

IFTranscript - Aplicação Móvel Acessível para Transcrição de Áudio

**Trabalho de Conclusão do Curso de
Tecnologia em Sistemas para Internet**

**Mateus Vagner Guedes de Almeida
Orientador(a): Silvia De Castro Bertagnolli**

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS)
Campus Porto Alegre
Av Cel Vicente, 281, Porto Alegre – RS – Brasil

mateus.vagner@gmail.com, silvia.bertagnolli@poa.ifrs.edu.br

Resumo. *O uso de métodos de ensino tradicionais apoiados em recursos visuais e auditivos, cria obstáculos para o aprendizado dos estudantes com deficiência. Alunos com deficiência visual, auditiva ou motora enfrentam desafios específicos, como a dificuldade de acompanhamento das aulas e a realização das anotações vinculadas às explicações dos professores. Com isso, a proposta desse trabalho visa diminuir as barreiras comunicacionais enfrentadas por alunos com deficiência no contexto educacional, por meio do desenvolvimento de uma aplicação móvel gratuita e acessível de transcrição de áudio. O aplicativo IFTranscript tem como finalidade fornecer uma ferramenta que permita tornar o conteúdo acessível a todos os alunos, sejam eles portadores de deficiência ou não.*

1. Introdução

A inclusão de alunos com deficiência no ambiente educacional enfrenta desafios significativos devido à escassez de materiais adaptados e métodos específicos de ensino, o que limita a capacidade desses estudantes de acompanhar o conteúdo das aulas. A necessidade de realizar anotações pessoais pode ser especialmente desafiadora. Estudantes com deficiência auditiva, por exemplo, podem se beneficiar de transcrições para acessar o conteúdo, enquanto aqueles com deficiência motora podem focar mais na exposição dos temas ao eliminarem a exigência de anotações manuais.

O objetivo principal deste trabalho é desenvolver uma aplicação móvel de transcrição de áudio, com foco em acessibilidade e gratuidade, para atender às necessidades de alunos com deficiência no contexto educacional. A transcrição pode tornar o conteúdo mais acessível, promovendo a inclusão desses alunos e oferecendo uma solução que permita acompanhar as aulas com maior autonomia.

A proposta desta aplicação surge em resposta à escassez de opções gratuitas e completas, uma vez que muitas alternativas disponíveis são pagas ou oferecem versões gratuitas com funcionalidades limitadas. Essa situação pode comprometer a igualdade de oportunidades e a autonomia dos alunos com deficiência. A solução consiste em um aplicativo dedicado à transcrição de áudio, com ênfase na acessibilidade e gratuidade, capaz de oferecer transcrição de gravações de aulas, podcasts e outras fontes de áudio relevantes.

Com a implementação de uma aplicação móvel acessível e gratuita para transcrição de áudio, espera-se contribuir para a redução das barreiras enfrentadas pelos alunos com deficiência no ambiente educacional. Para entender completamente o impacto dessa iniciativa, é essencial discutir a base legal que sustenta esses direitos, como a Lei Brasileira de Inclusão (LBI), além de explorar o conceito de Tecnologia Assistiva e suas aplicações na educação, analisando como essas tecnologias estão sendo integradas no contexto educacional para apoiar estudantes com deficiência.

2. Tecnologia Assistiva

A Lei Brasileira de Inclusão da Pessoa com Deficiência, também chamada de Estatuto da Pessoa com Deficiência, foi estabelecida com o objetivo de assegurar e promover, em condições de igualdade, o exercício dos direitos e das liberdades fundamentais pelas pessoas com deficiência, visando a sua inclusão social e cidadã. Segundo a mesma Lei, uma pessoa com deficiência é definida como aquela que apresenta um impedimento de longo prazo de natureza física, mental, intelectual ou sensorial. Este impedimento, quando combinado com uma ou mais barreiras ambientais ou sociais, pode resultar em restrições à sua participação plena e efetiva na sociedade em igualdade de condições com as demais pessoas. [Brasil 2015].

Além disso, a Lei apresenta definições essenciais para sua aplicação, destacando-se as seguintes, que são cruciais para a realização deste estudo:

- a) **Acessibilidade:** é definida como a possibilidade e condição de alcance, percepção e entendimento para a utilização com segurança e autonomia de espaços, mobiliários, equipamentos urbanos, edificações, transportes, informações e comunicações, inclusive seus sistemas e tecnologias, por pessoas com deficiência ou com mobilidade reduzida;
- b) **Barreiras:** são obstáculos que limitam ou impedem a participação das pessoas com deficiência na sociedade em igualdade de condições com as demais. Essas barreiras podem ser de diferentes tipos, como urbanísticas, arquitetônicas, nos transportes, nas comunicações e na informação, atitudinais e tecnológicas;
- c) **Tecnologia Assistiva ou Ajuda Técnica:** refere-se a produtos, equipamentos, dispositivos, metodologias, estratégias, práticas e serviços que visam proporcionar ou ampliar habilidades funcionais de pessoas com deficiência. O objetivo é promover a autonomia, a inclusão e a qualidade de vida dessas pessoas, facilitando a realização de atividades do dia a dia e a participação em diversos aspectos da vida social.

Nesse sentido, a Tecnologia Assistiva (TA), conforme definição dada pelo Comitê de Ajudas Técnicas (CAT) da Secretaria Especial de Direitos Humanos (SEDH), é uma área de conhecimento multidisciplinar que abrange produtos, recursos, estratégias, práticas, processos, métodos e serviços destinados a pessoas com deficiência, mobilidade reduzida e idosos, visando promover a autonomia, a qualidade de vida e a inclusão social em diferentes espaços [CAT 2007].

O conceito de TA começou a ser pensado no país através do Comitê de Ajudas Técnicas, estabelecido pela Secretaria de Direitos Humanos da Presidência da República, em 2006 [Galvão Filho 2013b].

A escolha do termo “Tecnologia Assistiva” em vez de “Ajudas Técnicas” ou “Tecnologia de Apoio” reflete uma preferência consolidada no meio acadêmico, em organizações de defesa dos direitos das pessoas com deficiência, setores governamentais e no mercado de produtos associados. Os principais motivos são a especificidade do termo, que melhor representa o conceito e o distingue de outras expressões genéricas, e a compatibilidade com a denominação de uma área de conhecimento. Este alinhamento conceitual também visa revisar a nomenclatura em legislações e documentos oficiais, consolidando a TA como uma disciplina oficialmente reconhecida [Galvão Filho 2013a].

No contexto educacional, alunos com deficiência frequentemente encaram obstáculos na educação, incluindo limitações na comunicação e mobilidade. Essas restrições podem isolar o aluno de interações, tanto com o ambiente quanto com outras pessoas, impactando negativamente seu desenvolvimento cognitivo e seu processo de aprendizagem. Por conta disso, a Tecnologia Assistiva desempenha um papel fundamental em contextos educacionais inclusivos onde seu uso é essencial no ambiente de ensino para garantir uma educação justa e acessível a todos [Galvão Filho 2012].

3. Transcrição de Áudio

Diversos aspectos merecem atenção especial para compreender a implementação e impacto no ambiente educacional da transcrição de áudio. Entre eles, destacam-se os desafios técnicos da transcrição de áudio, que incluem questões relacionadas à precisão, acessibilidade e custo das soluções disponíveis. Além disso, a adoção desse tipo de solução na educação deve ser discutida, buscando evidenciar se essa tecnologia pode transformar a experiência de aprendizagem para estudantes com deficiência, promovendo maior inclusão e autonomia.

3.1. Desafios da Transcrição de Áudio

Um dos principais desafios do reconhecimento de voz é a variação de sotaques intralinguísticos que complicam a precisão na interpretação da pronúncia, dialetos, sotaque e entonação individuais. Além disso, a qualidade do áudio também é um obstáculo importante, visto que a presença de ruído no local, como sons ou falas de fundo durante as gravações, pode afetar a capacidade dos sistemas em interpretar corretamente a fala humana. Esses desafios destacam a necessidade contínua de Sistemas de Reconhecimento Automático de fala (ASR, Automatic Speech Recognition no inglês), visando melhorar a precisão, inclusão e capacidade de lidar com as variações naturais da fala humana [Basak *et al.* 2022].

