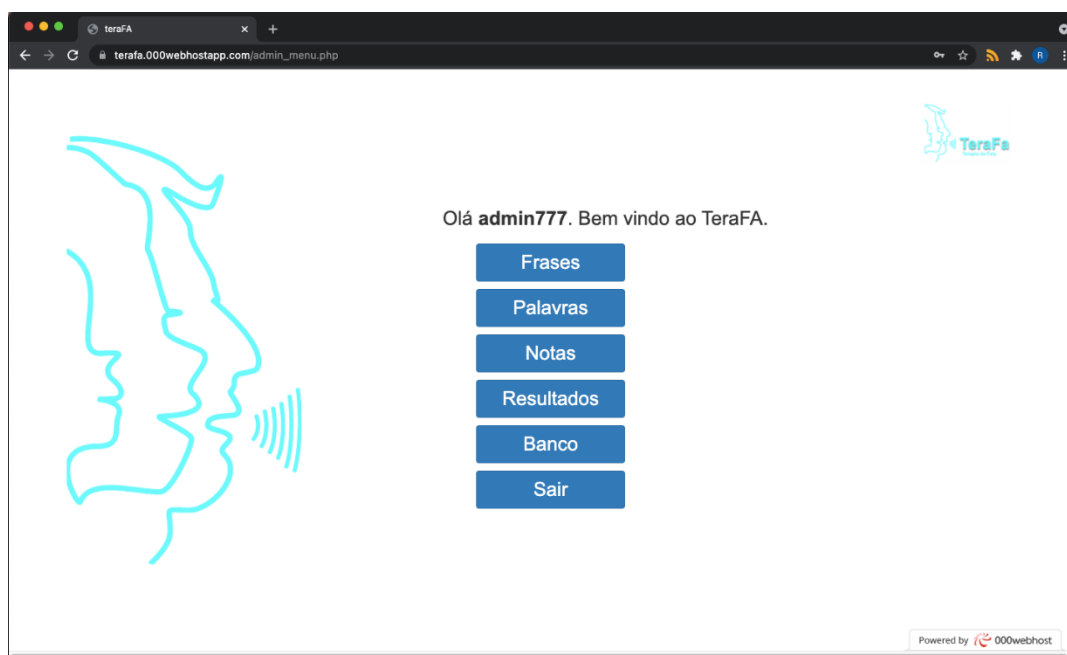


Figura 32 – Ecrã do menu principal utilizador terapeuta

Na figura 33 podem ser vistos as configurações das frases a serem lidas pelos pacientes. Existe uma tabela onde se pode ver o utilizador, a frase, nele também se pode editar a frase ou apagar. Mais abaixo existe a opção de adicionar uma nova entrada, para isso apenas é necessário introduzir o utilizador e a frase desejada. É importante mencionar que existe uma interface idêntica, mas apenas para palavras, ver figura 34.

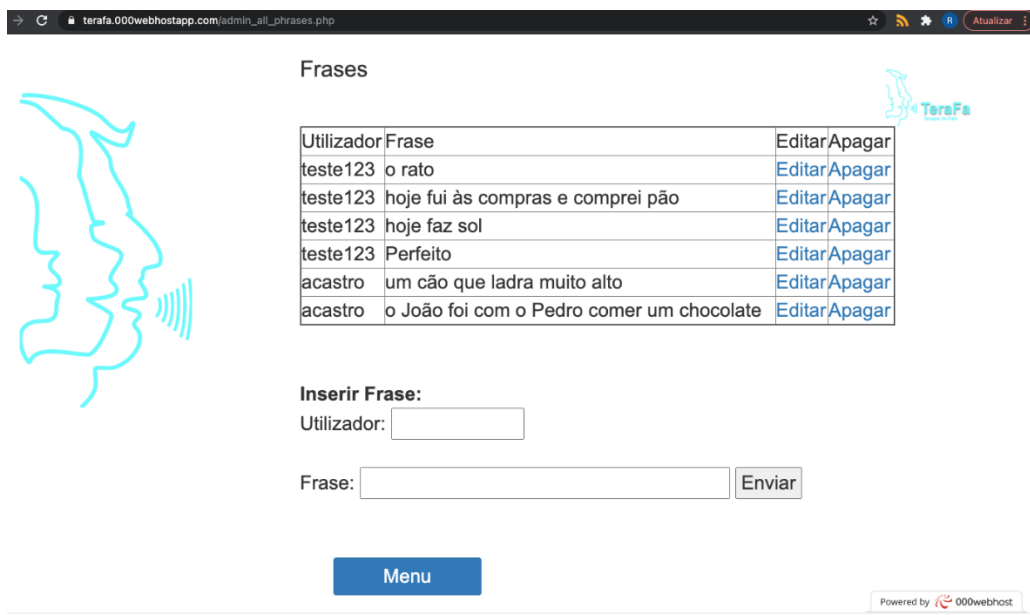


Figura 33 – Ecrã da opção “Frases” do utilizador Terapeuta

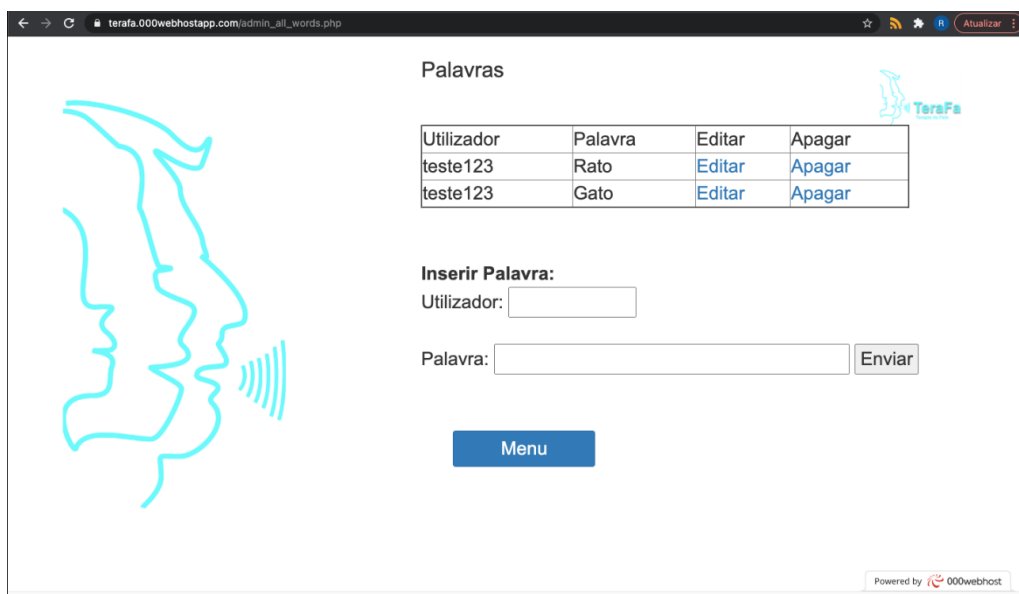


Figura 34 – Ecrã da opção “Palavras” do utilizador Terapeuta

Outra das opções disponíveis é a de “Notas”, aqui é o local onde os Terapeutas definem indicações a serem dadas aos pacientes. Por exemplo na figura 35 é possível ver que existem indicações para dois utilizadores. E que na parte inferior do ecrã existe a possibilidade de adicionar novas notas.

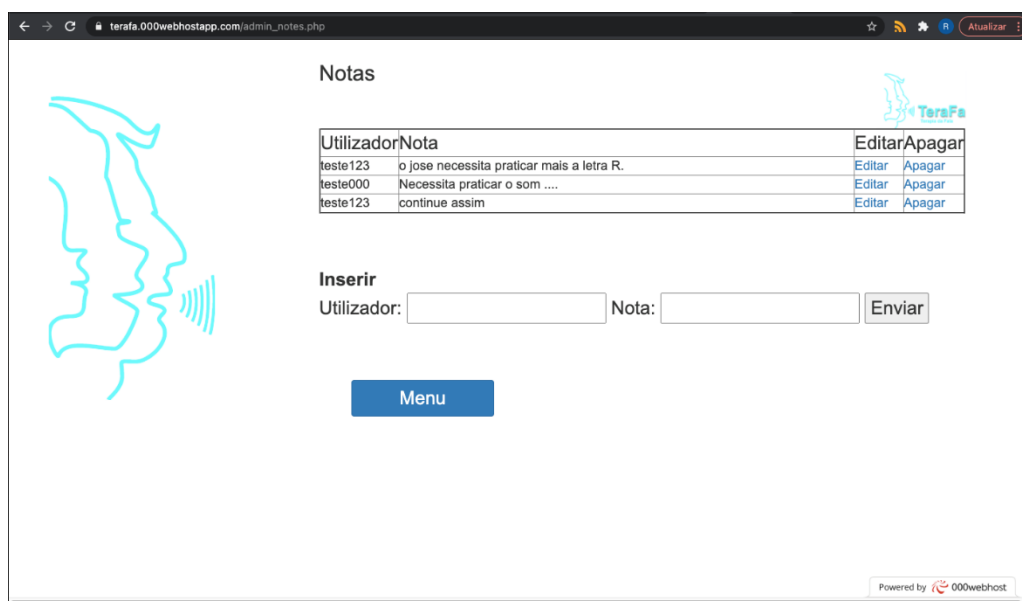


Figura 35 – Ecrã da opção “Notas” do utilizador Terapeuta

Na opção “Resultados”, o Terapeuta pode ver todos os dados dos utilizadores registados e os resultados obtidos de cada um que guardou a prática do reconhecimento de voz praticados na opção “Treinar” do perfil paciente, ver figura 36.

