

CENTRO UNIVERSITÁRIO SAGRADO CORAÇÃO – UNISAGRADO

GABRIELLI CORACINI DE FREITAS

RYAN GOMES VILLELA

DESENVOLVIMENTO DE UM SISTEMA INTEGRADO PARA TRADUÇÃO  
AUTOMÁTICA DE VÍDEOS EM LINGUAGEM DE SINAIS UTILIZANDO  
RECONHECIMENTO DE VOZ E AVATARES DIGITAIS INTERATIVOS

BAURU

2024

GABRIELLI CORACINI DE FREITAS  
RYAN GOMES VILLELA

DESENVOLVIMENTO DE UM SISTEMA INTEGRADO PARA TRADUÇÃO  
AUTOMÁTICA DE VÍDEOS EM LINGUAGEM DE SINAIS UTILIZANDO  
RECONHECIMENTO DE VOZ E AVATARES DIGITAIS INTERATIVOS

Trabalho de Conclusão de Curso apresentado  
como parte dos requisitos para obtenção do  
título de bacharel em Ciência da Computação -  
Centro Universitário Sagrado Coração.

Orientador: Prof.<sup>a</sup> Dr. Elvio Gilberto da Silva e  
Me. Roque Maitino Neto

BAURU  
2024

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

F862d

Freitas, Gabrielli Coracini de

Desenvolvimento de um sistema integrado para tradução automática de vídeos em linguagem de sinais utilizando reconhecimento de voz e avatares digitais interativos / Gabrielli Coracini de Freitas; Ryan Gomes Villela. -- 2024.

17f.

Orientador: Prof. Dr. Elvio Gilberto da Silva

Coorientador: Prof. M.e Roque Maitino Neto

Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Centro Universitário Sagrado Coração - UNISAGRADO - Bauru - SP

1. Libras. 2. Surdos. 3. Inteligência Artificial. 4. Aplicativo. 5. Tradução. I. Villela, Ryan Gomes. II. Silva, Elvio Gilberto da. III. Maitino Neto, Roque. IV. Título.

GABRIELLI CORACINI DE FREITAS

RYAN GOMES VILLELA

DESENVOLVIMENTO DE UM SISTEMA INTEGRADO PARA TRADUÇÃO  
AUTOMÁTICA DE VÍDEOS EM LINGUAGEM DE SINAIS UTILIZANDO  
RECONHECIMENTO DE VOZ E AVATARES DIGITAIS INTERATIVOS

Trabalho de Conclusão de Curso apresentado  
como parte dos requisitos para obtenção do  
título de bacharel em Ciência da Computação -  
Centro Universitário Sagrado Coração.

Aprovado em: \_\_\_/\_\_\_/\_\_\_.

Banca examinadora:

---

Prof.<sup>a</sup> Me. Saulo Silva Coelho  
Centro Universitário Sagrado Coração

---

Prof.<sup>a</sup> Me. Vinicius Santos Andrade  
Centro Universitário Sagrado Coração

## **AGRADECIMENTOS**

Em primeiro lugar, agradecemos a Deus, que esteve presente em cada etapa desta jornada, proporcionando força, sabedoria e resiliência para superarmos os desafios e alcançarmos os objetivos estabelecidos ao longo dos anos de estudo.

Às nossas famílias, que sempre foram nosso porto seguro, oferecendo suporte emocional, incentivo constante e compreensão nos momentos mais difíceis. Aos nossos amigos, que compartilharam conosco sorrisos, palavras de encorajamento e, muitas vezes, a paciência necessária para lidar com as adversidades.

Ao professor Elvio Gilberto da Silva e Roque Maitino Neto, expressamos nossa sincera gratidão por sua orientação atenta e pela paciência com que compartilhou seu conhecimento, ajudando a moldar nossa visão crítica e ampliando nossos horizontes acadêmicos. Seus conselhos e ensinamentos foram pilares importantes para a construção deste projeto.

Por fim, agradecemos a todos os professores e profissionais que, direta ou indiretamente, participaram dessa trajetória, transmitindo conhecimentos valiosos e incentivando nosso crescimento constante. Cada contribuição, por menor que tenha parecido, teve impacto profundo na construção deste trabalho e na formação de quem somos hoje.

“A maior invenção do mundo não é a minha tecnologia! É a morte! pois através dela, o velho sempre dará lugar para o novo!”  
(STEVE JOBS, 2005).

## **LISTA DE ABREVIATURAS E SIGLAS**

IBGE	Instituto Brasileiro de Geografia e Estatística.
Libras	Língua Brasileira de Sinais.
LBI	Lei Brasileira de Inclusão da Pessoa com.
FFmpeg	Fast Forward MPEG.
API	Application Programming Interface.
3D	Tridimensional.