Além dos desafios técnicos, a privacidade de dados é um problema para sistemas de reconhecimento de fala. A fala é uma forma privada de comunicação, e muitos se opõem a escutas e gravações. Usuários de serviços baseados em fala devem conceder acesso completo às suas gravações, arriscando a extração não autorizada de informações pessoais (gênero, etnia, estado emocional) e manipulação de gravações para criar conteúdo falsificado. A coleta de dados de voz sem consentimento pode identificar indivíduos em contextos não previstos, como em plataformas de mídia social. [Pathak *et al.* 2013].

3.2. Transcrição de Áudio na Educação

Estudos mostram que a transcrição de fala impacta positivamente a compreensão dos alunos com deficiência auditiva em cursos de educação inclusiva [Irdamurni *et al.* 2020]. Essas ferramentas são cruciais não apenas para esses estudantes, mas também para aqueles com dificuldades de aprendizagem. Podem ser utilizadas em qualquer idioma, auxiliando alunos multilíngues e reduzindo a ansiedade associada à leitura e escrita em novas línguas [Connelly and Doyle 2023].

Alunos com dificuldades de aprendizagem também melhoram seu desempenho na escrita ao utilizar ferramentas de transcrição de áudio, embora a imprecisão da tecnologia ainda seja um desafio [Matre and Cameron 2022]. As aplicações de fala para texto aumentam a acessibilidade dos recursos educacionais e enriquecem as experiências de aprendizagem, beneficiando alunos com deficiência e falantes não nativos [Yashina *et al.* 2022]. A transcrição de áudio e o avanço da inteligência artificial foram especialmente importantes durante a pandemia de COVID-19, melhorando a experiência de aprendizagem dos alunos [Abid *et al.* 2023].

No ensino médio, a tecnologia de transcrição de áudio beneficia estudantes durante tarefas de escrita. Estudantes com dificuldades de aprendizagem utilizam essas ferramentas com frequência, e seus professores têm percepções positivas sobre elas [Levine *et al.* 2023].

Por fim, o uso da tecnologia de voz para texto para aprimorar a escrita de crianças com necessidades educacionais especiais e deficiências apresenta resultados positivos tanto na quantidade quanto na qualidade do texto escrito à mão, além de aumentar a autoestima das crianças [Kambouri *et al.* 2023].

4. Trabalhos relacionados

Na análise dos trabalhos relacionados à transcrição de áudio, é fundamental explorar as diversas ferramentas e recursos que estão disponíveis atualmente. Neste contexto, serão discutidas as API's (Application Programming Interface) disponíveis para transcrição de áudio, que oferecem interfaces para a integração dessa tecnologia em diferentes aplicações, além das aplicações de transcrição de áudio e as características e limitações de cada uma delas. Por fim, as bibliotecas disponíveis para transcrição de áudio, que proporcionam aos desenvolvedores mais autonomia na implementação de soluções customizadas.

4.1. API's disponíveis para transcrição de áudio

Para a funcionalidade de transcrição de áudio é necessário criar um *backend* capaz de receber um arquivo de áudio ou vídeo e processar esse arquivo para gerar a transcrição. Foram pesquisados serviços via API's disponíveis no mercado que realizem essa funcionalidade. Nesse contexto, várias API's de transcrição de áudio estão disponíveis no mercado, sendo que cada uma delas também possui suas próprias características, vantagens e limitações. Neste trabalho, realizou-se uma análise comparativa das principais API's de transcrição de áudio, com foco em suas funcionalidades, preços e recursos adicionais (Apêndice A apresenta um estudo comparativo das ferramentas):

- a) **OpenAI Whisper**¹: possui preço de \$0,006 por minuto e suporte até 60 idiomas. Sua acurácia de transcrição pode variar dependendo do idioma utilizado. Uma característica interessante é sua capacidade de traduzir a transcrição para o Inglês;
- b) **Google Speech-to-Text**²: oferece uma opção gratuita com suporte para 125 idiomas, limitada a 60 minutos por mês. Além disso, fornece até 300 dólares em créditos para teste de outras ferramentas. Ela se destaca pela capacidade de transcrição em tempo real;
- c) **Amazon Transcribe**³: também oferece uma opção gratuita até 60 minutos por mês durante 12 meses, com suporte para 155 idiomas. Ela também oferece transcrição em tempo real;
- d) **Azure Speech-to-Text**⁴: apresenta uma opção gratuita de até 5 horas por mês, suportando 156 idiomas. Destaca-se pelo reconhecimento de locutor e modelos customizáveis de acordo com a voz do locutor, além de oferecer transcrição em tempo real;
- e) **IBM Watson Speech-to-Text**⁵: oferece uma opção gratuita até 500 minutos por mês, com suporte a mais de 14 idiomas. Destaca-se pelo reconhecimento de locutor e modelos customizáveis de acordo com a voz do locutor.

Além das grandes empresas de tecnologia pesquisadas para o acesso às APIs de transcrição de áudio, também foram procurados outros fornecedores de menor porte que disponibilizavam suas próprias APIs, tais como **Transkriptor**⁶, **Deepgram**⁷, **AssemblyAI**⁸ e **Fireflies**⁹.

4.2. Aplicações de transcrição de áudio

Atualmente existem diversas ferramentas para transcrição de áudio disponíveis para os usuários. Cada uma dessas ferramentas oferece uma combinação única de recursos, preços e limitações. Nesta análise, realizada para este trabalho de conclusão de curso, foram examinadas essas ferramentas, destacando seus pontos fortes, limitações e características distintivas (Apêndice B sintetiza um comparativo das aplicações pesquisadas).

- a) **Notta**¹⁰: se destaca como uma opção gratuita disponível nas plataformas Web, Android e iOS. Ela oferece suporte para 42 idiomas, resumo da transcrição e a capacidade de exportar transcrições. No entanto, seu uso é limitado a 120 minutos totais por mês, com um máximo de 5 minutos por áudio;
- b) **Transkriptor**¹¹: é compatível com mais de 100 idiomas, porém é necessário investir uma taxa mensal de \$4,99. Tem capacidade de edição de transcrições e criação de legendas, porém, seu limite de 5 horas por mês. Ela também está disponível nas plataformas Web, Android e iOS;

¹ Disponível em: <https://openai.com/index/whisper>.

² Disponível em: <https://cloud.google.com/speech-to-text>.

³ Disponível em: <https://aws.amazon.com/pt/transcribe>.

⁴ Disponível em: <https://azure.microsoft.com/en-us/products/ai-services/speech-to-text>.

⁵ Disponível em: <https://www.ibm.com/products/speech-to-text>.

⁶ Disponível em: <https://transkriptor.com>.

⁷ Disponível em: <https://deepgram.com>.

⁸ Disponível em: <https://www.assemblyai.com>.

⁹ Disponível em: <https://fireflies.ai>.

¹⁰ Disponível em: <https://www.notta.ai/pt>.

¹¹ Disponível em: <https://transkriptor.com>.

- c) **Happyscribe**¹²: possui um custo mensal de \$17,00, sendo que ela está disponível somente para o ambiente Web. Em contrapartida ela oferece suporte para 66 idiomas, exportação e edição de transcrições, além da criação de legendas. No entanto, seu limite de 120 minutos por áudio pode ser uma desvantagem para usuários com altas demandas de transcrição;
- d) **Otter**¹³: é uma opção gratuita disponível para as plataformas Web, Android e iOS. Oferecendo resumo da transcrição e exportação, com limite de 300 minutos por mês e 30 minutos por áudio. Sua maior restrição é que a ferramenta é limitada ao idioma inglês;
- e) **Fireflies**¹⁴: é outra opção gratuita, oferecendo suporte para mais de 69 idiomas, com um limite de 800 minutos de armazenamento, além de disponibilizar resumo da transcrição;
- f) **Rev**¹⁵: pode ser usada após realizar o investimento de \$1,50 por minuto, ela disponibiliza transcrições em 22 idiomas, com um limite de 5 minutos por áudio. O serviço de resumo da transcrição é pago separadamente e está disponível apenas em inglês;
- g) **Sonix**¹⁶: pode ser usada mediante o investimento de uma taxa de \$10,00 por hora pré-paga. Ela oferece suporte para mais de 49 idiomas e inclui a capacidade de exportar transcrições e criar legendas. Seu modelo pré-pago pode ser vantajoso para usuários que desejam maior flexibilidade em relação ao uso.

