

INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE SÃO PAULO
CAMPUS BRAGANÇA PAULISTA (IFSP-BRA)

Bruna Camargo Galindo
Fernanda Nascimento de Paula
Sarah Danniely Barbosa de Souza

**PARLA: Plataforma para auxiliar deficientes auditivos
e surdos na oralização e na leitura labial**

Bragança Paulista
2024

Bruna Camargo Galindo
Fernanda Nascimento de Paula
Sarah Danniely Barbosa de Souza

**PARLA: Plataforma para auxiliar deficientes auditivos
e surdos na oralização e na leitura labial**

Relatório solicitado como parte dos requisitos para a apresentação de trabalho de pesquisa científica para a 14ª BRAGANTEC, realizada de 17 a 19 de outubro de 2024, no IFSP *campus* Bragança Paulista.

Orientador: Prof. Dr. Rafael Prearo Lima
Co-orientadora: Profa. Me. Talita de P. Cypriano de Souza

Bragança Paulista
2024

RESUMO

O objetivo deste projeto é o desenvolvimento de uma plataforma para auxiliar no processo de oralização e de aprendizagem de leitura labial para surdos e para pessoas com surdez moderada ou profunda. À luz de três dos dezessete Objetivos de Desenvolvimento Sustentável (ODS) previstos pela ONU, justificamos o desenvolvimento do projeto pela necessidade de facilitar a comunicação entre pessoas com deficiência auditiva severa ou profunda e de pessoas surdas com pessoas do mundo ouvinte. Para implementar a plataforma, criamos um banco de dados com gravações de vozes e, a partir dele, um modelo de machine learning foi treinado para o reconhecimento de sons produzidos pela fala humana. O usuário, por sua vez, ao observar um gif com movimentos faciais, tenta reproduzir o som esperado, que é confirmado, ou não, pelo modelo treinado. Os testes preliminares, feitos com base nas vogais do português brasileiro e com pessoas do mundo ouvinte, obtendo um índice de acurácia de 80%. Os principais erros foram identificados no reconhecimento dos sons É e Ó, que foram confundidos com E e O, respectivamente. Após ajustes e calibrações na inteligência artificial (IA), o índice de acurácia aumentou para 90%, e, com ajustes adicionais, conseguimos alcançar quase 100% de precisão.

Palavras-chave: Oralização. Surdez. Leitura labial.

ABSTRACT

The objective of this project is the development of a platform to assist in the process of oralization and learning to read lips for deaf people and people with moderate or profound deafness. In light of three of the seventeen Sustainable Development Goals (SDGs) foreseen by the UN, we justify the development of the project by the need to facilitate communication between people with severe or profound hearing impairment and deaf people with people from the hearing world. To implement the platform, we created a database with voice recordings and, from it, a machine learning model was trained to recognize sounds produced by human speech. The user, in turn, when observing a gif with facial movements, tries to reproduce the expected sound, which is confirmed, or not, by the trained model. Preliminary tests, carried out based on in Brazilian Portuguese vowels and with people from the listening world, obtaining an accuracy rate of 80%. The main errors were identified in the recognition of the sounds É and Ó, which were confused with E and O, respectively. After adjustments and calibrations in artificial intelligence (AI), the accuracy rate increased to 90%, and, with additional adjustments, we were able to achieve almost 100% accuracy.

Keywords: Oralization. Deafness. Lip reading.

SUMÁRIO

1	INTRODUÇÃO	6
2	OBJETIVOS	7
	2.1 Objetivos gerais	7
	2.2 Objetivos específicos	7
3	DESENVOLVIMENTO DO PROJETO	9
4	RESULTADOS	19
5	CONCLUSÃO	21
	REFERÊNCIAS	22

1 Introdução

Dados do IBGE, do ano de 2020, divulgados pela Assembleia Legislativa de São Paulo (Alesp, 2021), apontam que, a cada duzentos bebês, três nascem com surdez total no Brasil. Indo além, quando pensamos no número dos que possuem algum grau de deficiência auditiva acima do moderado ou surdez no Brasil, o total já ultrapassa a marca de 10 milhões de brasileiros (Lemos, 2023).

Esses números impactam diretamente a questão da educação no Brasil, pois, de acordo com o Censo Escolar do Inep (Inep, 2023), dos 47,3 milhões de alunos da educação básica, 61.594 possuem alguma deficiência relacionada à surdez, o que indica que, talvez, um número grande esteja fora do ambiente escolar. Nesse sentido, um dos desafios é a questão dos estudos, visto que, por falta de ensino adequado, ou por falhas em sua adaptação, grande parte desses indivíduos acaba abandonando o ensino formal. Crianças, por exemplo, no ambiente escolar, podem encontrar barreiras que as impeçam de ser incluídas nos estudos, em brincadeiras e até mesmo para que sejam ajudadas por aqueles que não têm conhecimento da Língua Brasileira de Sinais, a Libras. Jovens e adultos, por sua vez, mesmo sabendo Libras, podem perder oportunidades de emprego por obstáculos na comunicação, como aponta Magno (2021), além de terem que enfrentar outras dificuldades em sua rotina diária.

A partir dessa necessidade, descobrimos que várias pessoas surdas conseguiram ser oralizadas. Uma delas foi Alex Bill (do perfil @alexbilljr). Diagnosticado com deficiência auditiva bilateral com um ano de idade, Alex é um surdo oralizado que aprendeu a falar, entre outros fatores, com a ajuda da leitura labial e, por causa disso, hoje trabalha como palestrante e *influencer* digital.

O caso de Alex Bill não é isolado. Alguns estudos, como o de Toffolo et al (2017) e o de Pinheiro et al (2020), apontam para os benefícios que o desenvolvimento das habilidades de leitura labial e da oralização trazem para a pessoa com surdez. Dentre eles, estão uma melhor comunicação dos surdos com o mundo ouvinte e uma maior facilidade no processo de leitura. Assim, tanto a oralização como a leitura labial podem auxiliar os surdos no ambiente escolar, o que é importante não apenas para seu futuro acadêmico, como também para o profissional.