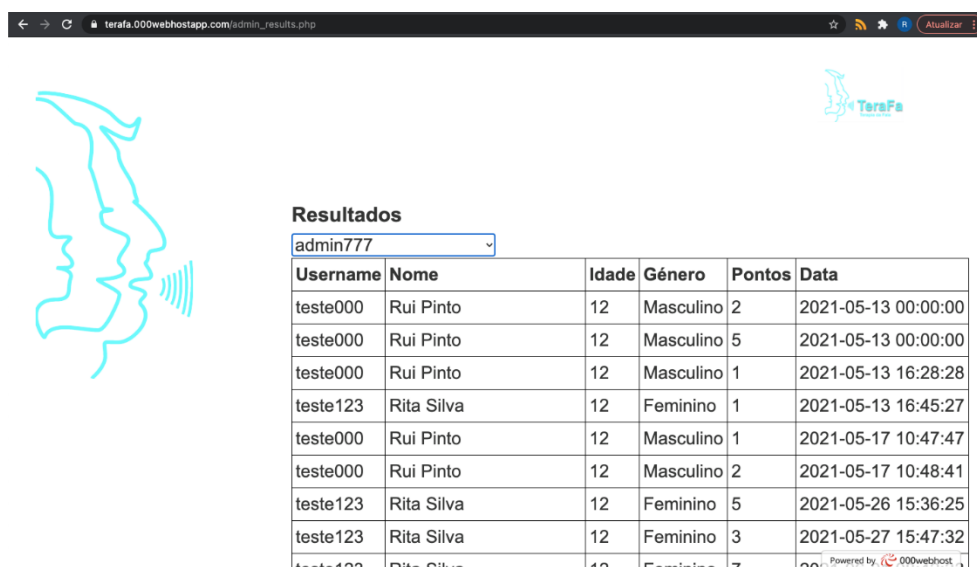


Figura 36 – Ecrã da opção “Resultados” do utilizador Terapeuta

Outra das opções que o Terapeuta tem acesso é à do banco de sons de palavras que este pode configurar e consultar de maneira a dar para os pacientes praticarem, ver figura 37.

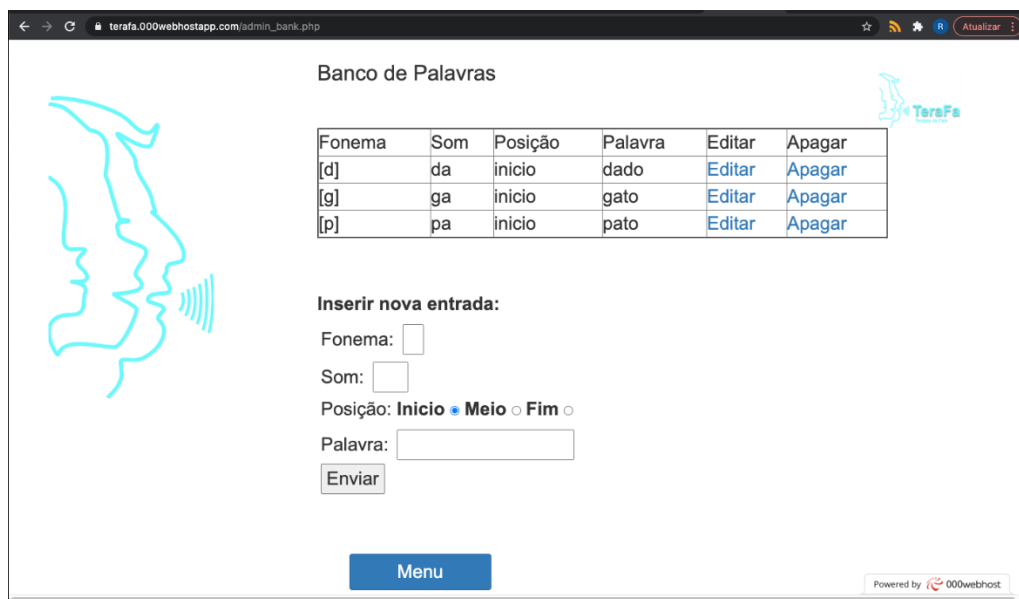


Figura 37 – Ecrã da opção “Banco” de palavras do utilizador Terapeuta

5.6 Planeamento dos testes

Durante o desenvolvimento do protótipo foram sendo realizados testes que consistiram em verificar o comportamento do protótipo e validar que o protótipo estivesse de acordo com os requisitos especificados. A fase planeamento de testes foi sendo realizada conjuntamente com os orientadores deste trabalho, através de reuniões para validar a aplicação via Skype. Sendo que os testes realizados permitiram melhorar a aplicação de uma forma considerável e corrigir bugs, permitindo assim ir fazendo alterações de forma rápida. Foram testados todos os formulários, links e botões para ver se existia alguma anomalia.

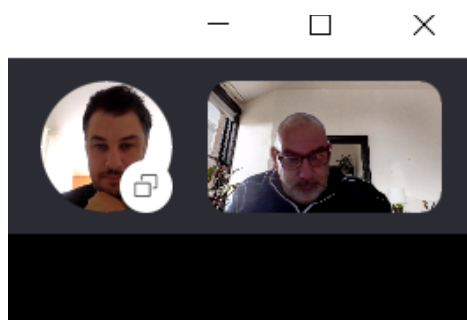


Figura 38 – Ecrã de reunião para validar protótipo realizada via Skype

Capítulo 6 – Avaliação

Neste capítulo são dados a conhecer as opiniões e as dificuldades encontradas pelos utilizadores durante a utilização do protótipo com o objetivo de avaliar a aplicação quanto às suas funcionalidades e à sua usabilidade. Para tal os utilizadores responderam a um questionário que tiveram de preencher depois de terem utilizado a aplicação. Durante o preenchimento do questionário, o utilizador da aplicação expressou a sua concordância ou discordância a um conjunto de afirmações de maneira a avaliar o protótipo.

Uma parte importante do processo de desenvolvimento de software é o da avaliação, de maneira a assegurar que a qualidade do produto final seja ótima.

A avaliação do protótipo pretendeu verificar que os requisitos funcionais definidos tinham sido desenvolvidos e estariam a funcionar corretamente. É importante mencionar que os utilizadores avaliados foram principalmente pessoas não afetadas à área da Terapia da Fala.

Para avaliar os utilizadores tiveram acesso ao protótipo e seguiram passo a passo o guia de testes de maneira a poderem responder ao questionário final.

O questionário por sua vez consistia em avaliar as apreciações dadas pelos utilizadores quanto a:

- Identificar os utilizadores relativamente à sua faixa etária, género e nível de escolaridade;
- Saber se os utilizadores já tinham tido contato com a área da Terapia da Fala;
- O que acharam do sistema de reconhecimento de voz;
- O que acharam da navegabilidade do site;
- Avaliar a facilidade de aprendizagem de utilização do protótipo;
- Avaliar a segurança da aplicação;
- Se pensa que o projeto pode ajudar pessoas com dificuldades na fala;
- Se demorou muito tempo na utilização da aplicação;
- Se foi capaz de completar as tarefas corretamente;
- O que achou do design da aplicação;
- Que melhorias realizaria na aplicação;
- E se pensa que a utilização do computador é uma mais-valia.

Foram feitas várias avaliações ao protótipo para aferir o seu potencial e funcionamento. Na figura seguinte apresenta-se um desses momentos.



Figura 39 – Um dos momentos de avaliação do TeraFA

A todos participantes foi dado acesso não só à aplicação, como distribuído um formulário de teste do (Anexo 1) e ao formulário de avaliação (Anexo 2).

Foram apuradas um total de 32 respostas em que 20 eram do género masculino e 12 do género feminino.

GÉNERO DOS INQUIRIDOS

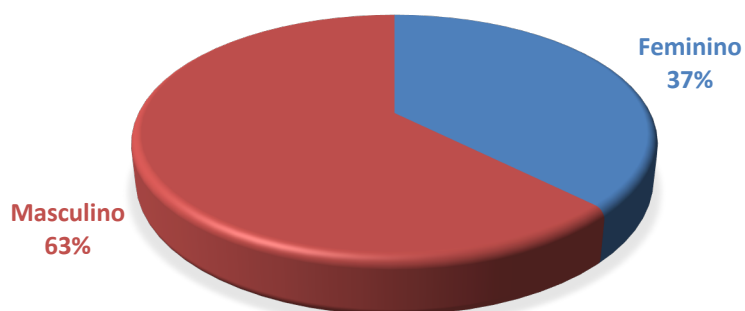


Figura 40 – Gráfico da amostra de sexo dos inquiridos

As idades dos utilizadores estiveram compreendidas entre os 20 e os 60 anos e tiveram a seguinte distribuição. Sendo a principal faixa etária a dos 20 aos 30 anos com 25 respostas a que avaliou mais o protótipo.

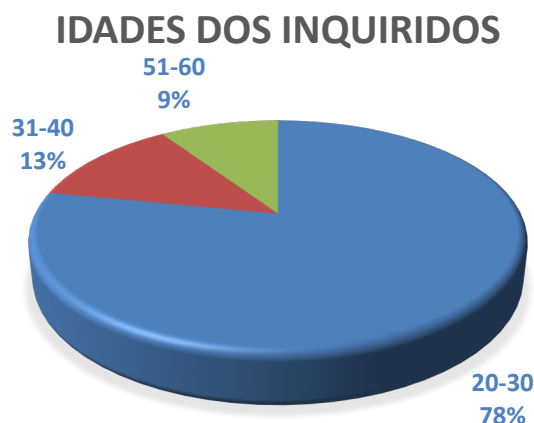


Figura 41 – Gráfico da amostra de faixa etária dos inquiridos

Em relação ao nível de escolaridade todos os utilizadores tinham pelo menos o 12º ano e até ao doutoramento.