4.3. Bibliotecas disponíveis para transcrição de áudio

Considerando que nenhuma das API's e aplicações pesquisadas é totalmente gratuita e sem limite de uso, foi realizada uma pesquisa de ferramentas *open source* para oferecer uma solução gratuita e sem limitações. A escolha da biblioteca adequada é crucial para garantir a precisão e eficiência da transcrição. Neste trabalho, realizou-se uma análise comparativa das bibliotecas de reconhecimento de fala encontradas, considerando fatores como idiomas suportados, data da última atualização, popularidade e recursos adicionais (Apêndice C apresenta algumas dessas bibliotecas):

- a) **Whisper**¹⁷: possui ampla gama de idiomas suportados, sua robusta base de código e a capacidade de traduzir transcrições para o inglês. Sua última atualização em 17 de novembro de 2023 e uma base de usuários significativa, com 7 mil *forks* na plataforma Github, destacam sua relevância e popularidade;
- b) **Kaldi**¹⁸: é focado no inglês, destaca-se por seu enfoque em reconhecimento de locutores. Uma atualização recente em 30 de abril de 2024 e pelos 5,3 mil *forks* mostra que é um *framework* popular e atual. No entanto, sua limitação linguística pode restringir sua aplicabilidade em ambientes multilíngues;
- c) **Mozilla's DeepSpeech**¹⁹: também suporta apenas o inglês, apresenta uma base de código sólida e uma comunidade ativa, com 3,9 mil *forks* e uma última atualização foi em 10 de dezembro de 2020;

¹² Disponível em: <https://www.happyscribe.com/pt>.

¹³ Disponível em: <https://otter.ai/>.

¹⁴ Disponível em: <https://fireflies.ai>.

¹⁵ Disponível em: <https://www.rev.ai>.

¹⁶ Disponível em: <https://sonix.ai/pt>.

¹⁷ Disponível em: <https://github.com/openai/whisper>

¹⁸ Disponível em: <https://github.com/kaldi-asr/kaldi>

¹⁹ Disponível em: <https://github.com/mozilla/DeepSpeech>

- d) **Flashlight**²⁰: foca em transcrever áudio diretamente para caracteres, sem o uso de alinhamento temporal. Isso o torna uma ferramenta limitada para o uso no presente trabalho. Sua última atualização em 19 de março de 2022 e possui aproximadamente 1,5 mil *forks*.

Após analisar as diferentes APIs, bibliotecas e aplicações, é proposta a ferramenta **IFTranscript**, que se diferencia das demais opções de transcrição de áudio disponíveis no mercado por seu compromisso com a acessibilidade e a gratuidade. A aplicação é projetada desde o início para atender às necessidades específicas de pessoas com deficiência, com o objetivo de oferecer uma experiência de uso inclusiva e sem barreiras. Com uma interface simplificada e acessível, aliada à gratuidade e ausência de limitações de uso, essa aplicação se torna uma escolha relevante na promoção da inclusão digital e da igualdade de acesso à transcrição de áudio.

5. Metodologia

No início do trabalho foi empregada a metodologia de pesquisa bibliográfica exploratória para esclarecer a definição de acessibilidade, além de buscar estudos sobre as tecnologias assistivas e sua empregabilidade no contexto educacional. A partir disso, procurou-se encontrar estudos sobre o uso da transcrição de áudio como ferramenta de auxílio para estudantes com deficiência

Após a pesquisa bibliográfica, foram feitas análises de outras aplicações de transcrição de áudio existentes, levantando suas principais características, custos e limitações. A partir dessa análise foi possível definir as principais características da aplicação em desenvolvimento no presente trabalho. Além disso, APIs e bibliotecas foram analisadas com intuito de verificar a viabilidade da criação de um sistema completo que englobe tanto aplicação móvel quanto o *backend* próprio.

Com base nos resultados da pesquisa bibliográfica e da análise das aplicações, APIs e bibliotecas, foi realizada a modelagem inicial do sistema. Inicialmente, foi definido o diagrama de casos de uso, seguido pela criação do diagrama de classes. Com esses dois diagramas, foi possível esboçar um diagrama inicial de funcionalidades, assim como idealizar um fluxo preliminar do usuário dentro do aplicativo, a partir do qual, algumas telas foram desenhadas (Apêndice E).

Com as principais características do aplicativo definidas, foi possível elaborar um modelo do sistema, que inclui o aplicativo Android juntamente com o *backend* responsável pela autenticação de usuários, cadastro de novos usuários, transcrição dos áudios e, finalmente, a criação de resumos (Apêndice F).

Por fim, os testes do aplicativo foram realizados por dois usuários com perfis distintos: uma pessoa cega e outra sem deficiência, especialista em acessibilidade. Ambos tiveram total liberdade para explorar o aplicativo, testando suas funcionalidades de acordo com suas próprias preferências e necessidades.

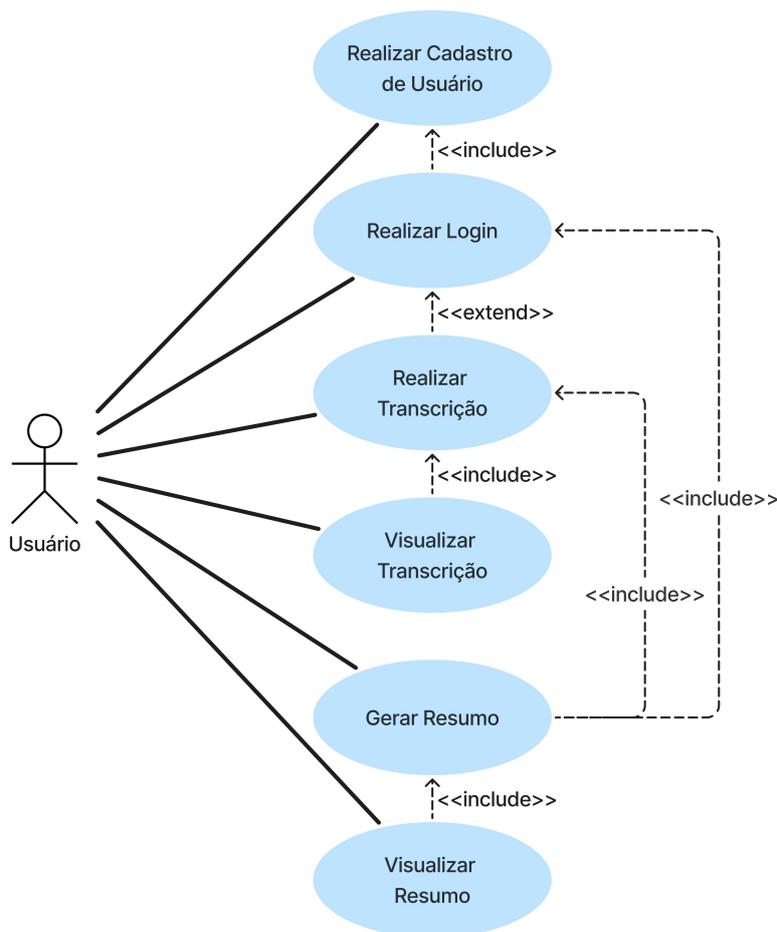
6. Modelagem e definição do Sistema

O processo de definição do sistema começou com a análise de trabalhos relacionados, a fim de identificar as funcionalidades essenciais para sistemas de transcrição de áudio. Em

²⁰ Disponível em: <https://github.com/flashlight/flashlight>

seguida, foi realizada a modelagem do sistema por meio de um diagrama de casos de uso, que determina as principais funcionalidades da aplicação. A funcionalidade central é a geração de transcrições de áudio de forma simples e rápida, sem a necessidade de cadastro do usuário. No entanto, para acessar o histórico de transcrições, o usuário terá a opção de se cadastrar. Além disso, a aplicação permitirá a geração de resumos das transcrições para os usuários cadastrados. A Figura 1 mostra o Diagrama de Caso de Uso, contendo os requisitos funcionais do sistema.

