Ministério da Educação
Universidade Tecnológica Federal do Paraná
Câmpus Curitiba
Oficina De Integração

João Victor Ferreira Castex da Motta
Pedro Schuves Marodin
Rafael Merchiori de Souza

08/04/2025

SignSpeak – Óculos tradutor de libras em áudio

Prof. Dr. César Manuel Vargas Benitez

1 Introdução

A comunicação entre pessoas surdas e ouvintes ainda representa um grande desafio, especialmente em situações cotidianas nas quais não há a presença de intérpretes de LIBRAS (Língua Brasileira de Sinais). Essa limitação compromete a inclusão social e restringe a autonomia das pessoas surdas em contextos como atendimentos médicos, repartições públicas, ambientes educacionais e interações sociais em geral.

Considerando essa realidade, este projeto propõe o desenvolvimento do SignSpeak, um sistema portátil de tradução automática de gestos do alfabeto manual de LIBRAS para áudio em português falado. A solução será implementada em um protótipo de óculos inteligente com câmera embutida, capaz de capturar os gestos realizados pelo usuário. O reconhecimento dos sinais será feito por meio de Redes Neurais Convolucionais (CNNs), que identificarão em tempo real as letras representadas. Após o reconhecimento, o sistema converterá o gesto identificado em áudio, utilizando uma biblioteca de texto-para-fala (TTS), sem a necessidade de conexão com a internet.

O SignSpeak integra técnicas de visão computacional, inteligência artificial e sistemas embarcados, oferecendo uma alternativa prática, acessível e discreta para melhorar a comunicação entre pessoas surdas e ouvintes. A proposta visa não apenas facilitar o diálogo em tempo real, mas também contribuir para a inclusão digital e social, promovendo maior independência e participação dos surdos na sociedade.

Para acompanhar o desenvolvimento do projeto, acesse o blog oficial do SignSpeak.

2 Escopo do Projeto

O projeto SignSpeak tem como objetivo o desenvolvimento de um sistema vestível, integrado a um par de óculos, capaz de realizar a tradução em tempo real dos gestos do alfabeto de LIBRAS para áudio. O sistema será projetado para operar de forma totalmente offline, assegurando seu funcionamento em ambientes sem conexão à internet e garantindo maior privacidade ao usuário.

O foco inicial do projeto está na tradução do alfabeto manual, uma vez que esse conjunto de gestos é padronizado e representa uma base sólida para o desenvolvimento e validação do reconhecimento por meio de CNNs. O sistema será composto por uma câmera embutida nos óculos para capturar os gestos realizados com as mãos, um módulo de processamento baseado em Raspberry Pi para interpretar os sinais por meio da CNN, e um sistema de conversão TTS para vocalizar a letra correspondente.

2.1 Itens dentro do escopo

- Captura de imagens por meio de câmera acoplada aos óculos;
- Processamento das imagens em tempo real com uso de CNNs para reconhecimento do alfabeto manual de LIBRAS;
- Conversão do gesto reconhecido em áudio falado utilizando uma biblioteca TTS offline;
- Integração de todos os componentes em um protótipo funcional e portátil, sem necessidade de conexão com a internet.

2.2 Itens fora do escopo

- Reconhecimento de frases completas ou estruturas gramaticais complexas da LIBRAS;
- Tradução de gestos corporais além do alfabeto manual (ex: expressões faciais, movimento corporal);
- Integração com serviços em nuvem ou banco de dados online.

3 Diagrama em Blocos

Para uma melhor compreensão da arquitetura funcional do sistema, as Figuras 1 e 2 apresentam dois fluxogramas complementares que descrevem as principais etapas do processo.

A Figura 1 traz um diagrama ilustrativo, com imagens que representam de forma intuitiva o fluxo de dados desde a aquisição da imagem até a emissão do áudio correspondente ao gesto identificado. Já a Figura 2 apresenta um fluxograma lógico, estruturado em blocos funcionais que detalham cada componente do sistema de forma mais técnica, facilitando a visualização da arquitetura e da integração entre os módulos.