## **RESUMO**

O projeto tem como prioridade ajudar a compreensão de pessoas surdas quando estiverem assistindo algo sem um intérprete de libras, por isso, o projeto tem como objetivo traduzir o áudio falado com som para a linguagem de sinais, utilizando uma tecnologia de reconhecimento de voz e um avatar para fazer os gestos com as mãos. A falta de inclusão de pessoas não ouvintes na sociedade é uma questão muito real que pode afetar até a qualidade de vida e a igualdade, hoje infelizmente, a língua brasileira de sinais (libras) não é vista como essencial para grande parte da população, segundo fonte do IBGE em 2021 apenas 10 milhões de pessoas no Brasil sabem libras fluente, sendo que 2,7 milhões não ouvem nada. A implementação de programas de IA para poder traduzir linguagem de sinais pode desempenhar um papel importantíssimo para a inclusão dessas pessoas entre as falantes, essa tecnologia tem a grande capacidade de mudar as vidas das pessoas surdas, ajudando também o acesso à educação, ao trabalho e aos serviços essenciais. Tomamos como referência para realizar esse projeto o aplicativo Vlibras, um aplicativo inovador cujo sua função é traduzir textos para a linguagem brasileira de sinais em tempo real, o aplicativo tem 3 avatares responsável por realizar a tradução.

**Palavras-chave:** libras; surdos; Inteligência Artificial; aplicativo; tradução

## ABSTRACT

The project prioritizes helping deaf individuals understand spoken content when there is no sign language interpreter available. Therefore, the project's goal is to translate spoken audio into sign language using voice recognition technology and an avatar to perform the hand gestures.

The lack of inclusion for deaf people in society is a very real issue that can affect their quality of life and equality. Unfortunately, Brazilian Sign Language (Libras) is not widely regarded as essential by much of the population. According to IBGE data from 2021, only 10 million people in Brazil are fluent in Libras, while 2.7 million people are completely deaf.

Implementing AI programs to translate into sign language can play a crucial role in including these individuals among speakers. This technology has the immense potential to change the lives of deaf people, improving access to education, work, and essential services.

We took inspiration for this project from the Vlibras application, an innovative app designed to translate text into Brazilian Sign Language in real time. The app uses three avatars responsible for performing the translations.

**Keywords:** Libras; deaf; Artificial Intelligence; app; translation

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	1
2	<b>OBJETIVOS</b> .....	1
2.1	OBJETIVO GERAL.....	2
2.2	OBJETIVOS ESPECÍFICOS .....	2
3	A SITUAÇÃO DAS PESSOAS COM DEFICIÊNCIA AUDITIVA NO BRASIL..	2
3.1	A BARREIRAS EDUCACIONAIS E SOCIAIS. ....	2
3.2	A IMPORTÂNCIA DA TECNOLOGIA COMO FERRAMENTA DE INCLUSÃO .....	3
4	TECNOLOGIAS DE RECONHECIMENTO DE VOZ.....	3
4.1	TECNOLOGIAS DE RECONHECIMENTO DE VOZ MODERNAS .....	3
4.2	APLICAÇÕES EM ACESSIBILIDADE .....	3
5	SISTEMA INTELIGENTE DE TRADUÇÃO DE ÁUDIO PARA LINGUAGEM DE SINAIS .....	4
5.1	VLIBRAS PARA TRADUÇÃO DE TEXTO EM LIBRAS.....	4
5.2	VLIBRAS PARA TRADUÇÃO DE TEXTO EM LIBRAS.....	4
5.3	WHISPER PARA RECONHECIMENTO DE VOZ. ....	4
5.4	FFMPEG PARA MANIPULAÇÃO DE ÁUDIO.....	4
6	<b>METODOLOGIA</b> .....	5
6.1	ETAPAS DO DESENVOLVIMENTO .....	5
6.1.1	PESQUISA E LEVANTAMENTO DE REQUISITOS .....	5
6.1.2	DEFINIÇÃO DA ARQUITETURA DO SISTEMA .....	5
6.1.3	DESENVOLVIMENTO DO SISTEMA .....	5
6.1.4	VALIDAÇÃO DO SISTEMA .....	5
6.1.5	ANÁLISE E AJUSTES FINAIS .....	6
6.2	FERRAMENTAS UTILIZADAS.....	6
6.3	CRITÉRIOS DE VALIDAÇÃO.....	6
6.4	FLUXO DO SISTEMA .....	6
6.5	LIMITAÇÕES DO MÉTODO.....	6
7	<b>RESULTADOS</b> . ....	7
7.1	PROCESSAMENTO DE ÁUDIO E TRANSCRIÇÃO .....	7
7.2	TRADUÇÃO PARA LIBRAS.....	7
7.3	DESEMPENHO E USABILIDADE .....	7
7.4	LIMITAÇÕES IDENTIFICADAS.....	8

7.5	DESENVOLVIMENTO DA API E INTEGRAÇÃO.....	8
7.6	CONCLUSÃO.....	8
8	<b>CONTRIBUIÇÕES DO PROJETO.....</b>	<b>8</b>
8.1	LIMITAÇÕES E TRABALHOS FUTUROS.....	9
	REFERÊNCIAS .....	9

## 1 INTRODUÇÃO

A inclusão de pessoas com deficiência auditiva na sociedade é um desafio persistente que impacta diretamente a qualidade de vida e a igualdade de oportunidades dessas pessoas. Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), em 2021, apenas 10 milhões de brasileiros eram fluentes na Língua Brasileira de Sinais (Libras), enquanto 2,7 milhões não possuem nenhuma capacidade auditiva (IBGE, 2021). Em paralelo, avanços tecnológicos têm gerado soluções inovadoras em áreas como tradução automática e acessibilidade, mas sua aplicação no campo da tradução de áudio para Libras ainda é limitada.