Figura 1. Diagrama de Caso de Uso.

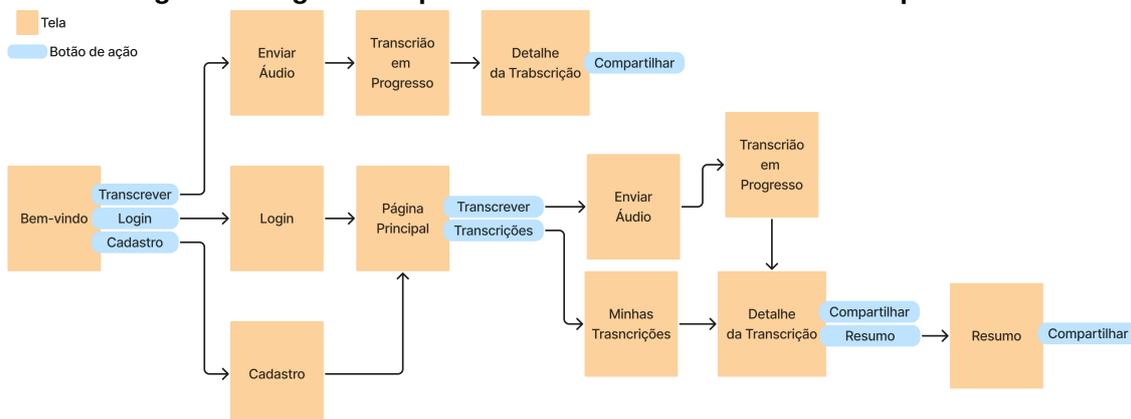


Fonte: Elaborado pelo autor (2024).

Além do diagrama de casos de uso, foi elaborado o diagrama de classes, o qual tem como finalidade definir as relações entre as classes que fazem parte do sistema, ele encontra-se descrito e ilustrado no Apêndice D.

A partir da premissa que o usuário deve ser capaz de gerar uma transcrição de maneira simples e rápida, e considerando-se o diagrama de caso de uso definido previamente, define-se primeiramente a jornada do usuário dentro do aplicativo. A Figura 2 mostra um diagrama esquemático do fluxo de uso do aplicativo pelo usuário, determinando quais telas e botões poderão ser utilizados.

Figura 2. Diagrama esquemático do fluxo de uso dentro do aplicativo.



Fonte: Elaborado pelo autor (2024).

Com o fluxo de uso da aplicação definido, bem como a quantidade de telas e o requisito não funcional de acessibilidade, foram elaborados protótipos das telas. Esses protótipos e suas descrições estão apresentados no Apêndice E.

Destaca-se como relevante para o sistema proposto o requisito não funcional de acessibilidade, visto que ele representa uma restrição necessária ao aplicativo, e que não está diretamente relacionado com as funcionalidades previstas. Para o projeto do aplicativo foi necessário seguir as diretrizes da WCAG (*Web Content Accessibility Guidelines*) e usar algumas recomendações vinculadas à tecnologia selecionada para o desenvolvimento do aplicativo.

Por fim, foi desenvolvida uma aplicação para a plataforma Android, que se comunica com um *backend* próprio, hospedado localmente e também desenvolvido no presente trabalho. Além disso, fez-se uso da metodologia ágil durante o desenvolvimento do sistema.

6.1. Modelagem do backend

Utilizando bibliotecas e *frameworks*, é possível projetar um *backend* que atenda a todas as necessidades da aplicação, garantindo a segurança dos dados dos usuários em um banco de dados próprio mantido no servidor local. Além disso, não se faz uso de nenhuma API externa para a transcrição de áudio e a geração de resumos.

Para a criação das transcrições, este trabalho optou pelo uso da biblioteca Whisper, devido à sua robustez, popularidade na comunidade de desenvolvedores, documentação abrangente e facilidade de integração. A biblioteca é capaz de processar diversos tipos de arquivos, incluindo MP3, MP4, MPGA, M4A, WAV e WEBM, além de oferecer suporte para transcrição gratuita em mais de 60 idiomas. Para garantir compatibilidade, foi implementada uma ferramenta que converte os arquivos dos formatos aceitos para o formato MP3 antes de serem armazenados no servidor.

A geração de resumos é realizada com o *framework* LangChain²¹ e a ferramenta

²¹Disponível em: <https://www.langchain.com/>

Ollama²². Essas ferramentas permitem a execução de modelos de linguagem de grande escala — *Large Language Models* (LLM) em inglês — de forma totalmente autônoma e local na máquina onde o *backend* estiver hospedado, possibilitando também a customização dos modelos para o uso específico do sistema. Além disso, essas tecnologias possibilitam a implementação de outras técnicas de Inteligência Artificial, permitindo novas funcionalidades no futuro além da geração de resumos. O Apêndice F ilustra e detalha o diagrama esquemático da arquitetura do Sistema.

Todo o *backend* do sistema foi desenvolvido utilizando a linguagem de programação Python²³ juntamente com o *framework* Flask²⁴. Essa escolha se deve pela compatibilidade da linguagem com as bibliotecas utilizadas na transcrição e geração de resumos.

6.2. Modelagem do aplicativo Android

Após a escolha da plataforma de desenvolvimento, foi possível selecionar a linguagem de programação. Para o presente estudo, optou-se pela linguagem *Kotlin*²⁵, recomendada pela Google para o desenvolvimento de aplicações Android. Além disso, utilizou-se o *framework* Jetpack Compose²⁶ para a construção da interface gráfica, também recomendado pela Google devido à sua flexibilidade e capacidade de criar interfaces modernas e dinâmicas.

O maior desafio no desenvolvimento da aplicação foi garantir que ela seja o mais acessível possível, de modo a atender o maior número de usuários possível, incluindo aqueles com deficiência visual, auditiva ou motora. Esse desafio é intensificado pela necessidade de adaptar conteúdos dinâmicos e elementos visuais complexos, além de garantir a compatibilidade com a ampla gama de dispositivos Android. O próprio *framework* Jetpack Compose oferece uma base sólida para a criação de interfaces acessíveis, com foco em boas práticas que facilitam a navegação e a interação por meio de tecnologias assistivas.

A documentação oficial traz três conceitos fundamentais para a acessibilidade [Google 2024]:

- a) **Semântica:** a forma como o significado dos elementos da interface é comunicado aos serviços de acessibilidade. Isso inclui descrições detalhadas, estados e ações que os usuários podem realizar;
- b) **Travessia:** a ordem de navegação que serviços como o TalkBack seguem ao percorrer os elementos da tela. Personalizar essa ordem pode melhorar significativamente a experiência do usuário;
- c) **Ações de acessibilidade:** são as interações que os usuários podem realizar em elementos específicos, como cliques ou rolagem. O app comunica essas ações aos serviços de acessibilidade, permitindo uma navegação mais intuitiva.

Neste trabalho, o foco principal foi a semântica, devido à sua importância em permitir que tecnologias como o *TalkBack* interpretem corretamente os componentes da

²²Disponível em: <https://ollama.com/>

²³Disponível em: <https://www.python.org/>

²⁴Disponível em: <https://flask.palletsprojects.com/>

²⁵Disponível em: <https://kotlinlang.org/>

²⁶Disponível em: <https://developer.android.com/compose>

interface. Durante o desenvolvimento, garantiu-se que os componentes essenciais, como botões, incluíssem descrições claras e intuitivas, informando sua função e a ação correspondente. Todos os ícones e imagens relevantes para a interação do usuário foram descritos de forma adequada, visando uma experiência mais inclusiva e acessível.

Além disso, também foi implementado *feedback* auditivo e tátil para melhorar a experiência de usuários com deficiência visual ou auditiva. Pequenas vibrações e sons foram utilizados para indicar ações de botões, oferecendo uma camada extra de acessibilidade.

O desenvolvimento também envolveu a execução contínua de testes de acessibilidade, utilizando ferramentas como o Accessibility Scanner²⁷, que ajudaram a identificar melhorias potenciais e assegurar que a interface seguisse boas práticas de acessibilidade.

