

Videoconferência Acessível para Deficientes Auditivos com Avatar de Libras Automático

José Leoberto Soares Filho



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, 2024

José Leoberto Soares Filho

Videoconferência Acessível para Deficientes Auditivos com Avatar de Libras Automático

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática do Centro de Informática, da Universidade Federal da Paraíba, como requisito para a obtenção do grau de Mestre em Informática.

Orientador: Guido Lemos de Souza Filho

Abril de 2024

Ficha catalográfica: elaborada pela biblioteca do CI.

Será impressa no verso da folha de rosto e não deverá ser contada.

Se não houver biblioteca, deixar em branco.



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

Dissertação de Mestrado do Programa de Pós-Graduação em Informática intitulado ***Videoconferência Acessível para Deficientes Auditivos com Avatar de Libras Automático*** de autoria de José Leoberto Soares Filho, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof. Dr. Guido Lemos de Souza Filho
Universidade Federal da Paraíba

Prof. Dr. Tiago Maritan Ugulino de Araújo
Universidade Federal da Paraíba

Prof. Dr. Cícero Inacio da Silva
Universidade Federal de São Paulo

Coordenador(a) do Programa de Pós-Graduação em Informática
Prof. Dr. Fernando Matos
CI/UFPB

João Pessoa, 31 de agosto de 2023

Centro de Informática, Universidade Federal da Paraíba
Rua dos Escoteiros, Mangabeira VII, João Pessoa, Paraíba, Brasil CEP: 58058-600
Fone: +55 (83) 3216 7093 / Fax: +55 (83) 3216 7117

DEDICATÓRIA

Dedico este trabalho a toda minha família, em especial para minha mãe, meu pai e minha irmã que, com muito amor e apoio, não mediram esforços para que eu alcançasse esta etapa da minha vida. Reconheço todo o empenho e ajuda da minha namorada e meus amigos próximos, os quais sempre apoiaram nas mais diversas decisões da vida, e me ajudaram trilhar o caminho da graduação e pós da melhor forma possível.

AGRADECIMENTOS

Agradeço a toda minha família, em especial a minha mãe, Lúcia, ao meu pai, Leoberto, minha irmã Renatha, que me promoveram todo o suporte e carinho para alcançar este sonho. Reconheço todo o auxílio da minha queridíssima namorada Yane e amigos que fiz durante todo o meu percurso na Universidade e que levarei para toda minha vida. Gostaria de agradecer de forma especial para o meu orientador Prof. Dr. Guido Lemos por todo o suporte e ajuda, um agradecimento especial a intérprete e mestre Dilainne Albuquerque em Computação que me auxiliou na realização dos experimento, e por fim, mas não menos importante, sou extremamente grato a banca avaliadora pelos conselhos e sugestões. No mais agradeço todo o corpo docente do Centro de Informática da UFPB que tive o prazer de aprender com eles.

RESUMO

Tendo em vista o aumento do uso de conferências síncronas nas atividades diárias da população brasileira, as mais diversas adversidades surgiram ou foram agravadas a partir desse novo, porém já consolidado, meio de comunicação. A falta de acessibilidade para pessoas com algum tipo de deficiência auditiva é um dos desafios mais notáveis, já que essa parcela da sociedade segundo o censo conduzido pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no ano de 2010, representa aproximadamente 9,7 milhões de pessoas na época. Existem diversas leis que garantem o acesso à informação para pessoas surdas, porém, profissionais capacitados para tradução do português para Libras ainda é um grupo escasso. Pensando nisso, este trabalho propõe a inserção do Avatar de Libras para tradução da linguagem oral portuguesa para a linguagem gestual de libras, acrescentando também apresenta a funcionalidade de legendas automáticas no interior da ferramenta Video For Health (V4H). Os resultados obtidos pelos experimentos propostos mostraram que a compreensão de surdos em videoconferência melhorou significativamente ao utilizar a ferramenta desenvolvida em comparação as plataformas disponíveis na internet.

Palavras-chave: <Acessibilidade>, <Libras>, <Avatar de Libras>, <Transcrição> e <Videoconferência>.