Com os constantes avanços tecnológicos e sua aplicação em favor da educação, é possível integrar alunos surdos a ferramentas de tecnologia assistiva que viabilizem o aprendizado (Rodrigues, 2024). Contudo, apesar dessa potencialidade, a popularização da internet como ferramenta educacional ainda enfrenta desafios relacionados à acessibilidade. Muitos vídeos usados para estudos não contam com intérpretes de Libras ou outros recursos inclusivos, o que pode criar barreiras significativas para a inclusão de pessoas surdas no aprendizado (Lima, 2020). Tal cenário reforça a necessidade urgente de soluções tecnológicas inovadoras que ampliem o suporte linguístico e eliminem obstáculos ao aprendizado dessa comunidade.

Diante desse cenário, este estudo tem como objetivo desenvolver um sistema inteligente capaz de traduzir áudio para Libras em tempo real, utilizando tecnologias de reconhecimento de voz e avatares animados para simular expressões faciais e corporais. O foco principal está em promover a acessibilidade de pessoas surdas em contextos educacionais.

A relevância desta pesquisa reside no impacto direto que pode gerar na inclusão de pessoas surdas na sociedade, dado que há uma necessidade crescente de ferramentas acessíveis para promover a comunicação. Esse fato enfatiza a importância de soluções tecnológicas que atendam às demandas dessa comunidade (Karnopp; Quadros, 2004). Alves et al. (2009) destacam a crescente demanda por soluções tecnológicas, especialmente na área educacional, onde a equidade é frequentemente comprometida. Além disso, pesquisas recentes exploram tecnologias assistivas, como o reconhecimento de voz (Huang et al., 2019) e o uso de avatares animados como soluções viáveis para melhorar a acessibilidade e inclusão (Kaur et al., 2020).

Este estudo apresenta contribuições significativas tanto no campo acadêmico quanto na prática social. No aspecto acadêmico, visa ampliar a literatura sobre tecnologias assistivas, oferecendo uma abordagem inovadora para a tradução de áudio para Libras em tempo real. Essa iniciativa dialoga diretamente com a necessidade de soluções que alinhem avanços tecnológicos à inclusão educacional, como apontado por Quadros; Karnopp (2004) e Alves et al. (2009). Socialmente, o sistema proposto busca impactar diretamente a vida de pessoas com deficiência auditiva, facilitando sua integração em ambientes educacionais e sociais. A aplicação de tecnologias como reconhecimento de voz e avatares animados, como explorado por Huang et al. (2019) e Kaur et al. (2020), reforça o potencial desta pesquisa em oferecer uma solução prática, acessível e de alta relevância para promover a equidade e a acessibilidade em diversas áreas, especialmente na educação.

## 2 OBJETIVOS

A seguir, serão apresentados o objetivo geral e os objetivos específicos que guiam o desenvolvimento deste projeto.

## 2.1 OBJETIVO GERAL

Desenvolver um sistema inteligente que capture o áudio de vídeos e traduza automaticamente para a Língua Brasileira de Sinais (Libras), utilizando tecnologias de ponta como o modelo Whisper para transcrição automática de áudio e a ferramenta VLibras para tradução e visualização, promovendo inclusão social e acessibilidade digital para pessoas surdas.

## 2.2 OBJETIVOS ESPECÍFICOS

- a) Implementar a captura e extração de áudio de vídeos provenientes de plataformas digitais como o YouTube.
- b) Integrar o modelo Whisper da OpenAI para transcrição automática de áudio em texto, garantindo alta precisão mesmo em contextos desafiadores, como sotaques e ruído de fundo.
- c) Utilizar a ferramenta VLibras para traduzir o texto transcrito para Libras, exibindo a tradução diretamente em uma interface acessível e interativa.
- d) Desenvolver uma interface web intuitiva para facilitar o upload de links ou arquivos de vídeo e exibir os resultados processados.
- e) Aprimorar a experiência do usuário utilizando ferramentas modernas de estilização e acessibilidade, como Bootstrap para interface responsiva e integração contínua com VLibras para suporte em tempo real.

## 3 A SITUAÇÃO DAS PESSOAS COM DEFICIÊNCIA AUDITIVA NO BRASIL

A população com deficiência auditiva no Brasil enfrenta diversos desafios sociais, educacionais e culturais, muitos dos quais estão profundamente enraizados na falta de acessibilidade e inclusão em diferentes esferas da sociedade. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), o Censo de 2010 identificou cerca de 9,7 milhões de pessoas com algum grau de deficiência auditiva no país, representando 5,1% da população à época. Entre essas pessoas, aproximadamente 2,1 milhões declararam ter grande dificuldade ou incapacidade total de ouvir (IBGE, 2010). Esses números evidenciam a relevância de debater as condições de vida e inclusão dessa parcela significativa da população.