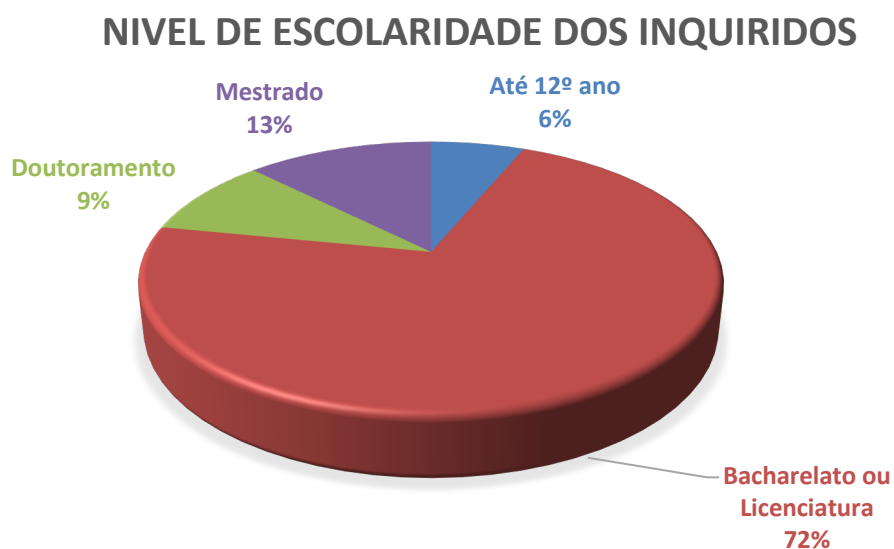


Figura 42 – Gráfico do nível de escolaridade dos inquiridos

Os utilizadores que responderam têm experiência em áreas como a Terapia da Fala, a informática ou o ensino. Todos os utilizadores conheciam a Terapia da Fala e tiveram

contato direto ou indireto com a mesma. A maioria dos utilizadores classificou o sistema de reconhecimento de fala positivamente com Bom, Muito bom ou Excelente.

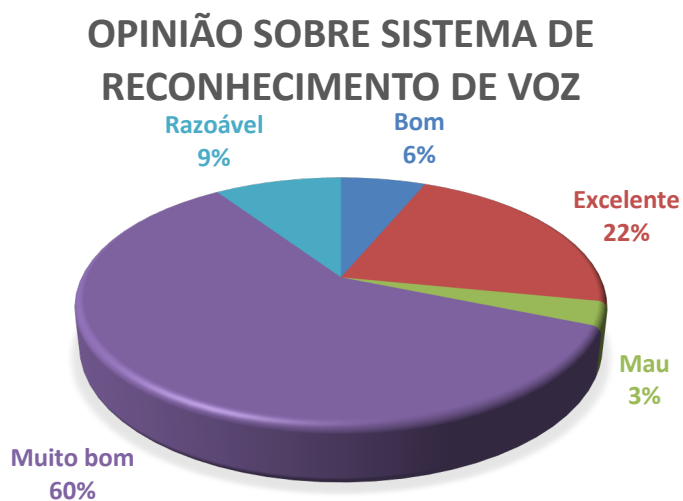


Figura 43 – Gráfico sobre avaliação do sistema de reconhecimento de fala

97% das respostas sobre a facilidade de navegar pela aplicação e se aprenderam facilmente as funcionalidades da aplicação foram positivas.



Figura 44 – Gráfico sobre facilidade de navegar pela aplicação e aprendizagem das funcionalidades

94% dos utilizadores sente que a aplicação pode ajudar a melhorar a fala, num total de 30 respostas afirmativas.

SENTE QUE A APLICAÇÃO PODE AJUDAR PESSOAS A MELHORAR A FALA

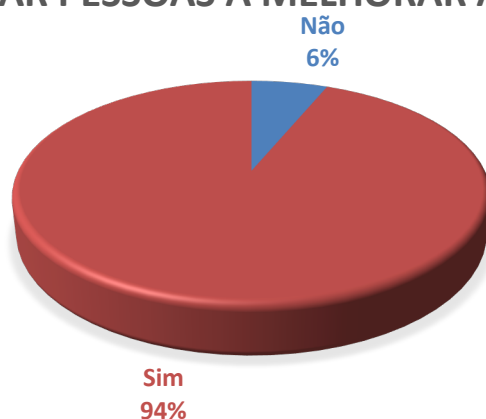


Figura 45 – Gráfico de opinião se a aplicação pode ajudar as pessoas com problemas de fala

Os inquiridos de uma maneira geral sentiram-se confortáveis a utilizar a aplicação e acham que a aplicação pode ser útil para ajudar as pessoas com dificuldades na fala a melhorarem a sua situação. 94% dos utilizadores acharam o design da aplicação lógico e recomendariam a aplicação.

Contudo durante a avaliação os utilizadores mencionaram que o protótipo continha ainda alguns erros, por exemplo no ecrã de resultados. Se não existirem resultados guardados aparece uma mensagem de erro do sistema, o que não deveria acontecer. Outro ponto que foi detetado na avaliação consistiu no problema de poder guardar resultados com o valor de 0. Alguns utilizadores mencionaram que seria bom se a aplicação estivesse disponível para outros navegadores para além do Google Chrome e que estivesse também disponível para dispositivos móveis. Alguns mencionaram que se poderia melhorar o design da aplicação. Ainda alguns utilizadores referiram que o melhor seria desenvolver mais as instruções para descrever mais detalhadamente o funcionamento da aplicação.

Capítulo 7 – Conclusões e trabalho futuro

Este capítulo pretende descrever as conclusões sobre a escrita deste trabalho. Aqui é referido se se conseguiu atingir os objetivos propostos, bem como indicar as limitações encontradas. Termina com a escrita sobre o possível trabalho futuro a realizar.

7.1 Conclusões e Critica de Resultados

Com este trabalho acima de tudo pretendeu-se desenvolver uma solução para auxiliar os TF com mais uma ferramenta que que ajudasse os seus pacientes a melhorarem a fala. Para tal foi realizado o levantamento do estado da arte e realizada uma análise de valor que possibilitou verificar a viabilidade do projeto e selecionar a API de reconhecimento de fala que melhor servia o propósito deste projeto. Numa primeira fase foram realizadas avaliações com 12 utilizadores mais afetos à área da Terapia da Fala. Posteriormente foram realizadas as restantes avaliações junto a alunos da disciplina de INDES do mestrado em Engenharia Informática do ISEP. Em relação à utilização da aplicação, o objetivo passou por motivar os utilizadores a utilizarem a aplicação com tecnologia de reconhecimento de fala e com uma interface simples, pratica e atrativa.

A solução criada pretendeu permitir aos utilizadores da mesma poderem praticar a fala praticamente sozinhos e assim poderem aprender novos fonemas, isto é, após estes terem sido instruídos e acompanhados pelos respetivos TF sobre o funcionamento do sistema. É possível afirmar que que o objetivo não foi inteiramente conseguido pois seria necessário desenvolver mais o protótipo, talvez recorrendo a outras *frameworks* que permitissem uma melhor manutenção e desenvolvimento, seria também interessante e de maior valia validar o projeto junto a Terapeutas da Fala e pessoas com dificuldades na fala para poder saber até que ponto a aplicação ajuda a melhorar a fala. Do ponto de vista funcional o protótipo está operativo, mas necessitaria ainda de algumas melhorias de maneira a poder ser mais atrativo aos utilizadores e não conter pequenos erros, para tal seria necessária uma maior rapidez do site ou uma melhor precisão do sistema de reconhecimento de fala.

Em suma podemos afirmar que se conseguiu atingir o objetivo a que nos havíamos proposto, através da apresentação de um protótipo.

7.2 Trabalho Futuro

Existem algumas considerações finais a serem tidas em conta em relação ao trabalho futuro. Tais como um desenvolvimento mais avançado do design das interfaces e de funcionalidades da aplicação. Seria igualmente importante adaptar o layout do sistema aos dispositivos móveis de maneira que funcionara melhor. Outro ponto seria o de desenvolver o sistema recorrendo a outras tecnologias, tais como por exemplo o da *framework* ReactJS ou outras de maneira a desenvolver a aplicação por componentes e assim facilitar a manutenção da aplicação a nível futuro.

Como referido anteriormente no futuro seria igualmente importante avaliar a aplicação junto a pacientes de TdF e não apenas a uma maioria de utilizadores não afetos à área. Assim poder-se-ia verificar melhor a efetividade e eficiência que a aplicação tem junto do público alvo. Também seria interessante incorporar princípios de acessibilidade e seguir as respetivas diretrizes no desenvolvimento. Outra mais valia passaria por divulgar esta ferramenta junto a TF para que estes pudessem adotá-la como uma ferramenta de ajuda com os pacientes. Se possível escrita de um artigo científico também seria algo positivo que ajudaria a dar um maior reconhecimento à aplicação.

Após a finalização do presente estudo pretendemos avançar para uma fase de escrita científica para o divulgar junto dos interessados no tema, como terapeutas de fala, e engenheiros com vontade de trabalhar soluções desta natureza.