Diante desse cenário, levantamos as seguintes perguntas-problema, que direcionam nosso projeto: como podemos auxiliar surdos e pessoas com deficiência auditiva severa ou profunda em sua comunicação com o mundo ouvinte? Como podemos treinar um modelo de aprendizado de máquina para ajudar na oralização e na aprendizagem de leitura labial desses indivíduos?

Assim, decidimos criar o Parla, uma plataforma que visa auxiliar na oralização e na aprendizagem de leitura labial de pessoas com deficiência auditiva severa ou profunda e de surdos. Treinado por meio de inteligência artificial (IA), o Parla serve como uma ferramenta de apoio para esses indivíduos que buscam outras formas, além Libras, de interagir com o mundo ouvinte.

Justificamos, pois, a criação do Parla como uma ferramenta disponível para que pessoas com deficiência auditiva severa ou profunda e também os surdos possam, caso queiram, ser oralizadas e aprendam a fazer leitura labial, a fim de facilitar sua comunicação com o mundo ouvinte.

Também justificamos a criação dessa plataforma a partir de três dos dezessete Objetivos de Desenvolvimento Sustentável (ODS) recomendados pela ONU (2024), a saber, os objetivos (3) Saúde e Bem-estar, entendendo que a facilidade na comunicação pode possibilitar novos relacionamentos, o que traz benefícios para a saúde mental dos indivíduos; (4) Educação de Qualidade, visto que a plataforma pode facilitar a comunicação no ambiente educacional, tornando-o mais inclusivo, o que pode garantir que todos os estudantes, independentemente de suas capacidades, tenham acesso às mesmas oportunidades de aprendizagem e desenvolvimento; e (10) Redução das desigualdades, pois promove a inclusão social e econômica de surdos ao mitigar barreiras de comunicação, permitindo que esses indivíduos tenham acesso igualitário a oportunidades e serviços.

Nossa hipótese é a de que, por meio do treino de uma inteligência artificial (IA) através do *machine learning* para reconhecer os sons de fala, e com a ajuda de gifs com indicação de movimento faciais (dos lábios, da boca, da língua), o Parla possa auxiliar pessoas com deficiência auditiva severa ou profunda e surdos no processo de oralização e de aprendizagem de leitura labial.

2 OBJETIVOS

2.1 Objetivo geral

O objetivo geral do Parla é desenvolver uma plataforma para auxiliar no processo de oralização e de aprendizagem de leitura labial para surdos e para pessoas com deficiência auditiva grave ou profunda.

2.2 Objetivos específicos

Com base no objetivo geral, estabelecemos os seguintes objetivos específicos:

(i) desenvolver uma plataforma online, dividida em níveis, fases e etapas; (ii) montar um banco de dados composto por gravação de vozes de vogais, consoantes, sílabas simples e sílabas complexas; (iii) treinar um modelo de aprendizado de máquina , a partir de um banco de dados com vozes, para reconhecer os sons produzidos pela fala humana, e criar gifs com movimentos faciais e labiais para ensinar os usuários como os sons são produzidos.

3 DESENVOLVIMENTO DO PROJETO

Primeiramente, realizamos um levantamento das necessidades específicas na sociedade que gostaríamos de suprir e, alinhados com os Objetivos de Desenvolvimento Sustentável (ODS), decidimos focar no auxílio a pessoas com deficiência auditiva severa ou profunda, incluindo surdos. Em seguida, elaboramos perguntas-problema para direcionar nosso trabalho e, a partir delas, estabelecemos nossos objetivos.

Aprofundamos nossas pesquisas para coletar informações relevantes e entender como poderíamos ajudar essas pessoas. Após compreendermos parte dos desafios, dificuldades e métodos já existentes para nosso público-alvo, começamos a desenvolver uma plataforma para a oralização e o ensino de leitura labial para surdos e pessoas com deficiência auditiva severa ou profunda.

3.1 Tecnologias

O Parla foi desenvolvido utilizando HTML, CSS, JavaScript e PHP, e está interligado ao banco de dados no MySQL Workbench.

Começamos o desenvolvimento da base do protótipo da plataforma Parla criando nossa base de dados. Para gerenciar os cadastros, utilizamos o MySQL. Em seguida, criamos as telas de evolução, de contato, de atividades e a tela inicial (Figura 1, abaixo). Para isso, utilizamos HTML para o desenvolvimento do corpo do protótipo, CSS para estilizar, posicionar e limitar o tamanho dos elementos, JavaScript para tornar as páginas web interativas e dinâmicas, o que possibilitou diferentes funcionalidades, como validação de formulários e integração com a inteligência artificial (IA), entre outras. Além disso, utilizamos o *Teachable Machine*, uma ferramenta essencial para o desenvolvimento e treino do modelo *machine learning* do projeto, a fim de que a IA, desenvolvida com técnicas de *machine learning*, pudesse reconhecer os sons emitidos pelo usuário e filtrar ruídos indesejados com a devida precisão.