### 3.1 A BARREIRAS EDUCACIONAIS E SOCIAIS.

A educação inclusiva ainda é um dos maiores desafios para pessoas com deficiência auditiva no Brasil. A Lei Brasileira de Inclusão da Pessoa com Deficiência (LBI) – Lei nº 13.146/2015 – estabelece que o Estado deve garantir igualdade de condições de acesso à educação, promovendo o uso da Língua Brasileira de Sinais (Libras) como forma de comunicação e instrução. Contudo, a implementação dessa lei enfrenta dificuldades práticas, como a escassez de intérpretes de Libras nas escolas e universidades e a falta de preparo de professores para atender alunos surdos (Quadros; Karnopp, 2020).

Além disso, muitos indivíduos com deficiência auditiva relatam dificuldades em acessar serviços básicos, como saúde e transporte público, devido à ausência de medidas acessíveis, como legendas em tempo real ou profissionais capacitados em Libras. Essa exclusão limita a autonomia dessas pessoas e reforça a desigualdade estrutural presente no país (Pereira, 2019).

## 3.2 A IMPORTÂNCIA DA TECNOLOGIA COMO FERRAMENTA DE INCLUSÃO

Nos últimos anos, a tecnologia tem desempenhado um papel crucial na superação de algumas dessas barreiras. Ferramentas de reconhecimento de voz, sistemas de legendagem automática e avatares animados que traduzem falas em Libras têm surgido como soluções promissoras. Essas inovações não apenas facilitam a comunicação, mas também promovem maior participação das pessoas surdas em espaços sociais e profissionais (Oliveira; Costa, 2021). No entanto, a implementação dessas tecnologias depende de investimentos e da conscientização sobre sua importância.

## 4 TECNOLOGIAS DE RECONHECIMENTO DE VOZ

O reconhecimento de voz é uma área de pesquisa e desenvolvimento que tem evoluído rapidamente, impulsionada pela crescente demanda por interfaces de usuário mais acessíveis e naturais. Essa tecnologia permite que sistemas interpretem e convertam linguagem falada em texto, criando possibilidades para aplicações em diversas áreas, como automação, saúde, educação, e, especialmente, na acessibilidade para pessoas com deficiência auditiva.

O mesmo é definido como "a capacidade de um sistema computacional identificar e processar palavras ou frases faladas e convertê-las em dados legíveis ou comandos acionáveis" (Rabiner; Juang, 1993). Os primeiros estudos sobre o tema surgiram na década de 1950, com dispositivos como o "Audrey System" da Bell Labs, que reconhecia dígitos falados. No entanto, foi apenas com o avanço de técnicas de aprendizado de máquina e maior poder computacional nas décadas seguintes que essa tecnologia ganhou relevância prática (Young et al., 2002).

### 4.1 TECNOLOGIAS DE RECONHECIMENTO DE VOZ MODERNAS

Com os avanços em aprendizado profundo, o reconhecimento de voz atingiu níveis de precisão e acessibilidade impressionantes. Ferramentas modernas, como o modelo Whisper, desenvolvido pela OpenAI, representam o estado da arte nesse campo. Whisper utiliza uma abordagem baseada em redes neurais de transformação (transformers) para transcrever áudio com alta fidelidade, mesmo em cenários desafiadores, como presença de ruído, diferentes sotaques e múltiplos idiomas. Este modelo suporta mais de 50 idiomas e é amplamente reconhecido por sua robustez e capacidade de transcrição em áudio previamente gravado ou em tempo real (Radford et al., 2022).

Outra tecnologia complementar é o FFmpeg, uma biblioteca de código aberto poderosa que facilita a manipulação de áudio e vídeo. Com o FFmpeg, é possível pré-processar arquivos de áudio, extraindo trechos relevantes e melhorando a qualidade do sinal, o que potencializa a eficácia do reconhecimento de voz. Esse pré-processamento é crucial em ambientes onde o áudio é comprometido por ruído de fundo ou baixa qualidade de gravação (FFmpeg, 2023).

### 4.2 APLICAÇÕES EM ACESSIBILIDADE

No contexto da inclusão, o reconhecimento de voz tem se destacado como uma ferramenta poderosa. Segundo Goldstein (2018), "a transcrição automática de fala é um divisor de águas para comunidades que enfrentam barreiras de comunicação". Ao ser integrado com APIs como o Whisper e ferramentas como a Vlibras, o reconhecimento de voz possibilita a tradução automática de áudio para Libras, promovendo a inclusão de pessoas com deficiência auditiva.

## 5 SISTEMA INTELIGENTE DE TRADUÇÃO DE ÁUDIO PARA LINGUAGEM DE SINAIS

O desenvolvimento de um sistema inteligente para tradução de áudio em linguagem de sinais requer a integração de tecnologias modernas, como Python, Whisper, VLibras e FFmpeg. Estas ferramentas desempenham papéis fundamentais para garantir a funcionalidade, precisão e acessibilidade do sistema.

### 5.1 VLIBRAS PARA TRADUÇÃO DE TEXTO EM LIBRAS

Python é amplamente utilizada para o desenvolvimento de sistemas de inteligência artificial e processamento de áudio devido à sua sintaxe acessível e ampla biblioteca de pacotes. Ferramentas como MoviePy, asyncio e PyDub permitem manipular áudio e integrar serviços de terceiros, enquanto o FFmpeg-python oferece acesso direto às funções da biblioteca FFmpeg.