Pretendemos fazer evoluir o estudo aqui iniciado e encontram-se neste momento em finalização uma proposta de um projeto associado ao tema precisamente com o nome “TERAFA” que irá ser submetido Fundação para a Ciência e Tecnologia (FCT),

Referências

- (AKAO 2014) Akao, Y. (2014). The Method for Motivation by Quality Function Deployment (QFD). *Nang Yan Business Journal*, 1(1), 1–9. <https://doi.org/10.2478/nybj-2014-0001>
- (BATISTA 2011) Apple. (n.d.). *Siri does more than ever*. Siri. <https://www.apple.com/siri/>
- (BATISTA 2011) Batista, J. S. N. (2011). *O PERFIL DO TERAPEUTA DA FALA EM PORTUGAL*.
- (BERDASCO 2019) Berdasco, López, Diaz, Quesada, & Guerrero. (2019). User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana. *Proceedings*, 31(1), 51. <https://doi.org/10.3390/proceedings2019031051>
- (CARNEIRO 2014) Carneiro, I., Fernandes, J., Sousa, M., & Kovalchuk, N. (2014). *Funny child talk: Sistemas Interativos em Saude*. 1–27.
- (COOK 2002) Cook, S. (2002). *Speech Recognition HOWTO*.
- (DAREJEH 2019) Darejeh, A., Salim, S. S., & Asemi, A. (2019). Speech pronunciation practice system for speech-impaired children: a systematic review of impacts and functionality. *Universal Access in the Information Society*, 18(1), 169–189. <https://doi.org/10.1007/s10209-017-0573-5>
- (DEVELOPERS 2020) Developers, A. (2020). *O que é Alexa?* <https://developer.amazon.com/pt-BR/alexa>
- (ERRATTAHI 2018) Errattahi, R., Hannani, A. El, & Ouahmane, H. (2018). Automatic speech recognition errors detection and correction: A review. *Procedia Computer Science*, 128, 32–37. <https://doi.org/10.1016/j.procs.2018.03.005>
- (FURLONG 2017) Furlong, L., Erickson, S., & Morris, M. E. (2017). Computer-based speech therapy for childhood speech sound disorders. *Journal of Communication Disorders*, 68(February 2016), 50–69. <https://doi.org/10.1016/j.jcomdis.2017.06.007>
- (GALVÃO 2009) Galvão Filho, T. (2009). A Tecnologia Assistiva: De Que Se Trata? *Porto Alegre: Redes Editora*, 207–235. https://www.mendeley.com/catalogue/b636e3ae-4f2e-3ff1-937c-a9d19971345d/?utm_source=desktop&utm_medium=1.19.4&utm_campaign=open_catalog&userDocumentId=%7B51238ab0-a590-413d-abfc-204456bca67b%7D
- (HERSTATT 2014) Herstatt, C. (2014). *The ‘ Fuzzy Front End ’ of Innovation The „ FuzzyFront End “ of Innovation The “ Fuzzy Front End ” of Innovation Prof . Dr . Cornelius Herstatt. January 2004*. <https://doi.org/10.1057/9780230512771>
- (HOY 2018) Hoy, M. B. (2018). Alexa , Siri , Cortana , and More : An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- (HUNICKE 2004) Hunicke, R., Leblanc, M., & Zubek, R. (2004). *MDA : A Formal Approach to Game Design and Game Research*.
- (KEPUSKA 2017) Këpuska, V., & Bohouta, G. (2017). *Comparing Speech Recognition Systems (Microsoft API , Google API And CMU Comparing Speech Recognition Systems (Microsoft API , Google API And CMU Sphinx). March*. <https://doi.org/10.9790/9622-0703022024>
- (KOEN 2015) Koen, P. A., Bertels, H. M. J., Kleinschmidt, E., Koen, P. A., Bertels, H. M. J., & Managing, E. K. (2015). *Managing the Front End of Innovation — Part I : Results From a Three-Year Study Managing the Front End of Innovation — Part I Results From a Three-Year Study*. 6308. <https://doi.org/10.5437/08956308X5702145>
- (KOEN 2001) Koen, P., Ajamian, G., Burkart, R., Clamen, A., Davidson, J., Amore, R. D., Elkins, C., Herald, K., Incorvia, M., Johnson, A., Karol, R., Seibert, R., Slavejkov, A., & Wagner, K. (2001). *PROVIDING CLARITY AND A COMMON LANGUAGE TO THE “ FUZZY FRONT END .”* 46–55.
- (NIODAOCD 2015) National Institute on Deafness and Other Communication Disorders. (2020). *Statistics on Voice, Speech, and Language*. <https://www.nidcd.nih.gov/health/statistics/statistics-voice-speech-and-language>
- (OSTERWALDER 2013) Osterwalder, A. (2013). A better way to think about your business model. *Harvard*

Business Review, May 06, 2–4. <https://hbr.org/2013/05/a-better-way-to-think-about-yo/>

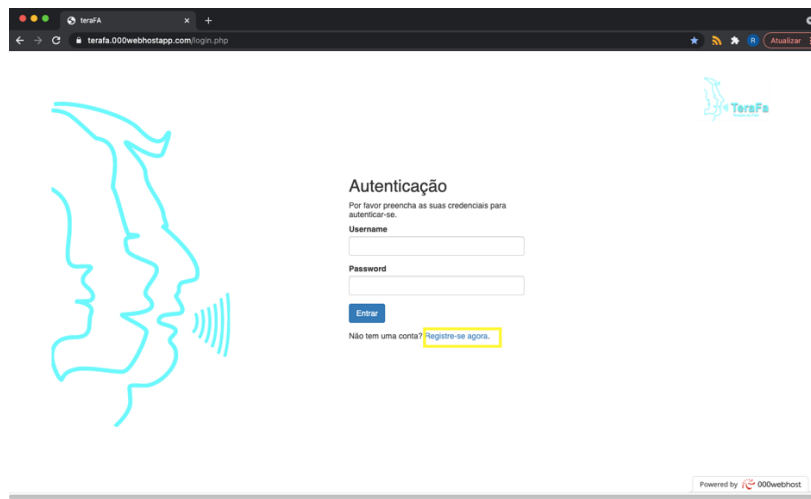
- (PENG 2018)** Peng, J. G., Xia, G., Sun, B. Q., & Wang, S. J. (2018). Systematical decision-making approach for quality function deployment based on uncertain linguistic term sets. *International Journal of Production Research*, 56(18), 6183–6200. <https://doi.org/10.1080/00207543.2018.1478462>
- (RIBEIRO 2019)** Ribeiro, R. M. P. V. (2019). *EasyTEFA - Solução para apoio à terapia da fala*.
- (RICH 2000)** Rich, N., & Holweg, M. (2000). Rich, BSc MBA Matthias Holweg, Dipl.-*Wirtschaftsing.(FH) MSc*. 32. https://www.urenio.org/tools/en/value_analysis.pdf
- (ROBSON 2015)** Robson, K., Plangger, K., Kietzmann, J. H., & Mccarthy, I. (2015). Is it all a game ? Understanding the principles of gamification. *Business Horizons*, 58(4), 411–420. <https://doi.org/10.1016/j.bushor.2015.03.006>
- (RUA 2015)** Rua, M. (2015). *Caraterização do desempenho articulatorio e oromotor de crianças com alterações da fala*. 67. [https://comum.rcaap.pt/bitstream/10400.26/9607/1/Trabalho Final de Projeto - Mafalda Rua 2012141.pdf](https://comum.rcaap.pt/bitstream/10400.26/9607/1/Trabalho%20Final%20de%20Projeto%20-%20Mafalda%20Rua%202012141.pdf)
- (SAATY 1990)** Saaty, T. L. (1990). *How to make a decision : The Analytic Hierarchy Process*. 48.
- (SAATY 2008)** Saaty, T. L. (2008). *Decision making with the analytic hierarchy process*. 1(1).

ANEXO 1 - Testes de usabilidade e funcionalidade do TeraFA

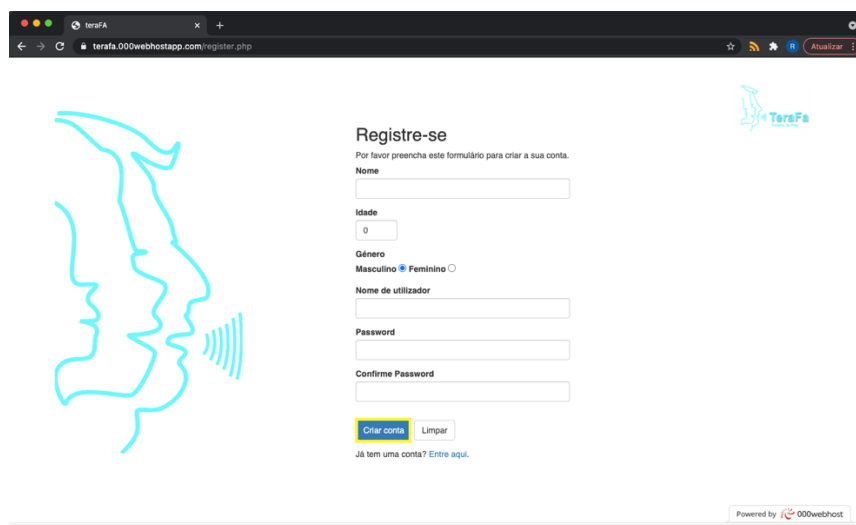
Ricardo Miguel Barros Ramiro

1 Registo de utilizador

- 1.1 Para realizar os testes é necessário utilizar um computador e aceder com o navegador Google Chrome.
- 1.2 Introduza a seguinte URL no navegador Google Chrome
<https://terafa.000webhostapp.com>
- 1.3 Seleccione o link “Registe-se agora” para se registar e assim poder autenticar no sistema.