ABSTRACT

Given the increased use of synchronous conferences in the daily activities of the Brazilian population, various adversities have arisen or been exacerbated through this new but already established means of communication. The lack of accessibility for individuals with some form of hearing impairment is one of the most notable challenges, because this portion of society, according to the census conducted by the Brazilian Institute of Geography and Statistics (IBGE) in the year 2010, represented approximately 9.7 million people at that time. There are several laws that ensure access to information for deaf individuals, however, professionals qualified for Portuguese to Brazilian Sign Language (Libras) translation are still a scarce group. With this in mind, this work proposes the incorporation of a Libras Avatar for translating spoken Portuguese language into sign language, while also adding the functionality of automatic captions within the Video For Health (V4H) tool. The results obtained from the proposed experiments showed that the understanding of deaf individuals in videoconferences significantly improved when using the developed tool compared to the platforms available on the internet.

Key-words: <Accessibility>, <Libras>, <Libras' Avatar>, <Videoconference> e <Transcription>.

LISTA DE FIGURAS

1	Alfabeto de Libras	24
2	Player do avatar de Libras	25
3	Camadas de Segurança adicionadas à videoconferência.	31
4	Interior da videoconferência	32
5	Sistema V4H Agenda	33
6	Pré-conferência: Câmera e microfone habilitados	34
7	Pré-conferência: Câmera e microfone desabilitados	34
8	Áreas lógicas	34
9	Visualização na disposição de blocos com três participantes	35
10	Modal do chat	36
11	Modal de convite	36
12	Modal de gravação	36
13	Fluxo de troca de mensagens via o <i>WebSocket</i>	38
14	Tela inicial para inserção de dados no <i>ChatGPT</i>	39
15	Exemplos de integração com o ChatGPT através de APIs	39
16	Resultados do questionário para o tópico de <i>Acesso a Ferramenta</i>	43
17	Resultados do questionário para o tópico de <i>Interagindo na Plataforma</i>	44
18	Resultados do questionário para o tópico de <i>Compreensão da Apresentação</i>	45
19	Resultados do questionário para o tópico de <i>Saindo da Reunião</i>	46
20	Resultados do questionário por ferramenta	46
21	Resultados do questionário por ferramenta	47
22	Proposta de inserção do intérprete humano na videoconferência (Araújo, 2021)	48
23	Proposta do menu de acessibilidade (Araújo, 2021)	48
24	Arquitetura de captura da fala e troca de mensagens (Araújo, 2021)	49
25	Tabela comparativa entre a solução proposta e demais ferramentas de videoconferência (Araújo, 2021)	50
26	Tela principal do Dicionário da Língua Brasileira de Sinais (Colling und Boscarioli)	51

27	Tela principal do Jogo Multi-Trilhas (Colling und Boscaroli)	51
28	Tradutores Automáticos Português-Libras (Colling und Boscaroli)	52
29	Resultado da Corretude dos Parâmetros da Libra Para os Tradutores Automáticos de Libras (Colling und Boscaroli)	53
30	Proposta de Processo Para Tradução Português-Libras (Colling und Boscaroli)	53
31	Estágios do conversor de fala para língua de sinais (Patel u. a., 2020)	54
32	Exemplo de análise e quebra de cada frase (Patel u. a., 2020)	54
33	Transição do Inglês para o Hindi (Patel u. a., 2020)	55
34	Dicionário de animações para <i>ISL</i> (Patel u. a., 2020)	56
35	Resultados da acurácia e tempo de processamento em comparação com sistemas presentes na literatura (Patel u. a., 2020)	56
36	Diagrama do funcionamento das legendas e tradução para avatar em tempo real (Fonte: Imagem própria)	58
37	Proposta do componente do Avatar de Libras (Fonte: Imagem própria)	59
38	Proposta do componente do Avatar de Libras quando o botão de "Visualizar texto" estiver ativado (Fonte: Imagem própria)	60
39	Proposta do componente de legenda (Fonte: Imagem própria)	60
40	Proposta da barra de ações (Fonte: Imagem própria)	61
41	Proposta da barra de ações quando acessibilidade automática estiver ativada (Fonte: Imagem própria)	61
42	Proposta da interface completa (Fonte: Imagem própria)	62
43	Resultado da implementação baseada na proposta da interface (Fonte: Imagem própria)	62
44	Legenda com no máximo três linhas e rolagem (Fonte: Imagem própria)	63
45	Troca de mensagens entre clientes e servidor (Fonte: Imagem própria)	65
46	Documentação <i>Swagger</i> da API do sumariador (Fonte: Imagem própria)	66
47	Documentação <i>Swagger</i> do <i>TSS</i> (Fonte: Imagem própria)	67
48	Visualização dos dados do <i>TSS</i> (Fonte: Imagem própria)	67
49	Visão geral do estágio de definição do experimento (Wohlin u. a., 2000)	68
50	Visão geral do estágio de planejamento do experimento (Wohlin u. a., 2000)	70