Segundo Rossum et al. (2011), "Python fornece uma base sólida para projetos complexos devido à sua clareza, extensibilidade e suporte à integração de bibliotecas externas".

### 5.2 VLIBRAS PARA TRADUÇÃO DE TEXTO EM LIBRAS

O VLibras é uma ferramenta essencial para acessibilidade, permitindo a tradução de textos do português para a Língua Brasileira de Sinais (Libras). A ferramenta utiliza um avatar interativo para exibir os sinais de forma clara, promovendo inclusão digital para pessoas surdas.

VLibras é uma solução prática e eficiente que torna a comunicação em Libras acessível no ambiente digital (Almeida, 2022).

### 5.3 WHISPER PARA RECONHECIMENTO DE VOZ.

O Whisper, desenvolvido pela OpenAI, é uma das tecnologias mais avançadas em reconhecimento de voz. Ele utiliza aprendizado profundo para transcrever áudio em texto com alta precisão, mesmo em condições adversas, como ruídos e sotaques variados. Suporta mais de 50 idiomas e fornece resultados consistentes para transcrição de áudio.

O Whisper é projetado para ser robusto, mesmo em ambientes acústicos desafiadores, oferecendo transcrições precisas e confiáveis (Radford, 2022).

### 5.4 FFMPEG PARA MANIPULAÇÃO DE ÁUDIO

O FFmpeg é uma biblioteca de código aberto utilizada para processamento de arquivos multimídia. Ele suporta uma ampla gama de formatos e codecs, permitindo tarefas como extração, conversão e edição de áudio e vídeo.

Segundo Bellard (2023), "o FFmpeg é amplamente adotado devido à sua flexibilidade e performance robusta, tornando-se uma escolha padrão para manipulação de mídia em sistemas complexos".

## 6 METODOLOGIA

A metodologia deste trabalho foi elaborada com uma abordagem prática e exploratória, abrangendo as etapas de pesquisa, planejamento, desenvolvimento, validação e refinamento de um sistema inteligente para tradução de áudio em Libras. Este capítulo descreve o processo de desenvolvimento, as ferramentas utilizadas e os critérios de validação, garantindo que os objetivos do projeto sejam atingidos de forma estruturada.

### 6.1 ETAPAS DO DESENVOLVIMENTO

O desenvolvimento do sistema foi dividido em cinco etapas principais:

#### 6.1.1 Pesquisa e Levantamento de Requisitos

Foi realizada uma análise sobre as necessidades de acessibilidade para pessoas com deficiência auditiva, revisando estudos como os de Quadros e Karnopp (2020), que destacam os desafios de inclusão digital para surdos. Além disso, ferramentas como Whisper, VLibras e FFmpeg foram avaliadas para integrar tecnologias modernas que atendam a essas necessidades.

#### 6.1.2 Definição da Arquitetura do Sistema

A arquitetura foi projetada para integrar Python, Whisper, VLibras, e FFmpeg. O fluxo operacional abrange as seguintes etapas: captura de áudio, transcrição em texto, tradução para Libras e exibição dos sinais traduzidos em um avatar interativo.

#### 6.1.3 Desenvolvimento do Sistema

O desenvolvimento foi realizado em Python devido à sua robustez e ampla compatibilidade com bibliotecas. As etapas incluem:

- **Captura de Áudio:** Realizada com bibliotecas como FFmpeg-python, que também auxilia no pré-processamento.
- **Reconhecimento de Voz:** Utilizando o modelo Whisper, que transcreve o áudio capturado com alta precisão, mesmo em cenários com ruído ou sotaques variados (Radford et al., 2022).
- **Tradução para Libras:** Com a ferramenta VLibras, que converte texto em sinais utilizando avatares.
- **Otimização:** Implementação de algoritmos para melhorar a precisão em ambientes desafiadores.

#### 6.1.4 Validação do Sistema

Testes foram realizados em diferentes cenários para avaliar a funcionalidade do sistema:

- **Cenários Controlados:** Testes com áudio em ambientes silenciosos, garantindo transcrições precisas.

- **Cenários Realistas:** Testes com áudio em ambientes ruidosos ou com sotaques diversos, avaliando a robustez do sistema.

### 6.1.5 Análise e Ajustes Finais

A partir dos testes, foram identificados pontos de melhoria, incluindo ajustes para maior precisão em ambientes adversos e aprimoramento da interface para facilitar o uso.

## 6.2 FERRAMENTAS UTILIZADAS

As principais ferramentas empregadas no desenvolvimento do sistema foram:

- **Python:** Linguagem principal para implementação.
- **Whisper:** Reconhecimento de voz baseado em aprendizado profundo (Radford et al., 2022).
- **VLibras:** Tradução de texto para Libras, exibida por meio de avatares (Almeida et al., 2022).
- **FFmpeg:** Processamento de áudio para redução de ruídos e ajuste de formato (Bellard, 2023).
- **Bibliotecas Python:** Incluindo requests, asyncio e FFmpeg-python para integração e manipulação de dados.