Figura 1: Parla – tela inicial



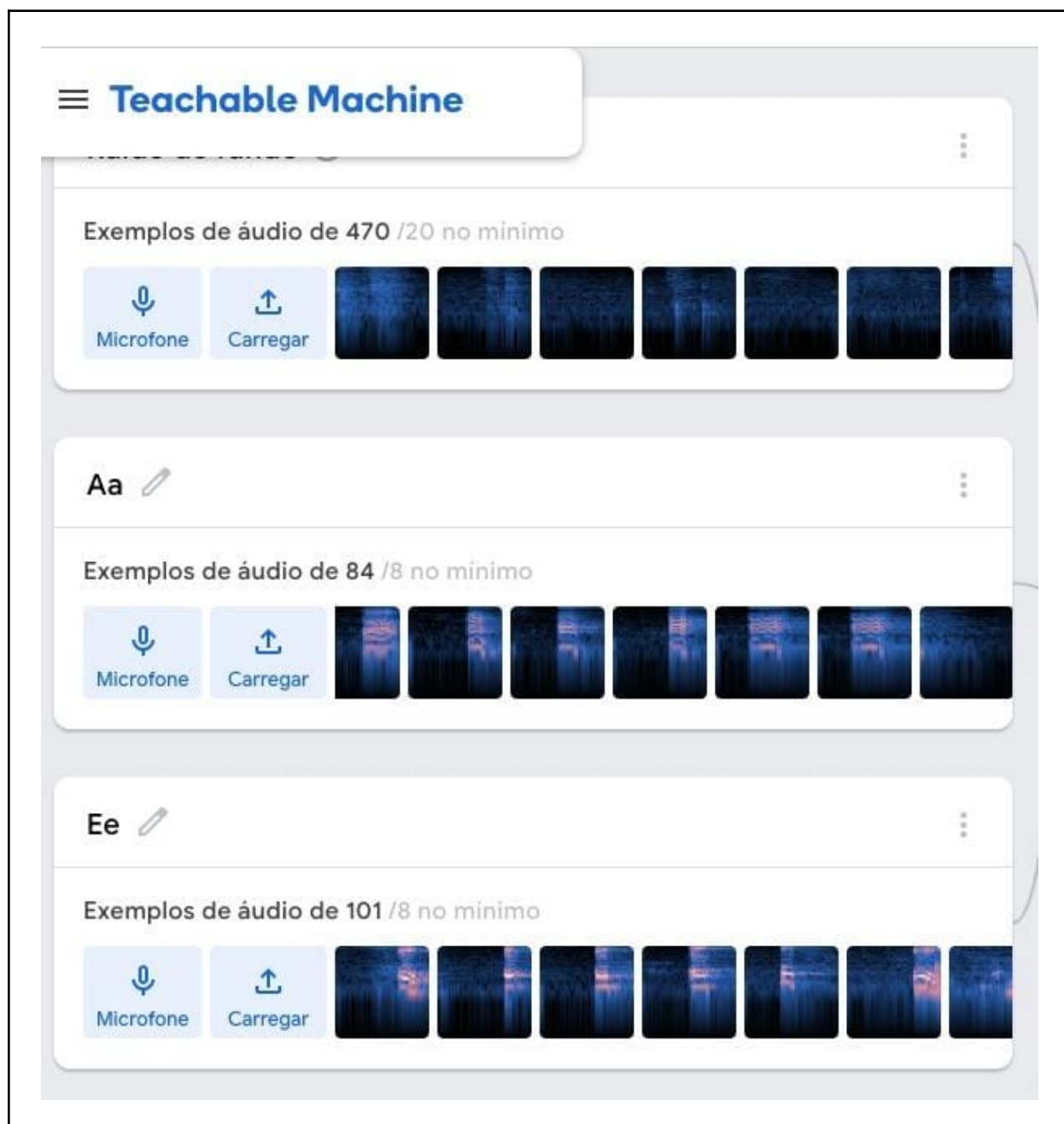
Fonte: autoria própria.

3.2 Machine learning e Teachable Machine

Machine learning é o processo de otimização do modelo para que este possa prever a resposta correta com base nas amostras de dados de treinamento. Nesse sentido, quanto mais amostras de treinamento o algoritmo de *machine learning* receber, mais preciso será o modelo.

Como ferramenta de *machine learning* usamos o *Teachable Machine* (Figura 2, abaixo). A fim de treinar o algoritmo, contamos com a colaboração de várias pessoas para gravar os níveis iniciais com diferentes formas, entonações e tipos de voz, a fim de construir um banco de dados extenso que permitisse à inteligência artificial (IA) realizar o reconhecimento e a correção de sons da maneira mais eficaz possível.

Figura 2: *Teachable machine* – exemplos de áudio



Fonte: <https://teachablemachine.withgoogle.com/models/oE20awNax/>. Acesso em 29 abr. 2024.

Assim, para garantir um funcionamento satisfatório, começamos as gravações a partir de diversos tipos de ruídos de fundo que pudessem eventualmente interferir na experiência do usuário em situações normais, como ruídos brancos, chiados,