51	Nível de deficiência auditiva dos avaliadores	73
52	Aplicação do experimento para o solução ICA	75
53	Quantidade de avaliadores que afirmaram conseguir habilitar o avatar dentro da solução proposta	77
54	Contribuição do Avatar de Libras para a compreensão da apresentação	78
55	Velocidade do Avatar de Libras	78
56	Velocidade do Avatar de Libras	79
57	Gráfico <i>box plot</i> para visualização de dispersão dos resultados dos experimentos	81

LISTA DE TABELAS

1	Exemplo de tradução do Português para Glosa	26
2	Web Speech API: Configurações	30
3	Funcionalidades na interface de videoconferência	37
4	Configuração utilizada da OpenAI	66
5	Resultados do total de participantes que marcaram "sim" para cada questão na respectiva solução	76

LISTA DE ABREVIATURAS

TCP – Transmission Control Protocol

Libras - Língua Brasileira de Sinais

HTTP - Hypertext Transfer Protocol

IHM - Interação Homem-Máquina

V4H - Video For Health

LS - Língua de Sinais

IA - Inteligência Artificial

ISL - Língua de Sinais Indiana

ITV - Interfaces Tradicionais de Videoconferência

ICI - Interface com Intérprete

ICA - Interface com Avatar

Sumário

1 INTRODUÇÃO	19
1.1 Definição do Problema	20
1.2 Premissas e Hipóteses	21
1.2.1 Objetivo geral	22
1.2.2 Objetivos específicos	22
1.3 Estrutura da monografia	23
2 REFERENCIAL TEÓRICO	24
2.1 Libras	24
2.2 Avatar de Libras - WebAssista	25
2.3 Acessibilidade	28
2.4 Transcrição de Áudio	29
2.4.1 Web Speech API - Speech Recognition	30
2.5 Video For Health (V4H)	31
2.6 WebSocket	37
2.7 OpenAI ChatGPT	38
3 REVISÃO BIBLIOGRÁFICA	40
3.1 Protocolo de Busca	40
3.2 Revisão Qualitativa dos Principais Trabalhos	40
3.2.1 Pesquisa sobre acessibilidade das ferramentas de videoconferência em plataforma web	40
3.2.2 Uma proposta de interface de videoconferência acessível para pessoas com deficiência auditiva	47
3.2.3 Avaliação de Tecnologias de Tradução Português-Libras Visando o Uso no Ensino de Crianças Surdas	51
3.2.4 ES2ISL: An advancement in speech to sign language translation using 3D avatar animator	54
4 Descrição da Solução Proposta	57

4.1	Arquitetura	57
4.2	Proposta de Interface Gráfica	58
4.2.1	Componente do Avatar de Libras	58
4.2.2	Componente de legenda	60
4.2.3	Barra de tarefas	61
4.2.4	Proposta final	61
4.3	Avatar de Libras	62
4.4	Legendas	63
4.5	Servidor WebSocket	64
4.6	Serviço de Sumarização de texto	65
4.7	Serviço de Armazenamento de Transcrições	66
5	CENÁRIO E CONFIGURAÇÕES DOS EXPERIMENTOS	68
5.1	Definição do experimento	68
5.1.1	Objetivo do estudo	68
5.1.2	Propósito	68
5.1.3	Foco de Qualidade	69
5.1.4	Perspectiva	69
5.1.5	Contexto	69
5.2	Planejamento do experimento	69
5.2.1	Seleção de Contexto	70
5.2.2	Formulação de Hipótese	70
5.2.3	Seleção de Variáveis	71
5.2.4	Seleção de Sujeitos	71
5.2.5	Design do Experimento	71
5.2.6	Instrumentação	71
6	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	73
6.1	Execução do Experimento	73
6.2	Análise dos dados	75