Adicionalmente, nas telas de Detalhe da Transcrição e Resumo, foram integrados botões que permitem o ajuste dinâmico do tamanho da fonte, proporcionando uma leitura mais confortável e acessível para diferentes perfis de usuários. Além disso, foi adicionada uma seção de transcrições favoritas, facilitando o acesso rápido às transcrições preferidas. Para tornar a busca por transcrições específicas ainda mais ágil, uma barra de pesquisa também foi incluída. O aplicativo foi projetado para funcionar nos modos claro e escuro do sistema, garantindo uma experiência visual consistente em ambas as configurações.

Por fim, para garantir que a acessibilidade do aplicativo seja efetiva, recomenda-se a realização de testes manuais com usuários de diversos perfis. Esses testes são essenciais para identificar possíveis barreiras de usabilidade e validar a experiência de uso em cenários reais, permitindo ajustes refinados para atender a uma ampla variedade de necessidades.

7. Resultados e Discussões

Para avaliar a acessibilidade, usabilidade e utilidade do aplicativo, os testes iniciais do aplicativo IFTranscript foram realizados testes com duas pessoas, uma com deficiência visual e outra sem deficiência, proporcionando uma avaliação abrangente e multifacetada da ferramenta:

- a) **Usuário com deficiência visual:** este perfil de usuário é crucial para avaliar a acessibilidade real do aplicativo, especialmente considerando a interação com tecnologias assistivas como leitores de tela. A experiência deste usuário é particularmente valiosa para compreender como o aplicativo se comporta em condições reais de uso por pessoas com deficiência visual, que constituem um público-alvo importante para a ferramenta.
- b) **Usuário sem deficiência especialista em acessibilidade digital:** este perfil oferece uma perspectiva técnica e especializada, capaz de identificar nuances e aspectos de acessibilidade que podem passar despercebidos por usuários não especializados. A experiência deste avaliador é fundamental para uma análise crítica e detalhada dos aspectos técnicos e funcionais do aplicativo.

Esta combinação de perfis permite uma avaliação tanto da perspectiva de acessibilidade quanto da funcionalidade geral. A diversidade dos perfis selecionados possibilitou uma análise abrangente, desde a usabilidade básica do aplicativo até aspectos técnicos

²⁷Disponível em: <https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.auditor>

mais específicos, como a qualidade das transcrições. Para garantir autonomia durante os testes, os usuários receberam o arquivo do aplicativo acompanhado de um manual detalhado sobre suas funcionalidades. Após a instalação em seus dispositivos, eles tiveram total liberdade para explorar todas as funcionalidades, testando o aplicativo conforme suas preferências e necessidades, sem qualquer intervenção dos autores.

O primeiro teste foi realizado pelo usuário com deficiência visual, utilizando um smartphone Samsung Galaxy A24. Este usuário demonstrou particular entusiasmo com diversos aspectos do aplicativo, especialmente em relação à sua acessibilidade e funcionalidades básicas. Um ponto destacado foi a facilidade de instalação. A interface do aplicativo recebeu avaliação positiva quanto à sua acessibilidade, não apresentando problemas significativos na interação com o leitor de tela. Este aspecto é particularmente relevante, pois indica que as etiquetas e descrições dos elementos da interface foram implementadas adequadamente, permitindo uma navegação fluida e compreensível através de tecnologias assistivas. Uma característica especialmente apreciada foi a capacidade do aplicativo de manter um registro interno tanto dos áudios quanto das transcrições realizadas. Esta funcionalidade de armazenamento foi vista como um diferencial importante, pois permite que os usuários mantenham um histórico organizado de suas transcrições, facilitando o acesso posterior ao material. No entanto, uma limitação técnica significativa foi identificada durante este teste: o aplicativo apresentou dificuldades no processamento de arquivos de áudio mais extensos. O usuário relatou que conseguiu realizar apenas transcrições bem-sucedidas com áudios de poucos segundos de duração, o que representa uma restrição importante para seu uso em contextos educacionais, onde as gravações tendem a ser mais longas.

O segundo teste foi conduzido pelo especialista em acessibilidade digital, utilizando um smartphone Samsung Galaxy M21S. Este usuário identificou questões técnicas mais específicas relacionadas à qualidade das transcrições. Uma das principais observações foi a presença de erros no processo de reconhecimento de fala, nos quais o sistema identifica e transcreve palavras que não estão presentes no áudio original. Esses erros resultam, frequentemente, de interpretações incorretas de fonemas, gerando transcrições com palavras diferentes das pronunciadas, o que pode comprometer a compreensão do conteúdo transcrito. Esta questão é particularmente relevante no contexto educacional, onde a precisão do conteúdo transcrito é fundamental para o aprendizado efetivo. Outra limitação importante apontada foi a ausência de diferenciação entre locutores nas transcrições. Em situações de diálogo ou em contextos com múltiplos falantes, como é comum em ambientes educacionais, esta limitação pode dificultar a compreensão da dinâmica da conversação e a identificação de quem está falando em cada momento. A restrição quanto ao tamanho dos arquivos de áudio foi corroborada pelo segundo usuário, confirmando ser uma limitação sistemática do aplicativo e não um problema isolado. Esta convergência nas observações dos dois usuários enfatiza a necessidade de otimização do processamento de arquivos mais extensos.

Apesar das limitações identificadas, é importante ressaltar que o aplicativo demonstrou potencial significativo como ferramenta de tecnologia assistiva, especialmente considerando sua acessibilidade bem-sucedida para usuários com deficiência visual. A facilidade de uso e a boa integração com tecnologias assistivas estabelecem uma boa base para futuros desenvolvimentos e aprimoramentos do sistema.

As limitações identificadas durante os testes do aplicativo IFTranscript, embora significativas, apresentam soluções técnicas viáveis que podem ser implementadas em versões futuras do sistema. Duas principais áreas de melhoria foram identificadas: o processamento de arquivos de áudio mais extensos e a qualidade das transcrições geradas.

A limitação no processamento de arquivos de áudio mais longos está diretamente relacionada à infraestrutura computacional utilizada durante os testes. O *backend* do sistema foi executado em uma máquina virtual no IFRS Campus Porto Alegre, com recursos computacionais limitados. Esta restrição pode ser superada através da migração do sistema para uma infraestrutura com maior capacidade de processamento.

Quanto à qualidade das transcrições, as inconsistências identificadas podem ser minimizadas através de uma configuração mais refinada do *framework* Whisper. Esta ferramenta de código aberto oferece diversos parâmetros de configuração que podem ser ajustados para otimizar o processo de transcrição, incluindo:

- a) **Ajuste do tamanho do modelo:** o Whisper disponibiliza diferentes tamanhos de modelo (*tiny*, *base*, *small*, *medium*, *large* e *turbo*), cada um oferecendo um equilíbrio diferente entre velocidade de processamento e precisão. A utilização de modelos maiores pode resultar em transcrições mais precisas, embora requeira mais recursos computacionais;
- b) **Configuração de parâmetros de linguagem:** o *framework* permite a especificação do idioma esperado, o que pode melhorar significativamente a precisão do reconhecimento de palavras e estruturas gramaticais específicas do português;
- c) **Otimização do pré-processamento de áudio:** ajustes na forma como o áudio é processado antes da transcrição, incluindo normalização de volume e redução de ruído, podem contribuir para resultados mais precisos;
- d) **Implementação de pós-processamento:** a adição de uma etapa de pós-processamento pode ajudar a corrigir erros comuns e melhorar a formatação do texto transcrito, incluindo a pontuação e a estruturação de parágrafos.

A implementação dessas melhorias técnicas, combinada com uma infraestrutura computacional mais robusta, tem o potencial de resolver as principais limitações identificadas durante os testes, tornando o IFTranscript uma ferramenta mais efetiva para seu propósito educacional. Estas otimizações permitiriam que o aplicativo processasse arquivos de áudio mais longos com maior precisão, mantendo sua característica fundamental de acessibilidade, que já se mostrou bem-sucedida nos testes realizados.

8. Conclusões

O desenvolvimento desta aplicação é visto como um meio significativo de promover a inclusão e a autonomia dos estudantes com deficiência no ambiente educacional. Ao fornecer uma ferramenta acessível e gratuita para transcrição de áudio, espera-se reduzir a barreira comunicacional durante as aulas, permitindo que os alunos acompanhem o material de forma independente e eficaz. Ademais, essa iniciativa visa contribuir para o aprimoramento da qualidade do aprendizado desses alunos, capacitando-os a participar plenamente do processo educacional e alcançar seu pleno potencial acadêmico.