## 6.3 CRITÉRIOS DE VALIDAÇÃO

A eficácia do sistema foi avaliada com base em:

- **Precisão:** Avaliação da qualidade das transcrições geradas pelo Whisper e da tradução para Libras realizada pelo VLibras.
- **Desempenho:** Medição do tempo de resposta desde a captura do áudio até a exibição dos sinais traduzidos.
- **Usabilidade:** Feedback de usuários quanto à facilidade de uso e clareza dos sinais exibidos.
- **Robustez:** Testes em condições adversas, como ruído ambiental, para garantir a confiabilidade.

## 6.4 FLUXO DO SISTEMA

O sistema segue o seguinte fluxo operacional:

1. Captura do áudio utilizando FFmpeg para gravação em tempo real.
2. Pré-processamento do áudio para redução de ruídos.
3. Transcrição do áudio em texto utilizando Whisper.
4. Tradução do texto para Libras por meio do VLibras.
5. Exibição dos sinais traduzidos em avatares interativos.

## 6.5 LIMITAÇÕES DO MÉTODO

Apesar do sucesso em alcançar os objetivos, algumas limitações foram observadas:

- Dependência de conexão com a internet para integração com o VLibras.
- Dificuldades em traduzir termos técnicos ou jargões específicos.
- Redução na precisão de transcrição em ambientes com ruído excessivo.

## 7 RESULTADOS.

O sistema inteligente para tradução de áudio em Libras foi desenvolvido e validado em diferentes etapas, demonstrando a eficácia das tecnologias integradas e o cumprimento dos objetivos estabelecidos.

### 7.1 PROCESSAMENTO DE ÁUDIO E TRANSCRIÇÃO

O áudio foi extraído dos vídeos utilizando a ferramenta FFmpeg, que realizou o pré-processamento necessário, como redução de ruído e normalização. Esse processamento garantiu que o áudio estivesse em formato adequado para o modelo Whisper, que alcançou alta precisão na transcrição de áudio em textos.

Em cenários sem ruído, a transcrição foi altamente precisa, com taxas de acerto consistentes. Em ambientes moderadamente ruidosos, a precisão foi reduzida, mas técnicas adicionais de pré-processamento, como filtros de redução de ruído, melhoraram significativamente o desempenho.

### 7.2 TRADUÇÃO PARA LIBRAS

A integração com a ferramenta VLibras permitiu traduzir textos transcritos para Libras de forma fluida e interativa, utilizando avatares digitais.

- Para textos simples e cotidianos, os resultados foram amplamente satisfatórios, com tradução precisa e de fácil compreensão.
- Para termos técnicos ou jargões específicos, algumas limitações foram observadas, com sinais traduzidos de forma genérica ou incompleta.

Esses resultados destacam a adequação do VLibras para cenários de uso cotidiano e apontam a necessidade de adaptações ou extensões para atender a contextos mais especializados.

### 7.3 DESEMPENHO E USABILIDADE

A interface web desenvolvida para o sistema foi testada por usuários finais e mostrou-se intuitiva e eficiente. O feedback dos usuários enfatizou a facilidade de interação com o sistema e a clareza na exibição dos sinais traduzidos.

Além disso, a velocidade de processamento, desde a captura de áudio até a exibição dos sinais, foi avaliada como adequada para aplicações práticas.

#### **Pontos positivos identificados:**

- Clareza e organização da interface.
- Tempo de resposta satisfatório.
- Apresentação precisa e fluida dos sinais em Libras.

#### **Pontos de melhoria identificados:**

- A dependência de conexão com a internet para integração com o VLibras foi apontada como um desafio.
- Tradução limitada para termos técnicos ou menos comuns.

## 7.4 LIMITAÇÕES IDENTIFICADAS

Apesar dos resultados positivos, algumas limitações foram observadas ao longo do desenvolvimento e validação do sistema:

1. **Ambientes Ruidosos:** A precisão do Whisper foi afetada em cenários com excesso de ruído, mesmo após técnicas de pré-processamento.
2. **Termos Técnicos:** O VLibras enfrentou dificuldades na tradução de jargões e termos específicos, resultando em sinais menos compreensíveis.
3. **Conexão com a Internet:** A funcionalidade do sistema depende de uma conexão estável para acessar os serviços do VLibras, limitando sua aplicação em ambientes offline.

## 7.5 DESENVOLVIMENTO DA API E INTEGRAÇÃO

O sistema utilizou o módulo `http.server`, uma solução nativa do Python, para desenvolver um servidor HTTP simples. Esse servidor foi projetado para lidar com requisições HTTP e gerenciar a interação entre o backend e a interface web. A classe personalizada `CustomHTTPRequestHandler` foi responsável por definir as rotas e os comportamentos necessários para processar dados, gerenciar uploads de links de vídeos e exibir as páginas web.

A interface web foi desenvolvida com foco em acessibilidade e usabilidade, proporcionando uma experiência intuitiva para usuários com diferentes níveis de familiaridade tecnológica. A integração entre o servidor HTTP e a interface web foi realizada utilizando JavaScript, permitindo que o sistema verificasse o status do processamento em tempo real e apresentasse os resultados de forma clara e sincronizada.

A tela inicial do sistema foi projetada para ser simples e direta, permitindo que o usuário envie links de vídeos para processamento e, posteriormente, visualize os sinais traduzidos. Além disso, mensagens de progresso foram implementadas para informar o status das operações, como o processamento do vídeo e a transcrição do áudio.