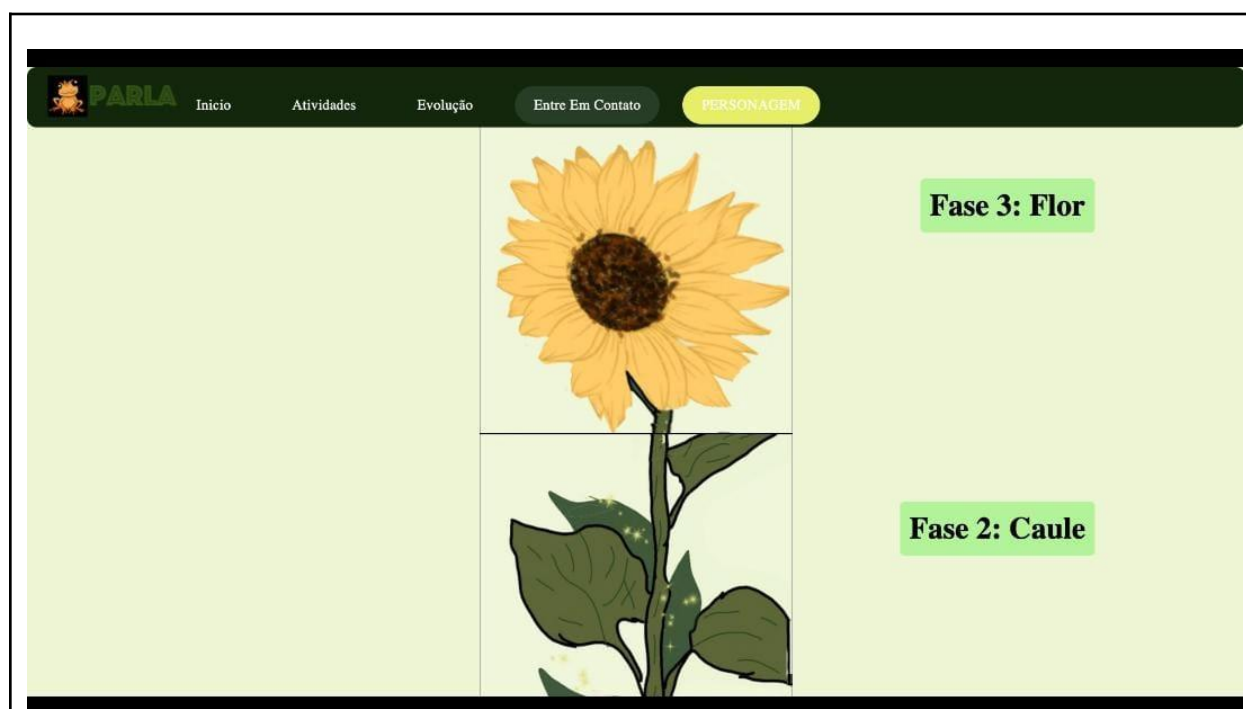
respirações, conversas paralelas, sons de animais de estimação, manuseio de materiais, buzinas, sons ambientes, entre outros.

3.3 Funcionamento

Ao utilizar a plataforma, o usuário aprenderá algo diferente, em um nível crescente de dificuldade: (a) vogais; (b) consoantes em sílabas simples; (c) sílabas complexas. Cada atividade incluirá uma imagem GIF com movimentos faciais e, ao lado dela, a imagem da letra correspondente para o usuário reproduzir.

Em cada uma das fases (a) a (c) mencionadas acima, haverá na plataforma três fases distintas: a base (inicial), o caule (intermediário) e a flor (avançado). Cada fase é subdividida em níveis e etapas específicas. Até o momento, as fases foram desenvolvidas da seguinte forma:

Figura 3: Parla – fases



Fonte: autoria própria.

O momento inicial envolverá o aprendizado das vogais (A, E, É, I, O, Ó, U), que são essenciais para a construção das primeiras palavras.

Em seguida, os usuários irão aprender a reproduzir o som das consoantes em sílabas simples, como Ba, Be, Bi, Bo, Bu, Da, De, Di, Do, Du, entre outras, cobrindo uma ampla variedade de combinações fonéticas. Essa etapa é fundamental para que o aluno compreenda a estrutura básica das sílabas e inicie a formação de palavras simples.

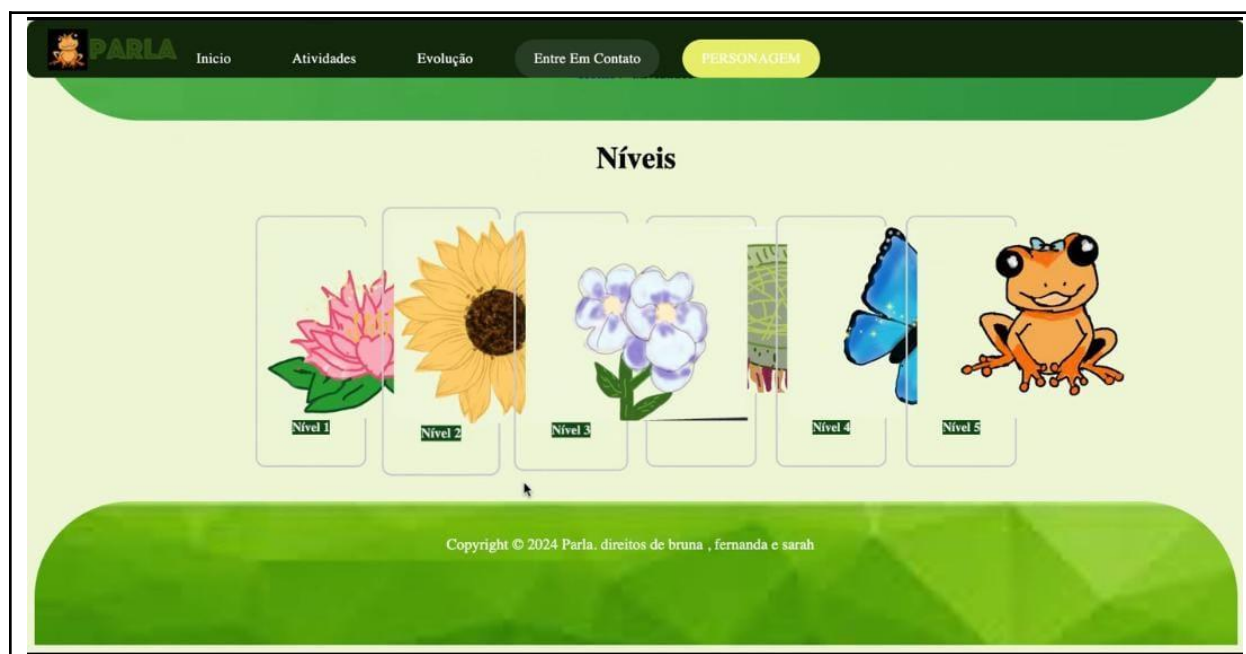
Posteriormente, serão introduzidas sílabas complexas, como Cha, Che, Chi, Cho, Chu, Lha, Lhe, Lhi, Lho, Lhu, além de combinações como Gua, Gue, Gui, Guo. Essas combinações ampliam o repertório linguístico do aluno, permitindo que ele entenda e pronuncie palavras mais elaboradas. A fase também abrange sílabas com encontros consonantais, como Bra, Bre, Bri, Bro, Bru, e sílabas nasais, como Ar, Er, Ir, Or, Ur, An, En, In, On, Un.