6.3 Avaliando hipóteses	79
7 CONCLUSÕES E TRABALHOS FUTUROS	82
REFERÊNCIAS	82
APÊNDICE A	86

1 INTRODUÇÃO

No ano de 2020, durante a pandemia de COVID-19, a utilização de conferências online tornou-se um dos principais meios de comunicação para trabalho, estudos e acesso a serviços de saúde, como apresentado nos estudos de (Khattar u. a., 2020). Sendo assim, serviços caracterizados por serem realizados de forma presencial, foram adaptados para o modelo remoto ou híbrido, a fim de atender às novas demandas.

Segundo Harry Moseley, CIO da Zoom, uma das principais empresas de videoconferências do mundo, em entrevista para (Richard, 2020), em dezembro de 2019 a companhia atendia 10 milhões de usuários diariamente e passou a atender cerca de 300 milhões de usuários por dia em abril de 2020, um aumento em torno de 2900% no uso da plataforma.

Bem como, em consequência do distanciamento social e *lockdowns* causadas pela pandemia, o Conselho Nacional de Justiça (CNJ) aponta que no período de 1º de maio de 2020 e 4 de agosto de 2020, o Poder Judiciário realizou mais de 360 mil videoconferências, como apurado pela repórter (Karine, 2020).

No contexto internacional da saúde, (Emilio, 2020) afere que na segunda semana de abril de 2020 foram realizadas aproximadamente 1 milhão de teleconsultas na França, de modo semelhante do que aconteceu no Reino Unido, onde das 1 milhão de consultas diárias, maior parte foi conduzida a distância.

No mundo pós-pandemia, as atividades que foram atualizadas para o modelo remoto, permaneceram em sua boa parte sendo realizadas assim, por exemplo, boa parte das audiências jurídicas permaneceram acontecendo no modelo remoto, bem como a prova de vida realizada pelo INSS. Consequentemente, as videoconferências permanecem presentes no dia a dia das pessoas. Sendo assim, alguns problemas, já conhecidos, começam a ressurgir. Dentre eles, podemos citar segurança, escalabilidade e acessibilidade como alguns dos principais percalços, inclusive legais.

Dentre as adversidades geradas pelo uso das videoconferências, a falta de acessibilidade para pessoas com algum tipo de deficiência auditiva se caracteriza entre as mais críticas. Essa parcela da sociedade, segundo o censo conduzido pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no ano de 2010, é composta por aproximadamente 9,7 milhões de pessoas (Felipe, 2020). No nível mundial, segundo a (Organization, 2021) (OMS), esse número pode ultrapassar mais de 500 milhões de pessoas.

Para a comunicação entre deficientes auditivos no Brasil, desde 2002, através da Lei nº 10.436, a Língua Brasileira de Sinais (Libras) é reconhecida legalmente como meio de comunicação e expressão (CRISTIANO, 2015). Esta linguagem é uma modalidade gesto-visual, na qual são utilizados gestos, expressões faciais e corporais com a finalidade

de viabilizar a comunicação entre pessoas surdas.

Atualmente, já existem ferramentas para a tradução automática da linguagem oral para as linguagens de sinais, porém a utilização dos intérpretes humanos ainda é a melhor maneira de realizar as traduções em tempo real. Porém, com o passar dos anos a tendência é que as ferramentas de tradução automática sejam constantemente aprimoradas, a fim de minimizar as perdas de informações entre a língua de origem e a linguagem de destino, ofertando assim, soluções para casos que o intérprete humano de Libras não esteja disponível.

Posto isto, o presente trabalho tem a finalidade de apresentar uma ferramenta de videoconferência acessível com tradução automática da linguagem oral no idioma português para a língua gestual de Libras. Além disso, esta dissertação busca fazer uma análise qualitativa da eficácia da ferramenta de tradução automática aplicada em reuniões remotas de propósito geral.

1.1 Definição do Problema

Pessoas surdas, mesmo com longos períodos de escolarização, demonstram dificuldades no uso da linguagem escrita, isso se dá ao fato das línguas de sinais (LS) serem a forma de comunicação natural entre os surdos (de Góes, 2020).