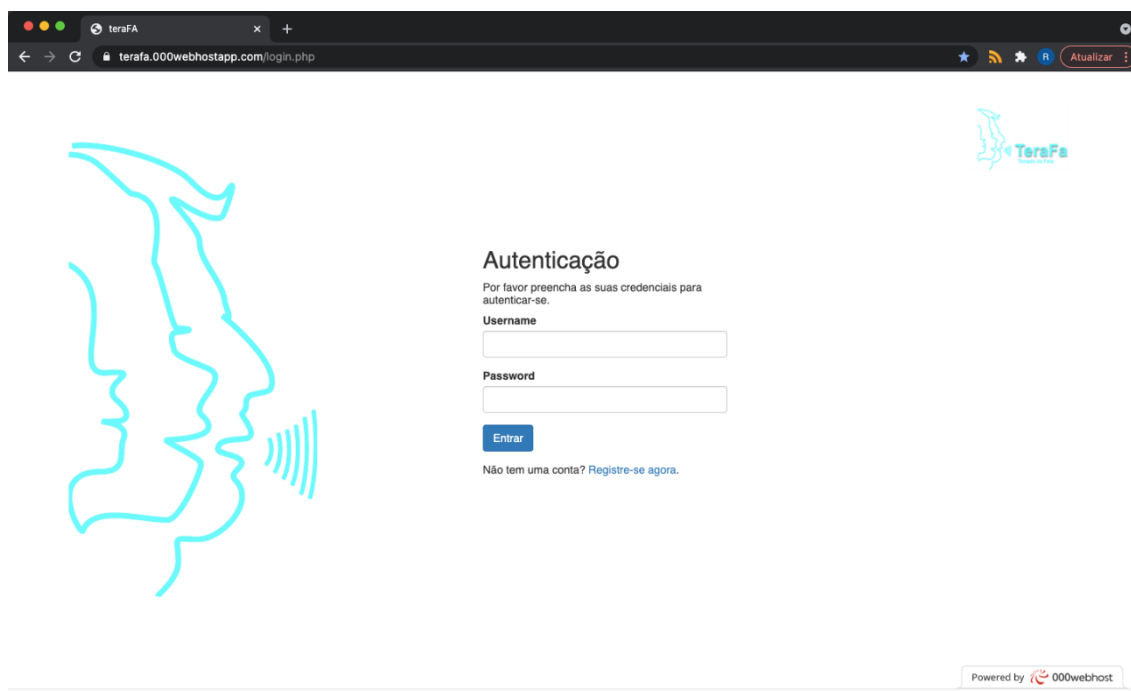


- 1.4 Preencha os campos requeridos no formulário e seleccione o botão “Criar conta”.



2 Acesso com autenticação

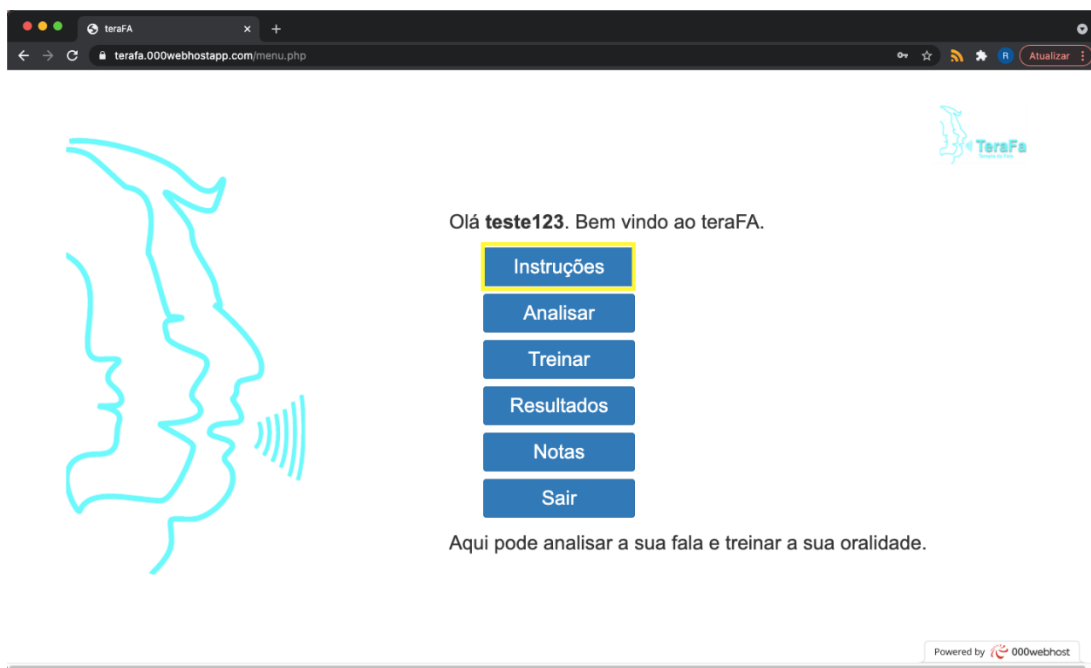
- 2.1 Aceda novamente ao website <https://terafa.000webhostapp.com> através do navegador Google Chrome do seu computador. Previamente dê permissão de utilizar o microfone no seu computador.
- 2.2 Aparecerá o ecrã de login da aplicação. Aceda com o seu username e password que registou anteriormente e selecione o botão “Entrar”.



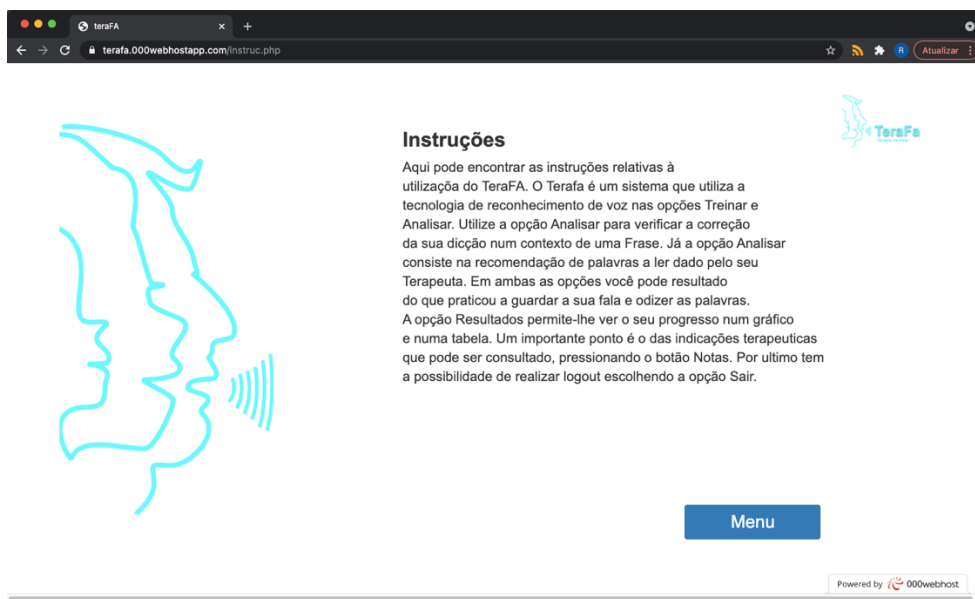
- 2.3 Depois de autenticado aparecerá o ecrã do menu de opções.

3 Acesso às instruções da aplicação

- 3.1 Pressione o botão **Instruções** para aceder e consultar como funciona a aplicação.



3.2 Leia as instruções que aparecem no ecrã, para conhecer a estrutura e funcionamento da aplicação.



3.3 Pressione o botão azul **Menu** para voltar ao ecrã principal.

Instruções

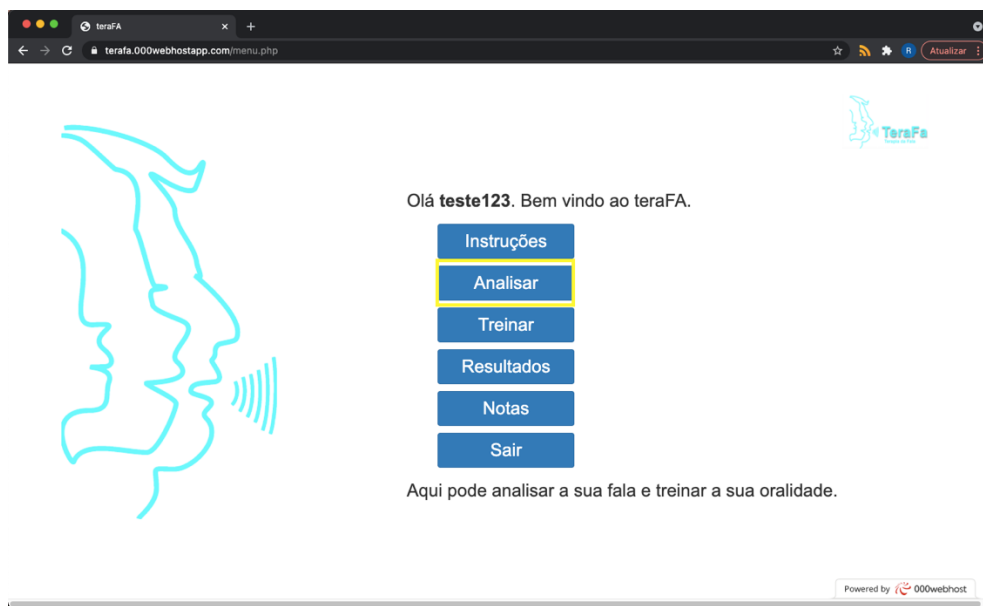
Aqui pode encontrar as instruções relativas à utilização do TeraFA. O Terafa é um sistema que utiliza a tecnologia de reconhecimento de voz nas opções Treinar e Analisar. Utilize a opção Analisar para verificar a correção da sua dicção num contexto de uma Frase. Já a opção Analisar consiste na recomendação de palavras a ler dado pelo seu Terapeuta. Em ambas as opções você pode resultado do que praticou a guardar a sua fala e odizer as palavras. A opção Resultados permite-lhe ver o seu progresso num gráfico e numa tabela. Um importante ponto é o das indicações terapeuticas que pode ser consultado, pressionando o botão Notas. Por ultimo tem a possibilidade de realizar logout escolhendo a opção Sair.