Para complementar o aprendizado, serão incluídas outras combinações e variações de sílabas que ajudam a diversificar o conhecimento fonético do aluno. Exemplos dessas combinações são encontros consonantais mais avançados, como Psa, Pse, Psi, Psu, e outras formas como Tcha, Tche, Tchi, Tcho, Tchu, que permitem um domínio mais completo das estruturas fonológicas do idioma.

Na tela de atividade, há recursos para apoiar a pronúncia do usuário. Elas incluem uma imagem GIF com movimentos orofaciais, para auxiliar na aprendizagem de leitura labial, uma imagem demonstrativa da onda sonora, para ajudar na entonação, e uma caixa de hertz que aparece após a pronúncia do usuário, funcionando como *feedback* da pronúncia realizada pelo usuário.

As atividades dessas fases será subdivida em níveis menores de atuação do usuário, conforme apresentado na imagem a seguir.

Figura 4: Parla – níveis



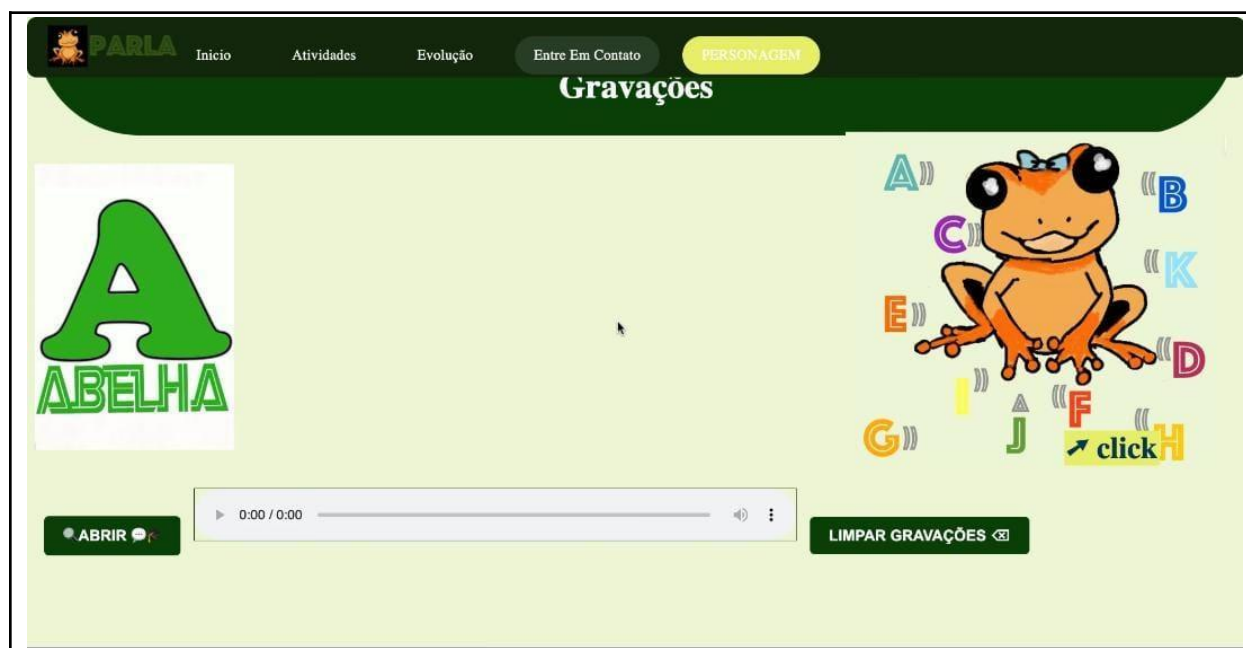
Fonte: autoria própria.

As atividades funcionarão da seguinte forma. Primeiramente, a tela exibirá a imagem do conteúdo que está sendo aprendido no nível, seja letra, seja sílaba. Com base nisso, o usuário tentará pronunciar o que lhe é mostrado; a inteligência artificial (IA) capturará e analisará o áudio em questão de segundos. Se a pronúncia estiver correta, o usuário avançará de fase; caso contrário, será encaminhado para a tela de análise de erros para tentar novamente.

É importante notar que tanto os erros quanto os acertos são salvos no banco de dados para uso futuro na tela de evolução, onde são apresentados os dados do usuário e indicadas atividades extras para as etapas em que o usuário teve maior dificuldade.