Espera-se que a implementação desta aplicação de transcrição de áudio proporcione uma experiência educacional significativamente aprimorada para alunos com de-

ficiência, conferindo-lhes maior autonomia e participação ativa nas aulas. Acredita-se que a transcrição de áudio possa permitir que esses alunos acompanhem o conteúdo de forma independente e eficaz, reduzindo barreiras comunicacionais e promovendo a inclusão no ambiente educacional.

Os testes iniciais do aplicativo demonstraram resultados promissores, com destaque especial para sua acessibilidade e facilidade de uso, particularmente para usuários com deficiência visual. A interface intuitiva e a boa integração com tecnologias assistivas evidenciaram o significativo potencial da ferramenta como tecnologia assistiva no contexto educacional. Durante os testes, foram identificadas algumas limitações técnicas, como restrições no processamento de arquivos de áudio longos e inconsistências nas transcrições geradas, que, embora presentes, não comprometeram a avaliação positiva geral do aplicativo. Estas limitações, de natureza principalmente técnica, podem ser plenamente superadas em trabalhos futuros através de otimizações no *backend* do sistema e ajustes mais refinados no *framework* de transcrição utilizado, permitindo que o aplicativo alcance seu pleno potencial como ferramenta de inclusão educacional.

Finalmente, para validar efetivamente a aplicabilidade e eficácia do IFTranscript no contexto educacional, faz-se necessária a realização de testes mais abrangentes, envolvendo um número maior de usuários em situações reais de sala de aula. Estes testes permitiriam avaliar o impacto real da ferramenta no processo de aprendizagem, identificar possíveis necessidades de adaptação específicas para diferentes contextos educacionais e verificar sua efetividade como instrumento de inclusão. A experiência em ambiente real de ensino também forneceria informações valiosas para futuras melhorias e adaptações do sistema, contribuindo para o desenvolvimento de uma ferramenta cada vez mais alinhada às necessidades dos estudantes com deficiência.

Referências

- Abid, B., Medani, A., Toktassyn, A., and Manni, Y. (2023). Reviewing audio transcription in education. *Journal of IoT and Intelligent Solutions*, 1(1):15–19.
- Basak, S., Agrawal, H., Jena, S., Gite, S., Bachute, M., Pradhan, B., and Assiri, M. (2022). Challenges and limitations in speech recognition technology: A critical review of speech signal processing algorithms, tools and systems. *Computer Modeling in Engineering & Sciences*, 135:1–37.
- Brasil (2015). Lei Nº 13.146, de 6 de julho de 2015. Diário Oficial da República Federativa do Brasil. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência).
- CAT, Comitê de Ajudas Técnicas (2007). Ata da Reunião VII, de dezembro de 2007. Secretaria Especial dos Direitos Humanos da Presidência da República (CORDE/SEDH/PR). Disponível em: https://www.assistiva.com.br/Ata_VII_Reuni%3a3o_do_Comite_de_Ajudas_T%3a9cnicas.pdf. Acessado em: 24 de jul. de 2024.
- Connelly, A. and Doyle, J. (2023). Speech to text, text to speech: Support for deaf and hard of hearing students. *Odyssey: New Directions in Deaf Education*, 23:64–67.
- Galvão Filho, T. (2012). Tecnologia Assistiva: favorecendo o desenvolvimento e a aprendizagem em contextos educacionais inclusivos. In Giroto, C. R. M., Poker, R. B., and

- Omote, S., editors, *As tecnologias nas práticas pedagógicas inclusivas*, pages 65–92. Cultura Acadêmica, Marília/SP.
- Galvão Filho, T. (2013a). A construção do conceito de Tecnologia Assistiva: alguns novos interrogantes e desafios. *Revista da FACED - Entreideias: Educação, Cultura e Sociedade*, 2(1):25–42.
- Galvão Filho, T. (2013b). Tecnologia Assistiva e educação. In Souza, R. C. S. and Barbosa, J. S. L., editors, *Educação inclusiva, tecnologia e Tecnologia Assistiva*, pages 15–38. Criação, Aracaju.
- Google (2024). Acessibilidade no Jetpack Compose. Google Developers. Disponível em: <https://developer.android.com/develop/ui/compose/accessibility>. Acessado em: 12 de out. de 2024.
- Irdamurni, I., Nurhastuti, N., Amini, R., and Taufan, J. (2020). Implementation of speech to-text application for deaf students on inclusive education course. In *Proceedings of the Third Workshop on Multidisciplinary and Its Applications, WMA-3 2019, 11-14 December 2019, Medan, Indonesia*. January 2020.
- Kambouri, M., Simon, H., and Brooks, G. (2023). Using speech-to-text technology to empower young writers with special educational needs. *Research in Developmental Disabilities*, 135:104466.
- Levine, S., Hsieh, H., Southerton, E., and Silverman, R. (2023). How high school students used speech-to-text as a composition tool. *Computers and Composition*, 68:102775.
- Matre, M. E. and Cameron, D. L. (2022). A scoping review on the use of speech-to-text technology for adolescents with learning difficulties in secondary education. *Disability and Rehabilitation: Assistive Technology*, 19(3):1103–1116. PMID: 36427182.
- Pathak, M. A., Raj, B., Rane, S. D., and Smaragdis, P. (2013). Privacy-preserving speech processing: cryptographic and string-matching frameworks show promise. *IEEE Signal Processing Magazine*, 30(2):62–74.
- Yashina, E., Rubanik, T., and Chukhray, A. (2022). Speech-to-text software design for the high education learning. In *Proceedings of the 2nd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2022), December 2-4, 2022, Łódź, Poland*, National Aerospace University «Kharkiv Aviation Institute», Chkalova St., 17, 61070 Kharkiv, Ukraine.

APÊNDICE A - API DE TRANSCRIÇÃO

Quadro A1. Comparação de API's de transcrição

Ferramenta	Preço	Idiomas	Obs.	Extra
OpenAI Whisper	\$0,006 /minuto	60	A acuracidade da transcrição depende do idioma.	Pode traduzir a transcrição para o Inglês
Google Speech-to-Text	Grátis*	125	Até 60 minutos por mês e \$300 em créditos para teste de outras ferramentas.	Transcrição em tempo real
Amazon Transcribe	Grátis*	155	Até 60 minutos por mês durante 12 meses.	API's separadas que compreendem telefonemas dos clientes e conversas médicas. Transcrição em tempo real.
Azure Speech-to-Text	Grátis*	156	Até 5 horas por mês.	Reconhecimento de locutor. Modelos customizáveis de acordo com a voz do locutor. Transcrição em tempo real.
IBM Watson Speech-to-Text	Grátis*	14+	Até 500 minutos por mês.	Reconhecimento de locutor. Modelos customizáveis de acordo com a voz do locutor.
Transkriptor	Pago	146	-	Não é tão sofisticada quanto as outras. Não tem muita opção de customização.
Deepgram	\$0,0043 /minuto	32	Até \$200 de uso.	Possibilidade de selecionar modelos.
AssemblyAI	Grátis*	99	Até 200 horas de áudio.	-
Fireflies	Grátis*	69+	Até 800 minutos de gravações e até 200 MB de tamanho de arquivo de áudio e 100 MB de vídeo.	-

Fonte: Elaborado pelo autor (2024).

APÊNDICE B - APLICAÇÕES DE TRANSCRIÇÃO

Quadro B1. Comparação de Aplicações de transcrição

Ferramenta	Preço	Plataforma	Idiomas	Obs.	Extra
Notta	Grátis*	Web / Android / iOS	42	Limite de 120 minutos por mês e 5 minutos por áudio	Resumo da Transcrição. Exportação das transcrições.
Transkriptor	\$4,99 /mês	Web / Android / iOS	100+	Limite de 5 horas por mês.	Exportação das transcrições. Editar transcrição. Cria Legendas.
Happyscribe	\$17,00 /mês	Web	66	Limite de 120 minutos por mês.	Exportação das transcrições. Editar transcrição. Cria Legendas.
Otter	Grátis*	Web / Android / iOS	Inglês	Limite de 300 minutos por mês e 30 minutos por áudio	Exportação das transcrições. Resumo da transcrição. Chat interno.
Fireflies	Grátis*	Web / Android / iOS	69+	Limite de 800 minutos armazenados	Resumo da transcrição.
Rev	\$1,50 /minuto	Web / Android / iOS	22	Limite de 5 minutos por áudio.	Resumo da transcrição (inglês somente e pago).
Sonix	\$10,00 /hora	Web	49+	Pré-pago	Exportação das transcrições. Resumo da transcrição. Cria Legendas.