Essa abordagem com `http.server` garantiu simplicidade e eficácia, sendo suficiente para atender aos requisitos do projeto.

## 7.6 CONCLUSÃO

Os resultados obtidos demonstram que o sistema é uma solução viável e funcional para a tradução de áudio em Libras, promovendo a inclusão social de pessoas surdas em ambientes digitais. Apesar de algumas limitações, o sistema oferece avanços significativos em termos de acessibilidade e usabilidade, atendendo a uma necessidade crescente por soluções acessíveis e inclusivas.

As melhorias futuras sugeridas, como suporte offline e aprimoramento na tradução de termos técnicos, podem ampliar ainda mais o impacto e a aplicabilidade do sistema em diversos contextos.

## 8 CONTRIBUIÇÕES DO PROJETO

O sistema oferece contribuições significativas no campo da acessibilidade:

1. **Inclusão Digital:** Facilitando o acesso de pessoas surdas a conteúdos audiovisuais em Libras.

2. **Automatização:** Reduzindo a dependência de intérpretes humanos em situações onde não estão disponíveis.
3. **Adaptação:** A solução pode ser expandida para suportar outros idiomas e contextos.

## 8.1 LIMITAÇÕES E TRABALHOS FUTUROS

Embora o sistema tenha demonstrado bons resultados, algumas limitações foram identificadas:

- **Dependência de Internet:** A integração com o VLibras requer conexão estável.
- **Precisão em Ambientes Ruídosos:** Em cenários com ruídos intensos, a precisão do reconhecimento de voz é reduzida.
- **Tradução de Jargões:** O sistema enfrenta dificuldades ao lidar com termos técnicos ou contextos altamente específicos.

Para trabalhos futuros, as seguintes melhorias são recomendadas:

- **Integração Offline:** Desenvolver funcionalidades que reduzam a dependência de conexão com a internet, utilizando modelos locais.
- **Treinamento Personalizado:** Ajustar o modelo Whisper para lidar com sotaques e ruídos específicos do público-alvo.
- **Expansão Multilíngue:** Adaptar o sistema para suportar traduções de diferentes idiomas para Libras.

Com os avanços contínuos em inteligência artificial e acessibilidade, o sistema proposto tem o potencial de impactar positivamente a vida de muitas pessoas, contribuindo para uma sociedade mais inclusiva e digitalmente acessível.

## REFERÊNCIAS

ALMEIDA, R. et al. **Vlíbras:** Uma abordagem para acessibilidade digital em Líbras, 2022. Disponível em: <https://vlibras.gov.br>. Acesso em: 21 nov. 2024.

ALVES, L. M.; MONTEIRO, G. B. M.; RABELLO, S.; GASPARETTO, M. E. R. F.; CARVALHO, K. M. Recursos de acessibilidade e o uso dos dispositivos móveis como tecnologia assistiva por pessoas com baixa visão. **Revista Brasileira de Educação Especial**, São Paulo, v. 15, n. 3, p. 415-430, set./dez. 2009. Disponível em: <https://www.scielo.br/j/rbee/a/C4GxYprjw5KMcTYB3nfcXSs>. Acesso em: 21 nov. 2024.

AMODEI, D., et al. Deep Speech 2: Reconhecimento de Fala de Fim a Fim em Inglês e Mandarim. **Anais da 33ª Conferência Internacional sobre Aprendizado de Máquina**. Disponível em: <https://arxiv.org/abs/1512.02595>. Acesso em: 21 nov. 2024.

BELLARD, F., **FFmpeg**: A Complete Solution to Record, Convert, and Stream Audio and Video, 2023. Disponível em: <https://ffmpeg.org>. Acesso em: 21 nov. 2024.

COSTA, A. **Tecnologia e inclusão**: O impacto da Hand Talk na acessibilidade. Revista Brasileira de Inclusão Social, 2021.

GOLDSTEIN, E. **Speech and Hearing Research**: From Signal to Speech. Academic Press, 2018.

GOOGLE CLOUD. **Speech-to-Text API Documentation**, 2023. Disponível em: <https://cloud.google.com/speech-to-text>. Acesso em: 21 nov. 2024.

GUARINELLO, A. C.; SANTANA, A. P.; ALMEIDA, M. A. Surdez e mercado de trabalho: Barreiras e possibilidades. **Cadernos de Educação**, Pelotas, n. 42, p. 89-102, 2013. Disponível em: [https://pepsic.bvsalud.org/scielo.php?pid=S1516-36872013000200007&script=sci\\_arttext](https://pepsic.bvsalud.org/scielo.php?pid=S1516-36872013000200007&script=sci_arttext). Acesso em: 21 nov. 2024.

HINTON, G.; DENG, L.; YU, D.; DAHL, G. E.; MOHAMED, A. R.; JAITLY, N. Redes neurais profundas para modelagem acústica no reconhecimento de fala: as visões compartilhadas de quatro grupos de pesquisa. **IEEE Signal Processing Magazine**, v. 29, n. 6, p. 82-97, nov. 2012. Disponível em: <https://ieeexplore.ieee.org/document/6296526>. Acesso em: 21 nov. 2024.