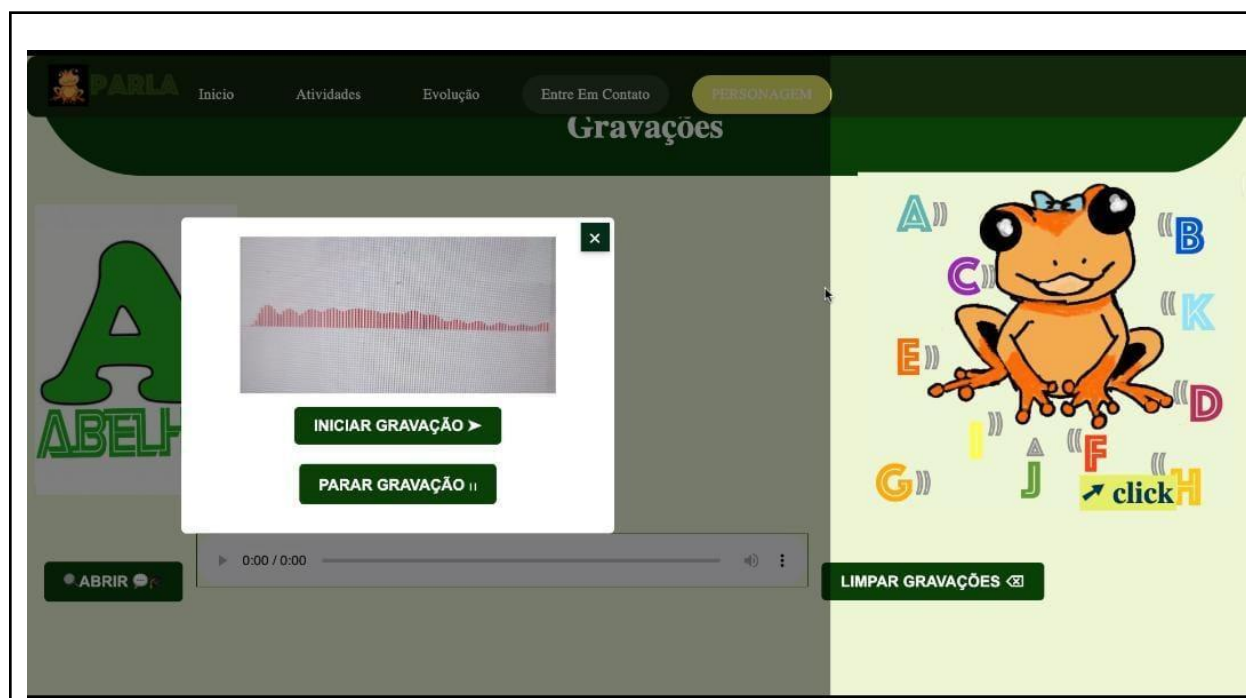
As imagens a seguir ilustram o funcionamento das atividades.

Figura 5: Parla – atividade (1)



Fonte: autoria própria.

Figura 6: Parla – atividade (2)

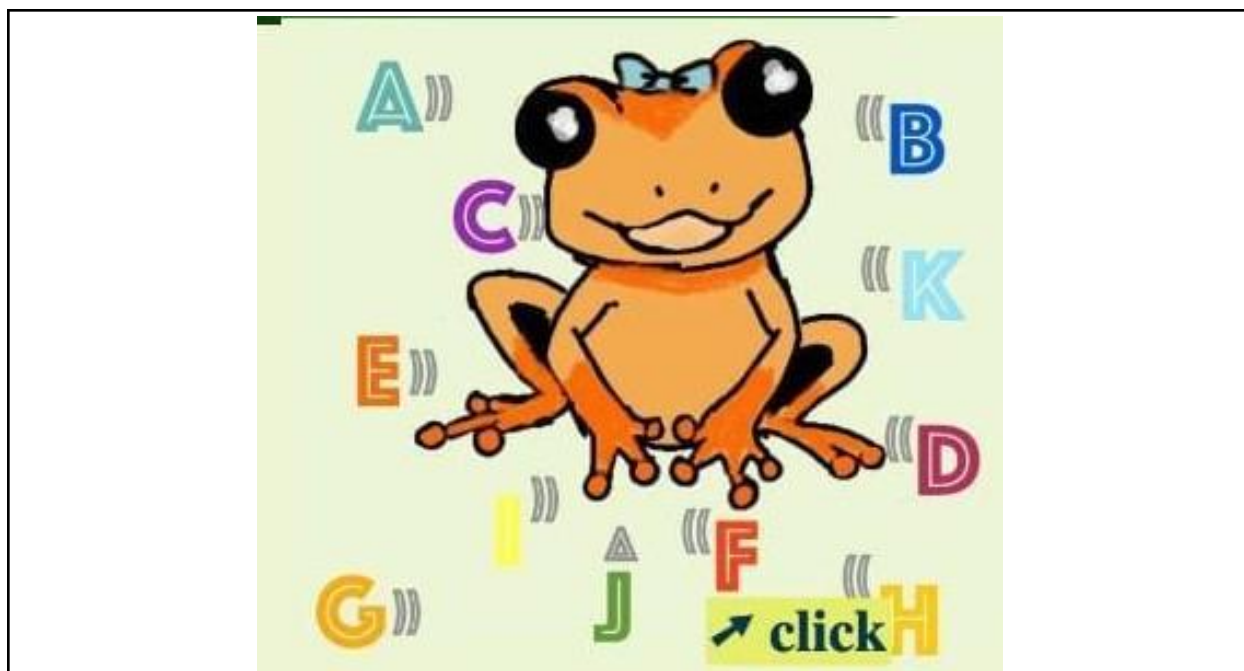


Fonte: autoria própria.