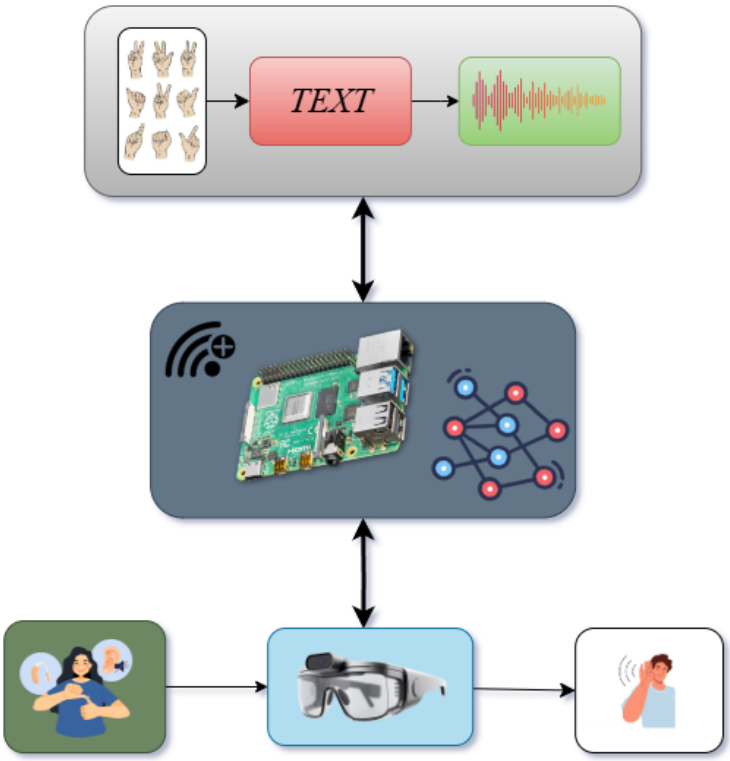

Figura 1 – Fluxograma ilustrativo do sistema SignSpeak.

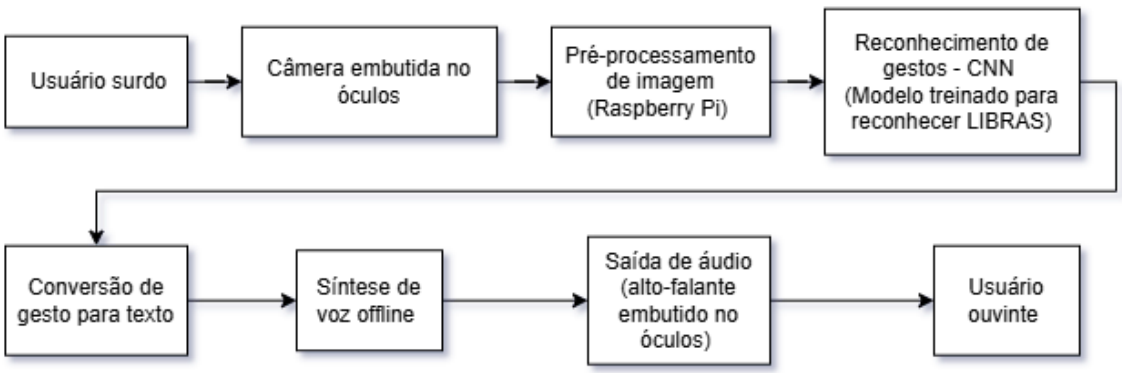

Figura 2 – Fluxograma lógico do sistema SignSpeak.

4 Requisitos Funcionais e Não-Funcionais

Esta seção descreve os requisitos fundamentais para o desenvolvimento do sistema SignSpeak, divididos em funcionais e não-funcionais. Os requisitos funcionais definem os comportamentos e funcionalidades esperadas do sistema, enquanto os não-funcionais estabelecem critérios de desempenho, qualidade e usabilidade.

4.1 Requisitos Funcionais

- Captura de imagem em tempo real: o sistema deve ser capaz de capturar, continuamente, imagens da mão do usuário por meio da câmera embutida no óculos.
- Reconhecimento de gestos do alfabeto manual de LIBRAS: utilizando CNNs, o sistema deve identificar, com precisão, o gesto realizado pelo usuário e associá-lo à letra correspondente.
- Conversão de gesto reconhecido em áudio: após o reconhecimento, o sistema deve transformar a letra detectada em áudio por meio de uma biblioteca de TTS.
- Execução local (offline): todo o processamento de imagem e conversão para áudio deve ser realizado localmente no dispositivo, sem depender de conexão com a internet.
- Reprodução do áudio pelo alto-falante embutido: o som gerado deve ser emitido em tempo real, permitindo a comunicação imediata com ouvintes.

4.2 Requisitos Não-Funcionais

- Baixo consumo de energia: o sistema deve ser energeticamente eficiente, permitindo seu uso prolongado com uma bateria portátil.
- Alta acurácia no reconhecimento dos gestos: o modelo de CNN deve apresentar uma taxa de acerto elevada para garantir a confiabilidade da tradução.
- Tempo de resposta em tempo real: o intervalo entre a captura do gesto e a reprodução do áudio deve ser suficientemente curto para manter a fluidez da comunicação.
- Portabilidade e ergonomia: o sistema deve ser leve, confortável e integrado de forma discreta ao formato dos óculos, favorecendo o uso contínuo no dia a dia.
- Facilidade de manutenção e escalabilidade: o projeto deve ser modular, permitindo futuras atualizações, como expansão do vocabulário ou inclusão de novas funcionalidades.