Menu

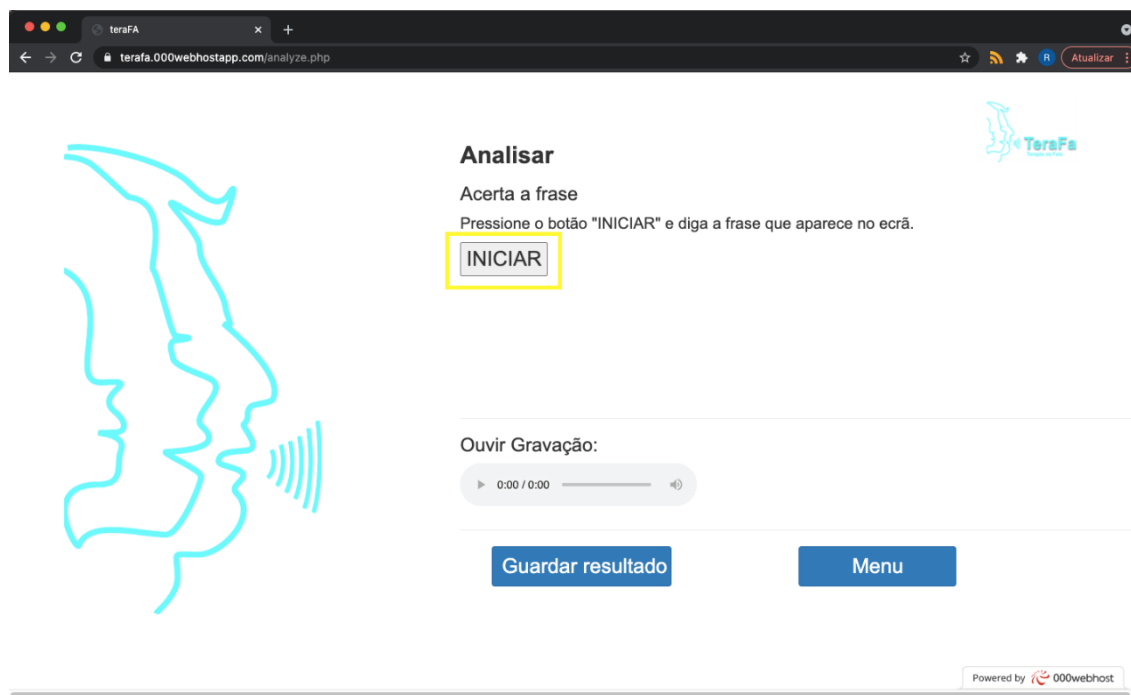
Powered by 000webhost

4 Aceder à opção de Análise

4.1 Pressione o botão **Analisar** para analisar a sua fala.



4.2 Siga as instruções que aparecem no ecrã, pressionando o botão INICIAR e diga a frase que vai aparecer.



4.3 Depois de dita a frase, carregue no botão play para ouvir a gravação do que disse.

terafa

terafa.000webhostapp.com/analyze.php

Atualizar

TeraFa

Analisar

Acerta a frase

Pressione o botão "INICIAR" e diga a frase que aparece no ecrã.

Iniciar

o rato roeu a rolha

Recebido: o rato roeu a rolha

Ponto(s): 1

Ouvir Gravação:

x:00

Guardar resultado

Menu

Powered by 000webhost

4.4 Carregue agora no botão "Guardar resultado" para guardar a pontuação e seleccione guardado OK.

terafa

terafa.000webhostapp.com/analyze.php

Atualizar

TeraFa

Analisar

Acerta a frase

Pressione o botão "INICIAR" e diga a frase que aparece no ecrã.

Iniciar

o rato roeu a rolha

Recebido: o rato roeu a rolha

Ponto(s): 1

Ouvir Gravação:

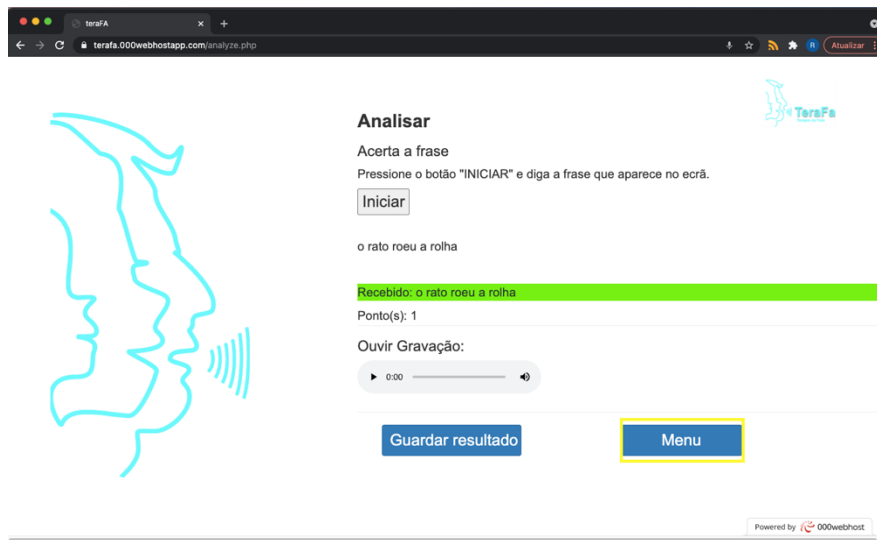
0:00

Guardar resultado

Menu

Powered by 000webhost

4.5 Carregue agora no botão “Menu” para regressar ao menu principal.



terafa

terafa.000webhostapp.com/analyze.php

Atualizar

Analisar

Acerta a frase

Pressione o botão "INICIAR" e diga a frase que aparece no ecrã.

Iniciar

o rato roeu a rolha

Recebido: o rato roeu a rolha

Ponto(s): 1

Ouvir Gravação:

0:00

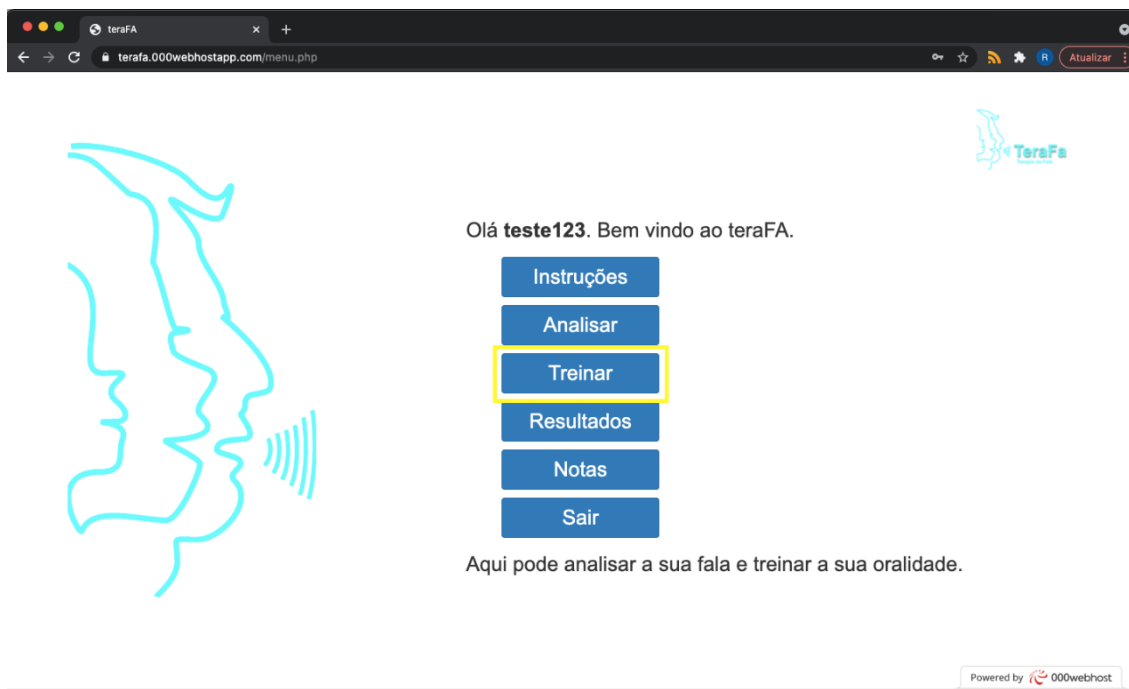
Guardar resultado

Menu

Powered by 000webhost

5 Aceder à opção de Treino

5.1 Pressione o botão **Treinar** para analisar a sua fala.



terafa

terafa.000webhostapp.com/menu.php

Atualizar

Olá teste123. Bem vindo ao teraFA.

Instruções

Analisar

Treinar

Resultados

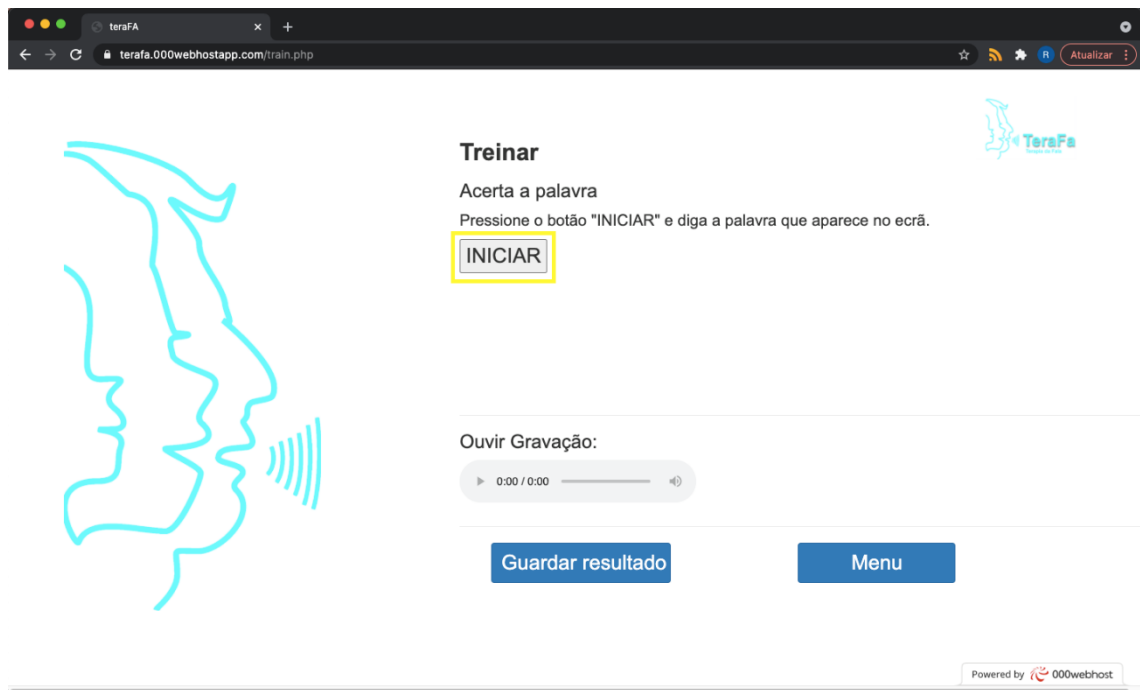
Notas

Sair

Aqui pode analisar a sua fala e treinar a sua oralidade.