3.4. Identidade visual

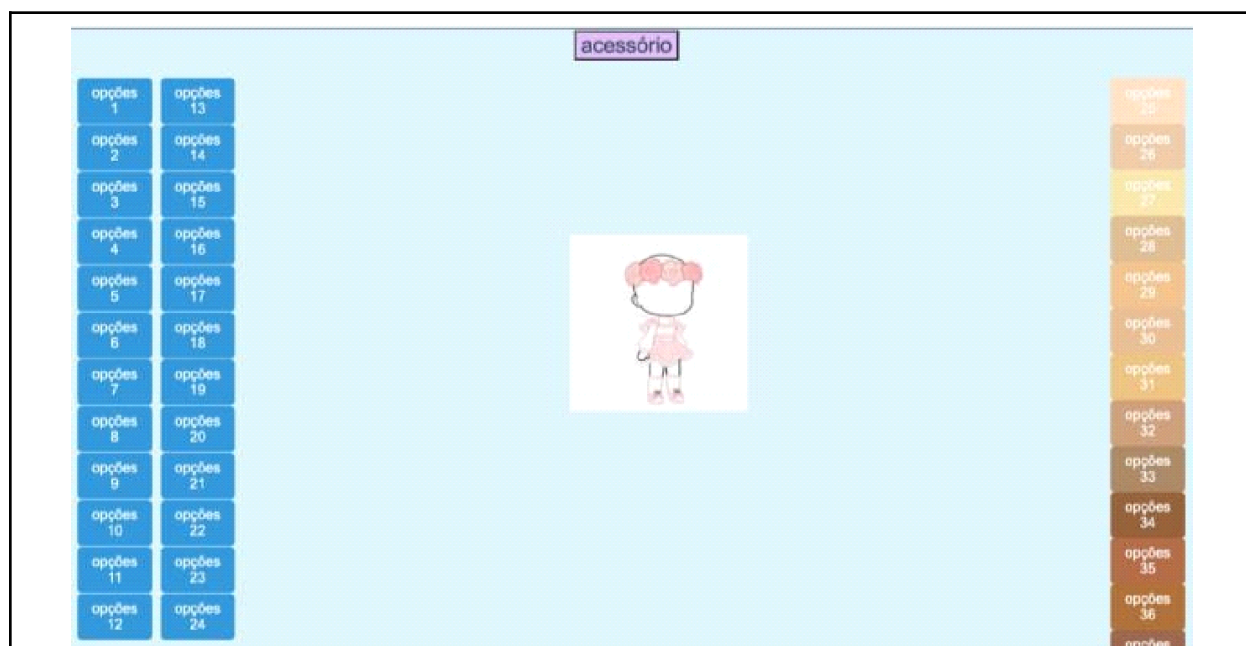
Ao concluir a base do protótipo, elaboramos uma identidade visual para nossa plataforma, incluindo a criação da mascote Parlinha, um sapo pingo-de-ouro (*Brachycephalus ephippium*), que produz sons, mas não é capaz de se ouvir. Essa mascote atua como assistente do usuário na plataforma.

Figura 7: Mascote - Parlinha



Fonte: autoria própria.

A plataforma possui 17 páginas desenvolvidas, incluindo: a tela de apresentação do site, a tela de cadastro, a tela de login, a tela de desenvolvimento do personagem interativo (demonstrativa), a tela inicial do usuário, a tela de fases (que dá acesso a três níveis: básico, intermediário e avançado), com atualmente 7 telas de níveis disponíveis para teste (A, E, É, I, O, Ó, U), a tela de falha, a tela de evolução (demonstrativa) e a tela de contato (demonstrativa).

Figura 8: Parla - criação do personagem

Fonte: autoria própria.

Figura 9: Parla – tela

Fonte: autoria própria.

3.5 Tela de evolução

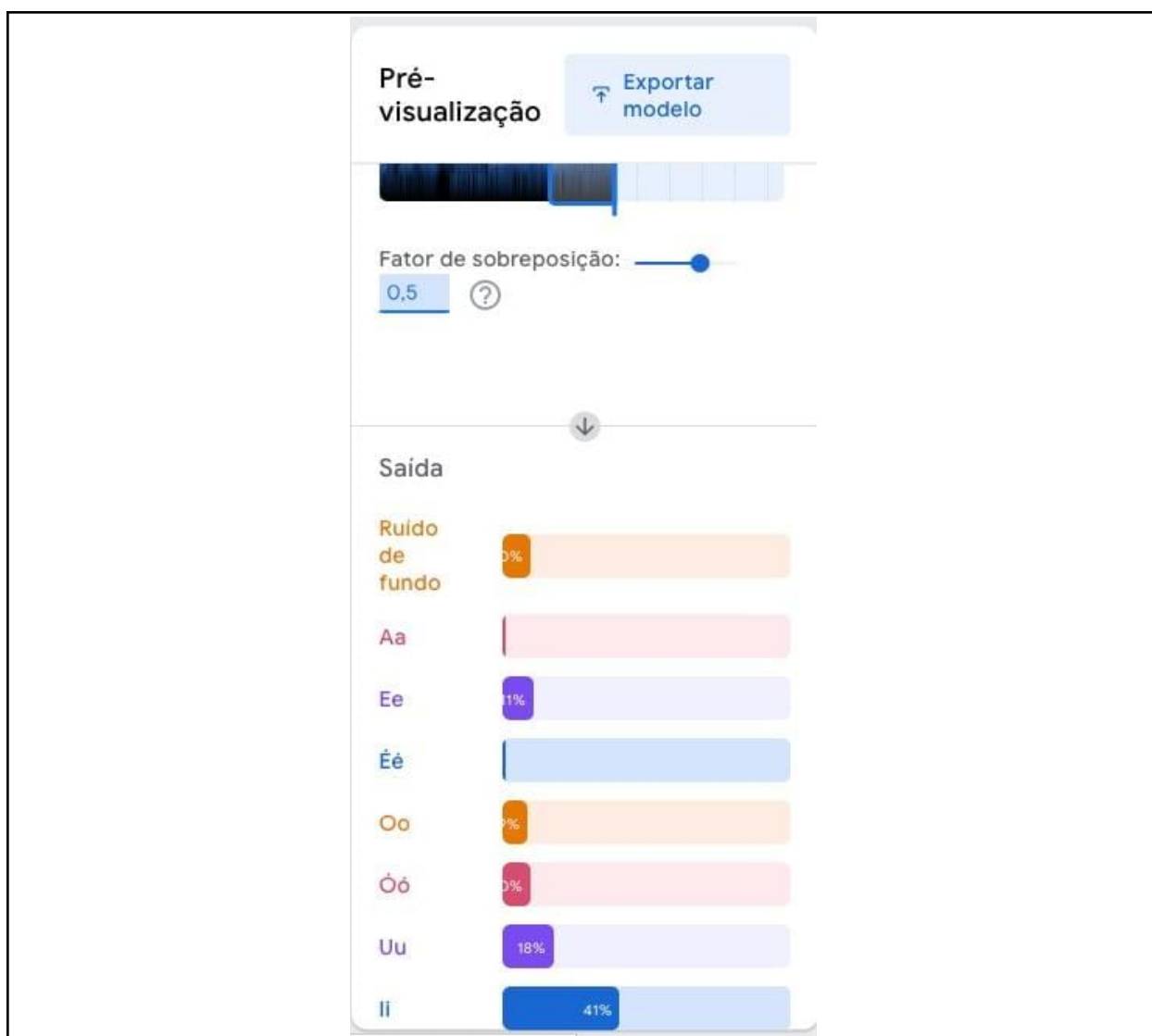
A tela de evolução armazena a frequência das atividades, os dias em que foram realizadas, o número de erros e acertos, e a quantidade de tentativas. Com essas informações, é gerado um gráfico de aprendizado do aluno e um boletim que pode ser impresso ou enviado por e-mail.