HUANG, J.; HARPER, F. M.; LI, X.; WANG, H. Compreendendo o papel do reconhecimento em agentes inteligentes controlados por voz. In: CONFERÊNCIA CHI SOBRE FATORES HUMANOS EM SISTEMAS COMPUTACIONAIS, 2019, Glasgow. **Anais da Conferência CHI de 2019 sobre Fatores Humanos em Sistemas Computacionais**. Nova York: ACM, 2019. p. 1-12. Disponível em: [https://www.researchgate.net/publication/353590548\\_Recursos\\_de\\_Acessibilidade\\_e\\_o\\_Uso\\_dos\\_Dispositivos\\_Moveis\\_como\\_Tecnologia\\_Assistiva\\_por\\_Pessoas\\_com\\_Baixa\\_Visao](https://www.researchgate.net/publication/353590548_Recursos_de_Acessibilidade_e_o_Uso_dos_Dispositivos_Moveis_como_Tecnologia_Assistiva_por_Pessoas_com_Baixa_Visao). Acesso em: 21 nov. 2024.

IBGE. (2010). Censo Demográfico 2010: Resultados Preliminares sobre Deficiência no Brasil. Instituto Brasileiro de Geografia e Estatística.

IBGE. "Censo 2021: População Surda no Brasil." Instituto Brasileiro de Geografia e Estatística, 2021.

KAUR, D.; SINGH, S.; KAUR, P. Reconhecimento de linguagem de sinais usando deep learning: uma revisão. **International Journal of Engineering Research & Technology**, v. 9, n. 6, p. 1-5, 2020. Disponível em: <https://typeset.io/pdf/lingua-de-sinais-brasileira-estudos-linguisticos-4ydygZR2im.pdf>. Acesso em: 21 nov. 2024.

LIMA, Eliane Maria dos Santos. Tecnologia assistiva no âmbito educacional para o aluno surdo. **Revista Científica Multidisciplinar Núcleo do Conhecimento**, ano 05, ed. 06, v. 06, p. 66-74, jun. 2020. ISSN: 2448-0959. Disponível em: <https://www.nucleodoconhecimento.com.br/educacao/aluno-surdo>. Acesso em: 21 nov. 2024.

LUTZ, M. **Aprendendo Python**. O'Reilly Media, 2013.

OLIVEIRA, T. F.; COSTA, E. S. Tecnologias assistivas e a inclusão de pessoas com deficiência auditiva: um panorama atual. **Revista de Inclusão Digital**, v. 7, n. 3, p. 56-72, 2021. Disponível em: <https://periodicos.utfpr.edu.br/recit/article/viewFile/5188/pdf>. Acesso em: 21 nov. 2024.

PEREIRA, M. S. Inclusão social e as barreiras enfrentadas por pessoas com deficiência auditiva no Brasil. **Revista Brasileira de Educação Especial**, São Paulo, 2019. Disponível em: <https://www.scielo.br/j/pusf/a/kcXkXchtQsnxsgcjTMCm7d/>. Acesso em: 21 nov. 2024.

QUADROS, R. M. de; KARNOPP, L. B. **Educação de Surdos: A aquisição da linguagem e o bilinguismo**. Florianópolis: Mediação, 2020.

QUADROS, R. M. de; KARNOPP, L. B. **Linguagem de sinais brasileira: Estudos linguístico**. Florianópolis: Artmed, 2004.

RABINER, L.; JUANG, B. **Fundamentos do reconhecimento de fala**. Prentice Hall, 1993.

RADFORD, A., KIM, J. W., HALLACY, C., et al. **Whisper: OpenAI's Automatic Speech Recognition System**, 2022. Disponível em: <https://openai.com/research/whisper>. Acesso em: 21 nov. 2024.

RODRIGUES, Marcelo. A UTILIZAÇÃO DO APLICATIVO HAND TALK PARA SURDOS, COMO FERRAMENTA DE MELHORA DA ACESSIBILIDADE NA EDUCAÇÃO. **Anais CIET:Horizonte**, São Carlos-SP, v. 4, n. 1, 2024. Disponível em: <https://ciet.ufscar.br/submissao/index.php/ciet/article/view/1470>. Acesso em: 7 nov. 2024.

TOMAR, S. Convertendo formatos de vídeo com FFmpeg. **Jornal Linux**, 2006. V.10 n.146

VAN ROSSUM, G., DRAKE, F. L., **Python**: Reference Manual, 2011. Disponível em: <https://docs.python.org>. Acesso em: 21 nov. 2024.

WU, Y., et al. O sistema de tradução neural do Google: superando a lacuna entre a tradução humana e a tradução por máquinas. **arXiv preprint**, arXiv:1609.08144, 2016. Disponível em: <https://arxiv.org/abs/1609.08144>. Acesso em: 21 nov. 2024.

WU, Y., SCHUSTER, M., CHEN, Z., et al., **Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation**, 2016. Disponível em: <https://arxiv.org/abs/1609.08144>. Acesso em: 21 nov. 2024.

YOUNG, S.; EVERMANN, G.; GALES, M.; HAIN, T.; KERSHAW, D.; MOORE, G. **The HTK book**. Universidade de Cambridge, 2002.