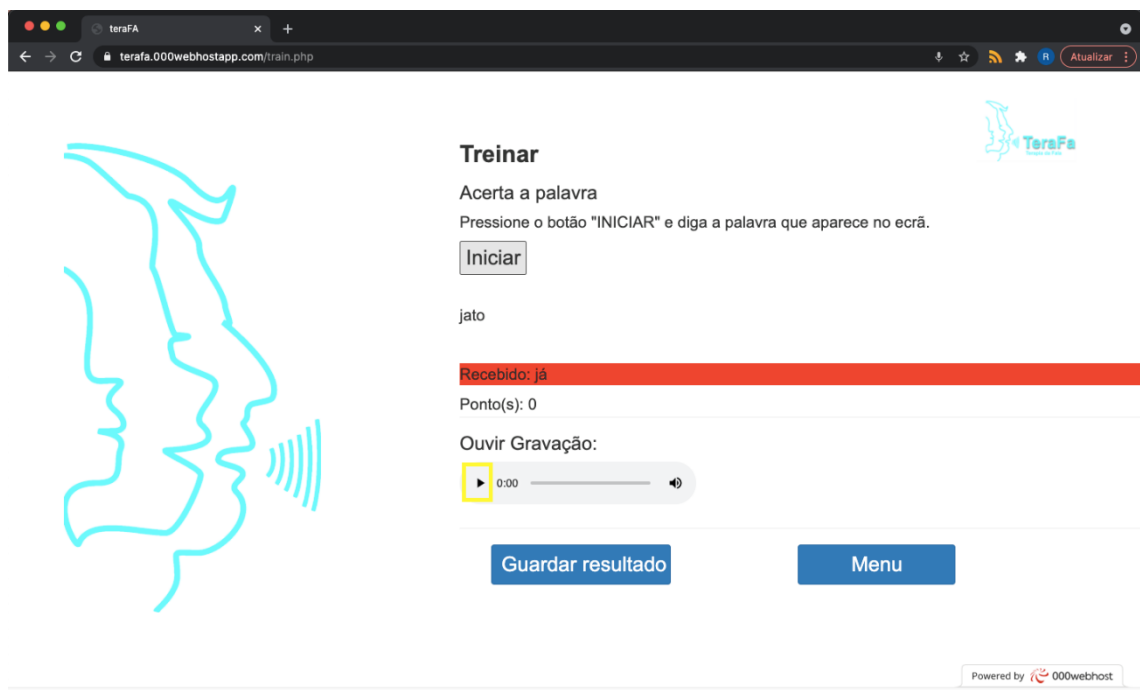
Powered by 000webhost

5.2 Siga as instruções que aparecem no ecrã, pressione o botão INICIAR e diga a palavra que vai aparecer no ecrã.



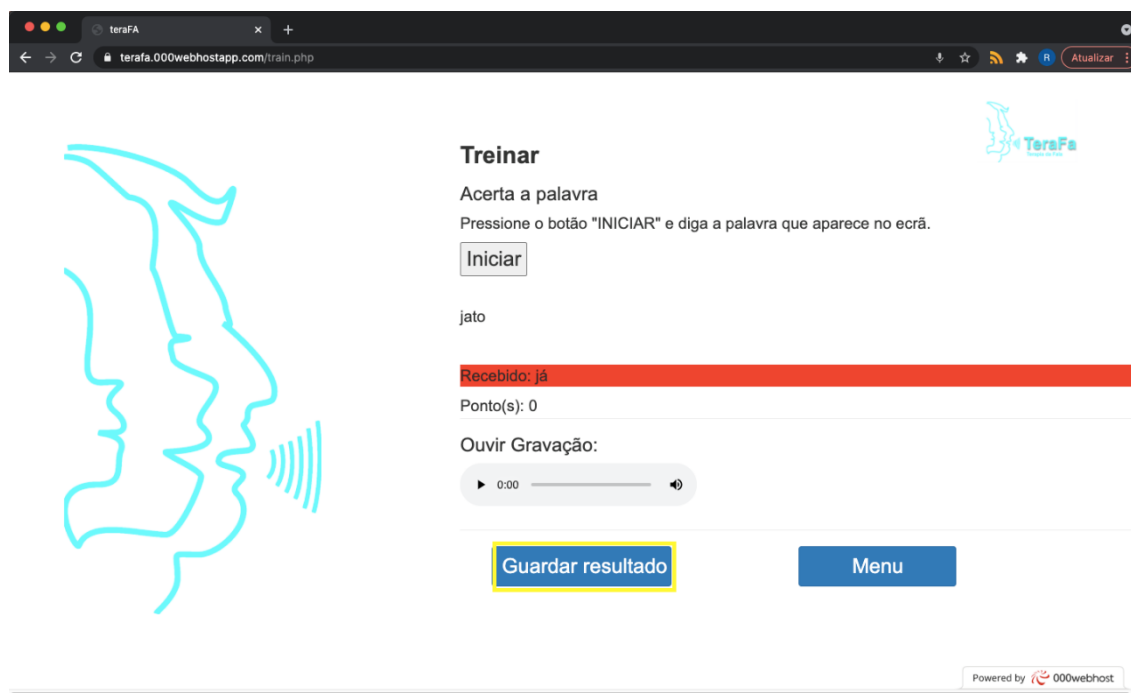
The screenshot shows a web browser window with the URL `terafa.000webhostapp.com/train.php`. The page features a large blue outline of a person's head and neck on the left. On the right, the heading "Treinar" is followed by the instruction "Acerta a palavra" and "Pressione o botão 'INICIAR' e diga a palavra que aparece no ecrã." The button "INICIAR" is highlighted with a yellow border. Below this is a "Ouvir Gravação:" section with a play button and a 0:00 / 0:00 timer. At the bottom, there are two blue buttons: "Guardar resultado" and "Menu". A "Powered by 000webhost" logo is visible in the bottom right corner.

5.3 Depois de dita a palavra, carregue no botão play para ouvir a gravação do que disse.



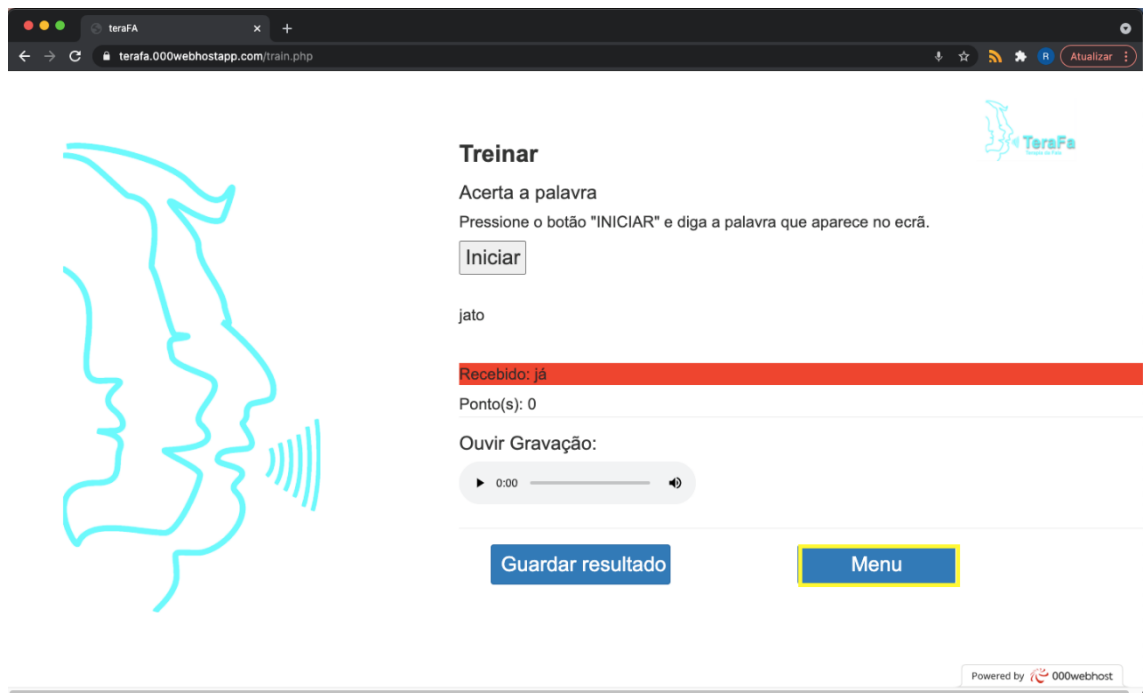
The screenshot shows the same web browser window as in 5.2. The "INICIAR" button has been replaced by "Iniciar". The word "jato" is displayed below the instruction. A red bar with the text "Recebido: já" is visible. The "Ouvir Gravação:" section now has a play button highlighted with a yellow border. The "Guardar resultado" and "Menu" buttons remain at the bottom. The "Powered by 000webhost" logo is still present in the bottom right corner.