Na tela de evolução, também são oferecidas atividades extras para recuperação nas áreas em que o usuário teve o maior número de erros, ajudando assim a melhorar os pontos de maior dificuldade. O boletim e as atividades de recuperação são disponibilizados mensalmente, permitindo o acompanhamento do usuário – ou de seus responsáveis.

4. RESULTADOS

Até ao momento, desenvolvemos a primeira fase da plataforma Parla, que corresponde ao ensino das vogais (A, E, É, I, O, Ó, U). Para isso, utilizamos um banco de dados com mais de 77 mil áudios gravados, em total de 42 horas e 7 minutos de gravações. Durante a fase de testes, que atualmente se encontra no estágio de protótipo, realizamos testes iniciais com usuários, obtendo um índice de acerto de 80%.

Figura 10: Gravação de áudio (*Teachable Machine*)



Fonte: autoria própria.

Os principais erros foram identificados no reconhecimento dos sons É e Ó, que eram confundidos com E e O, respectivamente. Após ajustes e calibrações na inteligência artificial (IA), o índice de acerto aumentou para 90%, e, com ajustes adicionais, conseguimos alcançar quase 100% de precisão.

5. CONCLUSÃO

A análise dos resultados indica que o Parla pode fazer uma diferença significativa para a inclusão educacional e social de pessoas com deficiência auditiva. A plataforma pode não apenas facilitar a comunicação com o mundo ouvinte, mas também contribuir para uma melhor integração dessas pessoas em ambientes escolares e profissionais. No entanto, reconhecemos que o estudo tem suas limitações, como a falta de testes com o nosso público-alvo e a necessidade de validação contínua da eficácia a longo prazo.

Com base nos resultados, pretendemos concluir a expansão da plataforma para incluir mais recursos interativos e a realização de testes com um público mais diversificado. A integração do Parla com programas educacionais existentes pode ampliar seu impacto e garantir que mais pessoas se beneficiem da tecnologia. Além disso, estamos no processo de desenvolvimento de novas funcionalidades que respondam às necessidades emergentes dos usuários.

Para os próximos passos do projeto, implementaremos as telas que atualmente são apenas demonstrativas e adicionar mais sete telas para tornar o site mais completo. As novas telas incluem a tela de recuperação, a tela para recuperação de senha, a tela dos pacotes Premium, a tela exclusiva de atividades Premium, a tela de perfil do usuário, a tela de recados/mensagens do sistema e a tela de medalhas e conquistas do Parla.

Após isso, para a próxima fase de testes, buscaremos autorizações para realizar testes com pessoas surdas e com deficiência auditiva grave ou profunda em escolas e instituições de nossa cidade.

REFERÊNCIAS

ALESP. Dia Internacional da Linguagem de Sinais procura promover a inclusão de pessoas surdas. 23 set. 2021. Disponível em: <https://www.al.sp.gov.br/noticia/?23/09/2021/dia-internacional-da-linguagem-de-sinais-procura-promover-a-inclusao-de-pessoas-surdas->. Acesso em: 29 abr. 2024.

INEP. Confira o panorama dos surdos na educação brasileira. 26 ago. 2023. Disponível em: www.gov.br/inep/pt-br/assuntos/noticias/censo-escolar/confira-o-panorama-dos-surdos-na-educacao-brasileira. Acesso em: 29 abr. 2024.

LEMOS, Simone. Mais de 10 milhões de brasileiros apresentam algum grau de surdez. **Jornal da USP**. 21 ago. 2023. Disponível em: <https://jornal.usp.br/?p=674626> . Acesso em: 29 abr. 2024.

MAGNO, Rodrigo. As dificuldades da pessoa surda na sociedade brasileira. **Jusbrasil**. 2024. Disponível em: www.jusbrasil.com.br/artigos/as-dificuldades-da-pessoa-surda-na-sociedade-brasileira/1176514129 . Acesso em: 29 abr. 2024.

ONU. Sobre o nosso trabalho para alcançar os Objetivos de Desenvolvimento Sustentável no Brasil. 2024. Disponível em: <https://brasil.un.org/pt-br/sdgs> . Acesso em: 29 abr. 2024.

O que é *machine learning*. Disponível em: O que é Machine Learning? Tipos e usos | Google Cloud. Acesso em: 29 abr. 2024.

PINHEIRO, Ângela Maria Vieira *et al.* Reading strategies for the profoundly deaf Libras users: benefits of speech and lip reading for strengthening linguistic skills. **Estudos de Psicologia**. Campinas, v.37. Disponível em: <http://dx.doi.org/10.1590/1982-0275202037e190003> . Acesso em: 29 abr. 2024.

TOFFOLO, Andreia Chagas Rocha *et al.* Os benefícios da oralização e da leitura labial no desempenho de leitura de surdos profundos usuários da Libras. **Revista Brasileira de Educação**. v.22, n.71, p.1-24, 2017. Disponível em: <http://dx.doi.org/10.1590/S1413-24782017227165> . Acesso em: 29 abr. 2024