Fonte: Elaborado pelo autor (2024).

APÊNDICE C - BIBLIOTECAS DE TRANSCRIÇÃO

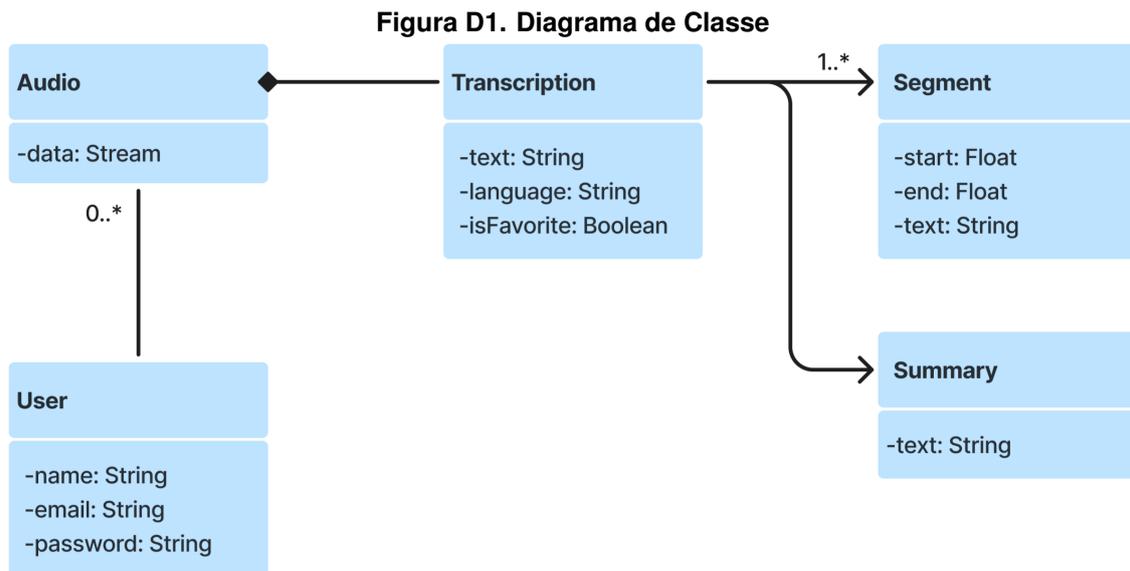
Quadro C1. Comparação de Bibliotecas de transcrição

Ferramenta	Idiomas	Última Atualização	Forks	Obs.
Whisper	60	17/11/2023	7k	Pode traduzir a transcrição para o Inglês
Mozilla's DeepSpeech	Inglês	10/12/2020	3.9k	-
Kaldi	Inglês	30/04/2024	5.3k	Focada em reconhecer locutores
Flashlight	Inglês	19/03/2022	1,5k	Transcreve áudio diretamente para caracteres, pulando a etapa de alinhamento temporal.

Fonte: Elaborado pelo autor (2024).

APÊNDICE D - DIAGRAMA DE CLASSE

O diagrama apresenta o arquivo de áudio que possui uma áudio transcrição, sendo que cada transcrição possui um resumo e está associada a um ou mais segmentos do áudio. Os segmentos representam trechos da transcrição organizados por intervalo de tempo. Destaca-se ainda que cada áudio precisa estar vinculado a um usuário, para que seja possível identificar os seus históricos, caso seja necessário. A Figura D1 ilustra o Diagrama de Classe definido.



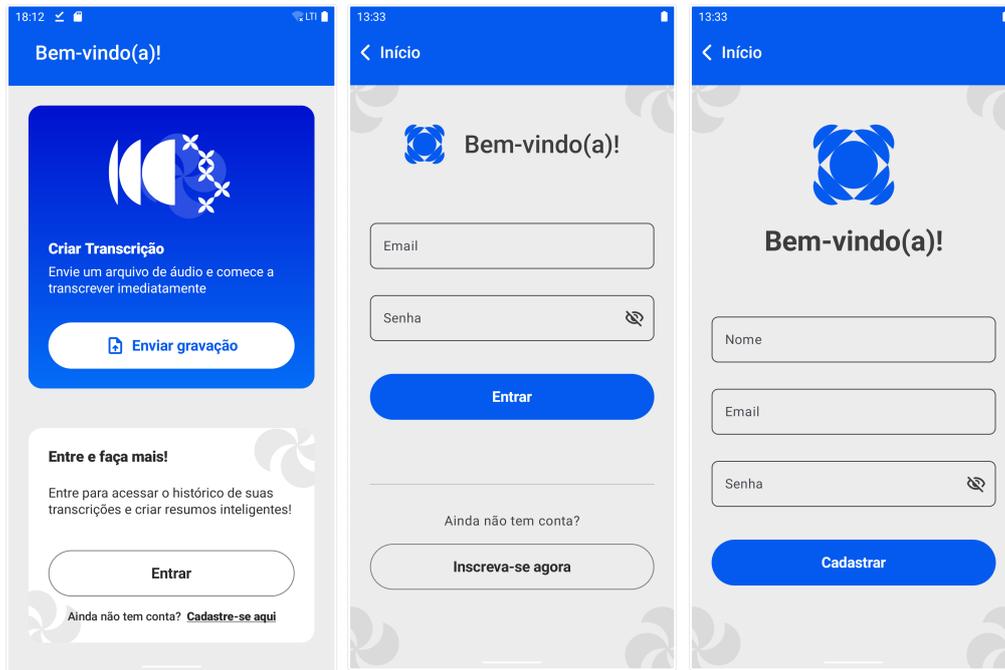
Fonte: Elaborado pelo autor (2024).

APÊNDICE E - PROTÓTIPO DAS TELAS

Foram definidas as seguintes telas:

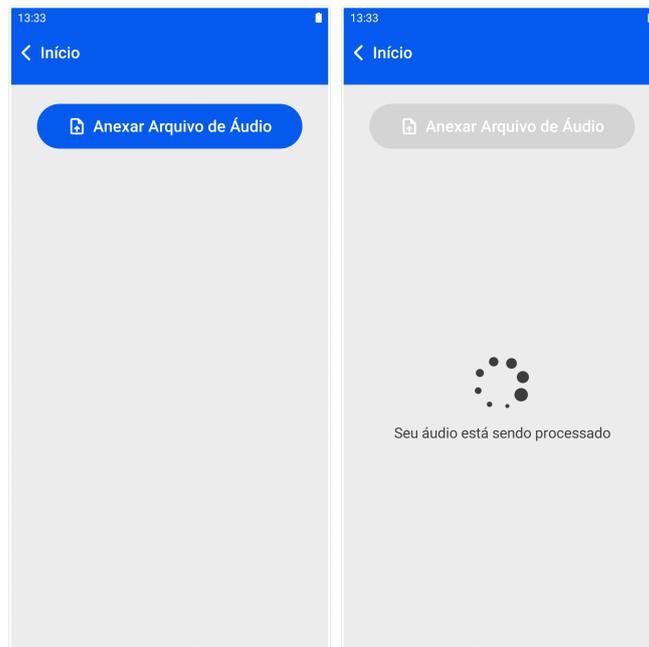
- a) **Boas-vindas:** é a tela inicial do aplicativo. Nela existem três ações possíveis. A primeira é gerar uma transcrição sem realização de cadastro ou login. As ações seguintes são de login e cadastro do usuário (Figura E1);
- b) **Login:** tela onde o usuário já cadastrado realiza seu login (Figura E1);
- c) **Cadastro:** tela onde o usuário poderá realizar um novo cadastro informando nome, e-mail e senha (Figura E1);
- d) **Enviar áudio:** tela dedicada a mostrar a opção ao usuário para anexar um arquivo de áudio a partir dos arquivos do aparelho (Figura E2);
- e) **Progresso da transcrição:** indica que a transcrição está em andamento usando uma barra de progresso, isso indica para o usuário que a ação ainda está sendo realizada (Figura E2);
- f) **Página principal:** esta é a tela inicial exibida para usuários logados. Nela, são mostradas as três transcrições mais recentes, facilitando o acesso rápido aos últimos conteúdos e um botão de ação destacado para iniciar uma nova transcrição (Figura E3);
- g) **Minhas transcrições:** a tela dedicada a exibir a lista de todas as transcrições armazenadas pelo usuário. A interface permite a visualização de cada transcrição e oferece a opção de excluir transcrições diretamente da lista (Figura E3);
- h) **Detalhe da transcrição:** a tela exibe o conteúdo completo da transcrição em formato de texto. Ela oferece duas funcionalidades principais: compartilhar o texto da transcrição e visualizar ou gerar um resumo. A opção de resumo é mostrada apenas para usuários cadastrados. Além disso, o usuário pode ajustar dinamicamente o tamanho da fonte conforme sua preferência, proporcionando uma leitura mais confortável. A tela também permite a reprodução do áudio correspondente à transcrição para aqueles usuários que estão logados (Figura E3);
- i) **Resumo da transcrição:** nesta tela, o usuário pode visualizar o texto resumido da transcrição. Há opções para ajustar dinamicamente o tamanho da fonte, permitindo uma leitura personalizada. Além disso, a tela oferece um botão de ação para compartilhar o conteúdo do resumo (Figura E4).