5 Integração

O projeto SignSpeak envolve a integração de diversos componentes de hardware e software, unindo áreas como eletrônica, inteligência artificial e processamento de imagem em um único sistema vestível. Do ponto de vista eletrônico, o sistema será construído em torno de um Raspberry Pi 4, responsável por controlar a câmera embutida nos óculos, realizar o processamento dos gestos e reproduzir o áudio resultante. A câmera capta imagens dos gestos feitos com as mãos, que são então processadas localmente por uma CNN previamente treinada para reconhecer o alfabeto manual da LIBRAS. Após o reconhecimento do gesto, o Raspberry Pi converte a letra identificada em áudio por meio de uma biblioteca de texto-para-fala offline. Esse áudio é reproduzido por um pequeno alto-falante embutido nos óculos, permitindo a comunicação de forma discreta e eficiente.

A integração eficiente entre hardware e software garante que todas as etapas – captura, processamento e síntese de fala, ocorram de maneira contínua e em tempo real, sem a necessidade de conexão com a internet. Essa abordagem modular também permite futuras expansões do sistema, como o reconhecimento de frases completas ou respostas por voz.

6 Análise de Riscos

Risco	Probabilidade	Impacto	Ação Preventiva / Mitigação
Baixa precisão no reconhecimento dos gestos	Alta	Alto	Melhorar dataset de treinamento, realizar testes com diferentes usuários e iluminações
Ambiente com baixa luminosidade	Média	Médio	Adotar câmera com boa captação em baixa luz; aplicar pré-processamento de imagem (aumento de contraste)
Interferência de ruídos no áudio	Média	Médio	Utilizar TTS offline com cancelamento de ruído e alto-falante direcional
Consumo excessivo de energia	Alta	Alto	Otimizar código, usar modos de economia de energia no Raspberry Pi e componentes de baixo consumo
Desalinhamento da câmera (posição errada no rosto)	Média	Médio	Projeto físico adaptável e fixação estável; testes ergonômicos com usuários reais
Atraso na execução de modelos de IA em tempo real	Média	Alto	Utilizar modelos mais leves, processamento otimizado
Usuários com variações no estilo de sinalização	Alta	Médio	Ampliar diversidade no treinamento; aplicar técnicas de generalização e personalização
Dificuldade de integração entre hardware e software	Média	Alto	Planejar testes modulares por etapas; usar bibliotecas bem documentadas
Falhas de hardware (câmera, speaker, placa)	Baixa	Alto	Realizar testes de estresse e manutenção preventiva; prever componentes de reposição

Tabela 1 – Análise de Riscos e Ações Preventivas

7 Cronograma Detalhado

Período	Atividade	Entregável
07/04 a 08/04	Reunião inicial e divisão de funções da equipe	N/A
08/04 a 09/04	Definição do escopo e requisitos do projeto	N/A
09/04 a 10/04	Criação do conteúdo para o Blog/Site	N/A
10/04 a 11/04	Finalização do Plano de Projeto e montagem do Blog	N/A
11/04	Apresentação do Plano de Projeto e do Blog/Site	Entregável 1 e 2
14/04 a 17/04	Pesquisa e definição dos componentes eletrônicos e mecânicos	N/A
17/04 a 21/04	Levantamento de materiais e início da modelagem da estrutura (óculos)	N/A
21/04 a 25/04	Montagem inicial da estrutura física	N/A
28/04 a 02/05	Montagem do circuito e testes de alimentação, câmera e alto-falante	N/A
05/05 a 09/05	Integração parcial do hardware com Raspberry Pi	N/A
12/05 a 16/05	Ajustes mecânicos e testes de portabilidade	N/A
19/05 a 22/05	Testes finais de hardware/mecânica e documentação	N/A
23/05	Apresentação do desenvolvimento e testes de hardware/mecânica	Entregável 3
24/05 a 30/05	Criação do dataset ou curadoria de base de imagens	N/A
31/05 a 06/06	Treinamento e validação da CNN	N/A
06/06 a 09/06	Implementação do sistema de texto-para-fala offline	N/A
10/06 a 12/06	Testes do software embarcado no Raspberry Pi	N/A
13/06	Apresentação do desenvolvimento e testes do software	Entregável 4
14/06 a 20/06	Integração final: câmera + CNN + TTS + áudio	N/A
21/06 a 27/06	Testes de uso real, ajustes de latência e melhoria de usabilidade	N/A
28/06 a 01/07	Gravação do vídeo demonstrativo e finalização do Blog	N/A
01/07 a 03/07	Escrita e finalização do relatório técnico	N/A
04/07	Demonstração do protótipo, entrega do relatório, vídeo e Blog	Entregável 5
07/07 a 09/07	Preparação da apresentação para banca (slides, roteiro, ensaios)	N/A
10/07	Ensaio final cronometrado com feedback entre os membros	N/A
11/07	Apresentação final com banca avaliadora (10 min apresentação + 10 min banca)	Avaliação Final da Banca