5.4 Carregue agora no botão “Guardar resultado” para guardar a pontuação, e seleccione guardado OK.



The screenshot shows a web browser window with the URL `terafa.000webhostapp.com/train.php`. The page title is "Treinar" and the instructions are "Acerta a palavra" and "Pressione o botão 'INICIAR' e diga a palavra que aparece no ecrã." There is an "Iniciar" button. The word "jato" is displayed. A red bar indicates "Recebido: já". The score is "Ponto(s): 0". There is a "Ouvir Gravação:" section with a play button and a 0:00 timer. Two buttons are visible: "Guardar resultado" (highlighted with a yellow box) and "Menu". A footer at the bottom right says "Powered by 000webhost".

5.5 Carregue agora no botão “Menu” para regressar ao menu principal.



The screenshot shows a web browser window with the URL `terafa.000webhostapp.com/train.php`. The page features a large cyan outline of a person's head in profile, facing right, with sound waves emanating from the mouth. To the right of the image, the text reads: **Treinar**, **Acerta a palavra**, and **Pressione o botão "INICIAR" e diga a palavra que aparece no ecrã.** Below this is a button labeled **Iniciar**. The word **jato** is displayed below the button. A red progress bar is labeled **Recebido: já**. Below that, it says **Ponto(s): 0**. There is a section for **Ouvir Gravação:** with a play button and a progress indicator showing **0:00**. At the bottom of the interface are two buttons: **Guardar resultado** and **Menu**. The footer of the page includes the text **Powered by 000webhost**.

6 Aceder à opção de Resultados

6.1 Seleccione a opção **Resultados** para consultar os seus resultados.

Olá **acastro**. Bem vindo ao teraFA.

Instruções

Analisar

Treinar

Resultados

Notas

Sair

Aqui pode analisar a sua fala e treinar a sua oralidade.

Powered by 000webhost

6.1 Consulte os **Resultados** obtidos.

Resultados

Username	Pontos	Data
acastro	0	2021-06-19 11:11:21
acastro	0	2021-06-19 11:11:33
acastro	0	2021-06-19 11:11:36
acastro	0	2021-06-19 18:48:28

Gráfico evolutivo

1

0

-1

2021-06-19 11:11:21 2021-06-19 11:11:33 2021-06-19 11:11:36 2021-06-19 18:48:28

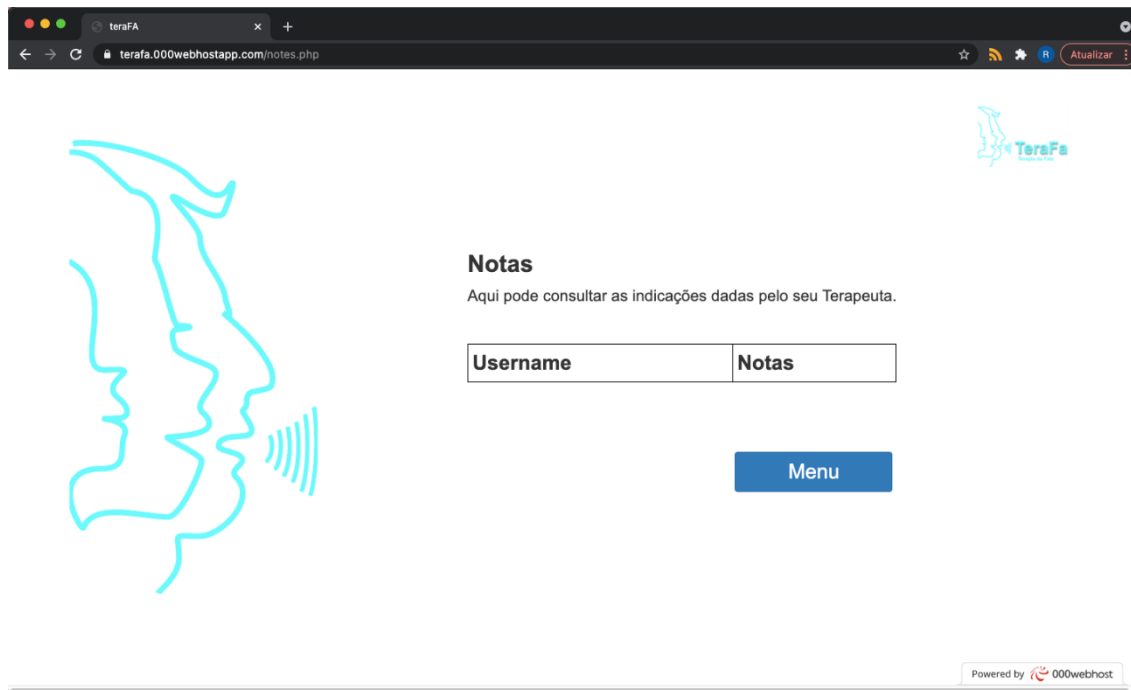
Menu

Powered by 000webhost

7 Aceder à opção de Notas do Terapeuta

7.1 No menu principal selecione a opção **Notas**.

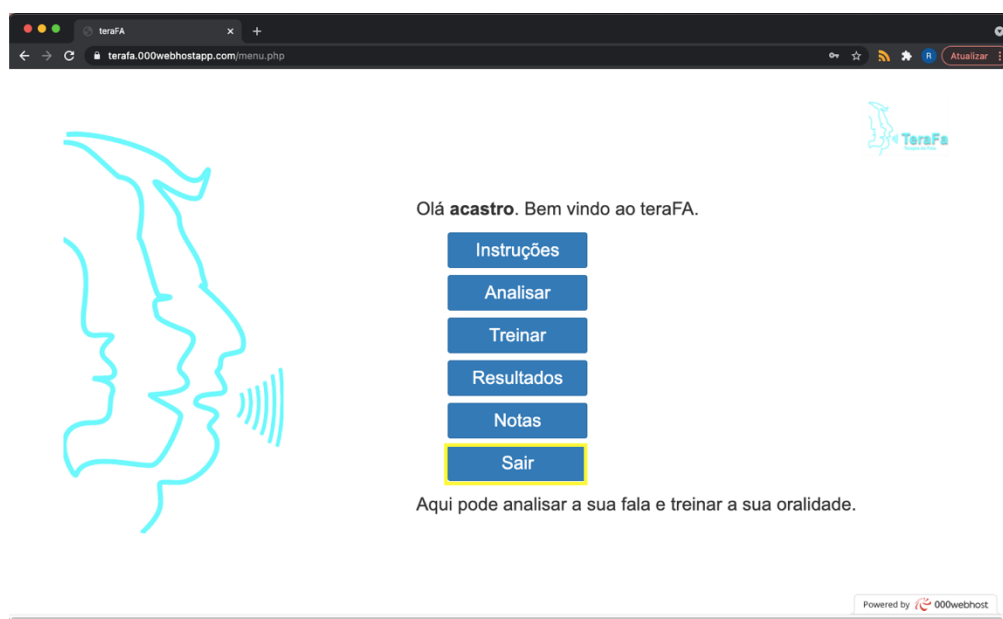
7.2 Verifique que consegue ver o seguinte ecrã.



7.2 Volte ao menu principal, selecionando a opção **Menu**.

8 Realizar logout

8.1 Selecione a opção **Sair** para realizar o logout da aplicação.



ANEXO 2 – Questionário



TeraFa
Terapia da Fala

QUESTIONÁRIO

ajude-nos a ajudar a terapia da fala



Sistemas de Interação e Inovação Social

Section 1 of 3

Bem vindo.

Este questionário é anónimo e confidencial. As suas respostas serão tratadas com o objectivo de identificar o potencial das tecnologias na Terapia da Fala, nomeadamente através da proposta pela solução TeraFA.

Deverá aceder à aplicação com browser Google Chrome num computador no URL <https://terafa.000webhostapp.com> para ter acesso à aplicação e executar testes de usabilidade e funcionalidades.

Agradecemos toda a atenção e tempo dispensados,

Cumprimentos,

Ricardo Ramiro, (ISEP / SIIS)

António Vieira de Castro (ISEP / SIIS)

Maria João Cunha (ESS)

Tânia Rocha (UTAD / SIIS)

* SIIS - Grupo de investigação em Sistemas de Interação e Inovação Social

Dados Gerais

Indique qual o seu sexo. *

(indicar Sim ou Não)

Masculino

Feminino

Indique a sua idade: *

(idade)

Choose

Habilitações literárias *

Até 4ªa classe

Até 12º ano

Bacharelato ou Licenciatura

Mestrado

Doutoramento

Outro

Já conhecia o termo "Terapia da Fala"? *

- Sim
- Não

Contacto com Terapia da Fala: Conhece alguém que tenha ou teve problemas de fala? *

- Sim
- Não

Tem ou teve problemas de fala? *

Choose ▼

Teste e avaliação inicial do TeraFA

Aceda a: <https://terafa.000webhostapp.com>

O que achou do sistema de reconhecimento de voz? *

- Excelente
- Muito bom
- Bom
- Razoável
- Mau

Achou fácil navegar pela aplicação? *

- Sim
- Não

Aprender novas funcionalidades foi fácil na aplicação? *

- Sim
- Não

Achou a aplicação segura? *

- Sim
- Não
- Talvez

De 1 a 5, quão confortável se sentiu a utilizar a aplicação? Sendo 5 "muito" e 1 "pouco". *

- | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 2 | 3 | 4 | 5 |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Sente que a aplicação pode ajudar pessoas a melhorar a fala? *

- Sim
- Não

O tempo dispendido para realizar as tarefas foi razoável? *

- Sim
- Não

Fui capaz de completar as tarefas corretamente? *

- Discordo completamente
- Discordo
- Neutro
- Concordo
- Concordo fortemente

Achou o design da aplicação lógico? *

- Sim
- Não

Detetou algum erro na aplicação? *

- Sim
- Não

Que melhoria realizaria na aplicação? *

Your answer _____

Recomendaria a aplicação? *

Sim

Não

Considera o uso de computador uma mais valia? *

Sim

Não