Figura E1. Protótipo das telas de Boas vindas, Login e Cadastro



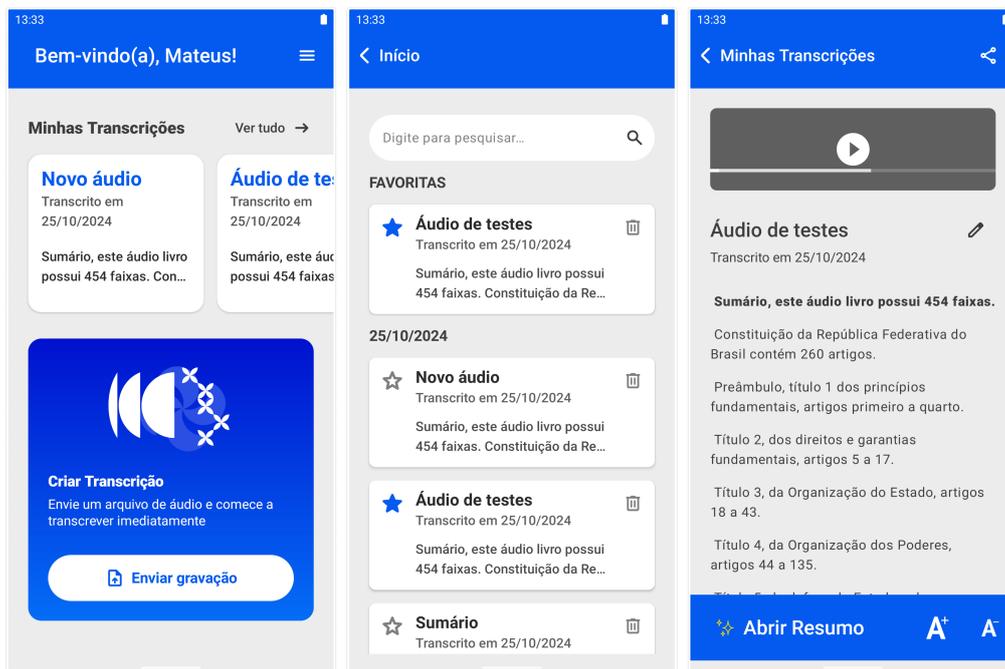
Fonte: Elaborado pelo autor (2024).

Figura E2. Protótipo das telas do fluxo de transcrição do áudio



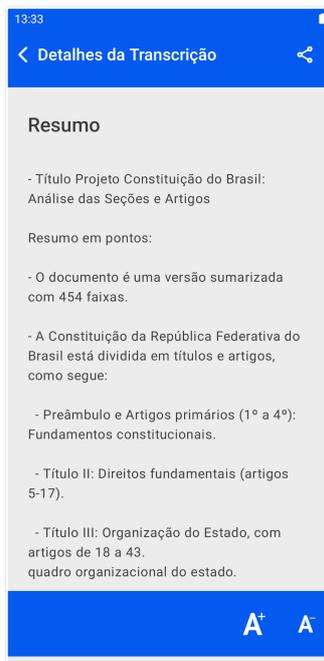
Fonte: Elaborado pelo autor (2024).

Figura E3. Protótipo das telas Principal, Transcrições e Detalhe da transcrição



Fonte: Elaborado pelo autor (2024).

Figura E4. Protótipo da tela de resumo



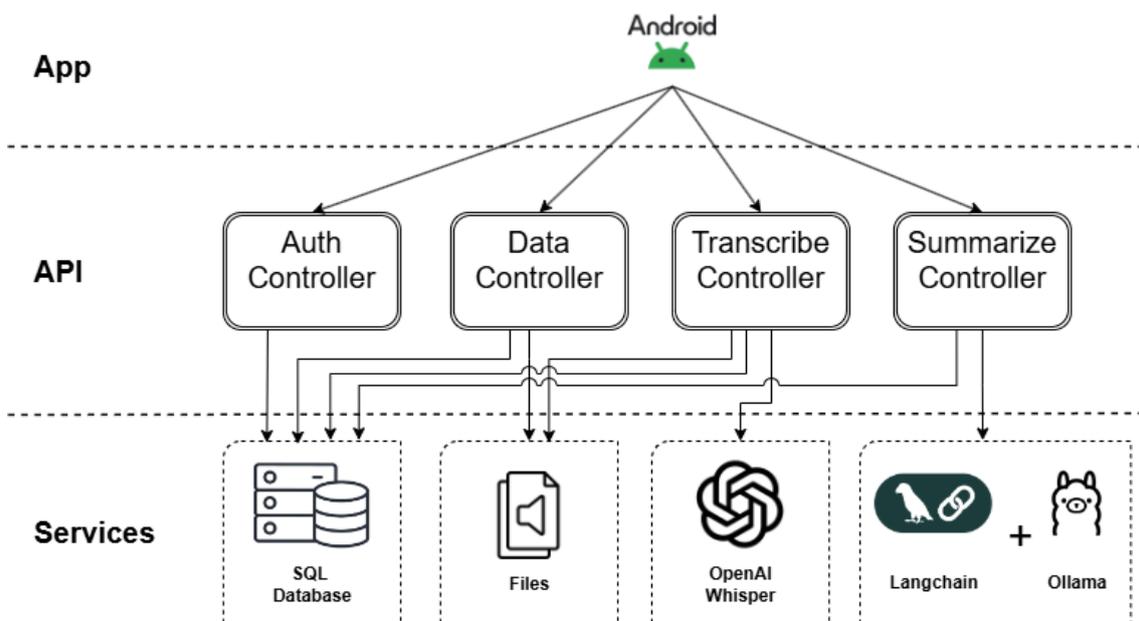
Fonte: Elaborado pelo autor (2024).

APÊNDICE F - DIAGRAMA ESQUEMÁTICO DA ARQUITETURA DO SISTEMA

A aplicação Android realiza requisições a uma API dividida em controllers, cada um responsável por diferentes funcionalidades do sistema. São eles:

- **Auth Controller:** contém os *endpoints* para cadastro e autenticação do usuário;
- **Data Controller:** contém os *endpoints* responsáveis por buscar os dados do usuário no banco de dados. Com ele é possível resgatar os áudios e transcrições do usuário;
- **Transcribe Controller:** é responsável por gerar a transcrição do arquivo de áudio, além de salvar a transcrição para o caso de usuários cadastrados;
- **Summarize Controller:** é responsável por gerar o resumo de transcrições previamente salvas.

Figura F1. Diagrama da Arquitetura do Sistema



Fonte: Elaborado pelo autor (2024).