Tabela 2 – Cronograma de atividades do projeto

8 Materiais e Métodos

8.1 Hardware

O protótipo será desenvolvido com componentes de baixo consumo energético e compatíveis com sistemas embarcados, visando portabilidade e integração eficiente. A lista de hardware inclui:

- **Raspberry Pi 4**: microcomputador que atuará como a unidade central de processamento, responsável pela execução da CNN, controle de periféricos e síntese de áudio;
- **Câmera compatível com Raspberry Pi**: utilizada para capturar, em tempo real, imagens dos gestos realizados em frente aos óculos;
- **Alto-falante mini embutido**: componente responsável pela saída de áudio, onde será reproduzida a tradução do gesto reconhecido;
- **Bateria portátil (powerbank)**: fonte de alimentação do sistema, permitindo o funcionamento do dispositivo de maneira autônoma e móvel.

8.2 Software

O sistema embarcado será programado utilizando linguagens e bibliotecas amplamente aplicadas em visão computacional e inteligência artificial. Os principais softwares e ferramentas utilizados são:

- **Python**: linguagem de programação principal, escolhida por sua sintaxe clara e ampla compatibilidade com bibliotecas de visão computacional e redes neurais;
- **OpenCV**: biblioteca utilizada para captura e pré-processamento das imagens obtidas pela câmera, como redimensionamento, normalização e realce de contraste;
- **TensorFlow/Keras**: frameworks de aprendizado profundo utilizados para o treinamento, validação e embarque do modelo de CNN;
- **pyttsx3 ou espeak**: bibliotecas de TTS que operam de forma offline, responsáveis por converter a saída do modelo em áudio compreensível.

8.3 Métodos

A metodologia adotada será dividida em três etapas principais:

1. **Treinamento do modelo de reconhecimento**: será criado ou curado um dataset contendo imagens dos gestos do alfabeto manual de LIBRAS. A CNN será treinada com essas imagens, com ajustes de hiperparâmetros visando maximizar a acurácia da classificação;
2. **Integração embarcada**: o modelo treinado será exportado e embarcado no Raspberry Pi. A câmera capturará os gestos, que serão processados localmente. Após a identificação da letra correspondente, o sistema acionará a biblioteca TTS para sintetizar o som e reproduzir a fala pelo alto-falante;

3. **Testes e validação**: serão realizados testes de desempenho do sistema em diferentes condições de iluminação, ângulo e distância da câmera, a fim de validar sua robustez e confiabilidade em ambiente real.

Essa abordagem permite que todo o processo de tradução, desde a captura do gesto até sua conversão em áudio, ocorra de forma offline, conferindo ao protótipo características como privacidade, portabilidade e independência de conexão com a internet.

9 Conclusão

O projeto SignSpeak propõe uma solução tecnológica inovadora voltada à inclusão social de pessoas surdas, ao permitir a tradução de gestos do alfabeto manual de LIBRAS para áudio em tempo real. A proposta consiste na utilização de CNNs embarcadas em um sistema portátil acoplado a óculos inteligentes, possibilitando o reconhecimento de sinais visuais captados por uma câmera e sua posterior conversão em fala por meio de bibliotecas TTS.

A implementação local e totalmente offline do sistema assegura maior privacidade, independência de conexão com a internet e maior confiabilidade em situações cotidianas. Além disso, a escolha por componentes de baixo consumo energético e de fácil integração, como o Raspberry Pi, torna o protótipo viável tanto tecnicamente quanto economicamente.

Com essa abordagem, o SignSpeak busca não apenas facilitar a comunicação entre surdos e ouvintes, mas também promover maior autonomia, mobilidade e acessibilidade em contextos diversos, como atendimentos públicos, interações sociais e ambientes educacionais.

O protótipo serve como base para futuras expansões, como o reconhecimento de frases completas e gestos corporais mais complexos, além de possíveis implementações de resposta por voz para interação bidirecional. Dessa forma, o projeto representa um avanço significativo na aplicação de tecnologias emergentes em benefício da inclusão e da equidade social, refletindo o compromisso da engenharia com o bem-estar e os direitos humanos.