Segundo a legislação brasileira disposta na lei 10.436/02 e Decreto 5.626/05, todos os discentes que comprovadamente possuam algum tipo de deficiência auditiva devem ser acompanhados de intérpretes durante o exercício das suas atividades como alunos das instituições de ensino público federal. Porém, como citado em reportagem da (Ribeiro, 2019) vinculada pela Agência Câmara de Notícias, por questões burocráticas ou pela falta de mão de obra qualificada para exercer essa função, esse grupo de pessoas acaba por não receber a devida assistência garantida por lei.

No entanto, o âmbito dessa lei só inclui a obrigatoriedade para as instituições federais, para as demais escolas, públicas e privadas, apenas é recomendada a utilização desses profissionais. Nessas instituições, muitos alunos com deficiência auditiva precisam enfrentar diariamente a falta de intérpretes e a falta de preparo na tradução para língua de sinais (LS). Como consequência, a taxa de evasão escolar de surdos chegou a ser de 97% no censo do ano 2000 realizado pelo IBGE. Atualmente esse número vem caindo ano após ano, com ajuda principalmente das políticas públicas.

Com a pandemia de COVID-19, os desafios de ensino a distância para alunos com deficiência auditiva se mostraram claros. Isto pode ser observado através da pesquisa realizada pela *Datafolha*, na qual afirma que durante a pandemia 90% dos alunos com algum tipo de deficiência não tiveram nenhuma aula com recursos de acessibilidade.

No contexto onde muitas interações interpessoais são realizadas a distância, o projeto (V4H, 2019) tem como objetivo construir uma plataforma de vídeo síncrono para a área da saúde, implementando todos os aspectos de segurança necessários: transmissão confidencial, gravação criptografada, assinaturas digitais e integridade dos metadados das conferências. O qual utiliza o (Jitsi Meet, 2020) como servidor de videoconferência de propósito geral e de código aberto como ferramenta base.

O *Jitsi Meet*, assim como outros servidores de videoconferência, como o Google Meet¹ possui apenas legendas geradas em tempo real, e muitas vezes, esses recursos só aparecem para planos ou versões pagas das plataformas. Esta funcionalidade colabora para a utilização de videoconferência em ambientes ruidosos e para pessoas com algum tipo de deficiência auditiva.

Porém, segundo o autor (Capovilla u. a., 2005), uma pessoa surda, principalmente os do grupo não oralizado, ou seja, que apenas teve contato com a língua de sinais (LS), demonstram um desempenho inferior na compreensão de textos escritos ao serem comparados com surdos oralizados ou alunos ouvintes, devido à dificuldade de surdos em fazer conferência perilexical, sua leitura mostrou-se dependente de mecanismos visuais diretos de reconhecimento e acesso ao significado.

Levando em consideração que a língua de sinais (LS) é o principal e mais eficaz meio de comunicação entre e com os surdos; a falta de mão de obra capacitada para a realização de traduções; e a utilização diária de videoconferências para diversos contextos em um mundo pós-pandemia. Levanta-se a necessidade de buscar ferramentas de tradução automática para Libras e transcrição em tempo real aplicadas no contexto das videoconferências, deste modo, motivando o presente trabalho a apresentar uma solução adequada e eficaz para esta demanda.

1.2 Premissas e Hipóteses

Este trabalho, baseado na definição do problema apresentadas na seção anterior, busca constatar que implementação proposta irá melhorar a inclusão de um usuário surdo em uma videoconferência de propósito geral, em vista disto, a seguinte premissa será utilizada:

1. Língua de sinais (LS) é o modo mais eficaz de comunicação para surdos

Dessa forma, as hipóteses que serão desenvolvidas e testadas durante a presente dissertação pode ser definida formalmente da seguinte forma:

¹<https://meet.google.com/>

Hipótese nula H_0 : A implementação proposta não ajudará na inclusão e compreensão do conteúdo pelos de surdos em videoconferências através de um avatar de Libras automático

Hipótese alternativa H_1 : A implementação proposta ajudará na inclusão e compreensão do conteúdo pelos de surdos em videoconferências através de um avatar de Libras automático

Na seção 5, será realizado um conjunto de experimentos a fim de testar as hipóteses apresentadas anteriormente, apontando os resultados e considerações finais nas seções subsequentes.

1.2.1 Objetivo geral

O objetivo geral deste trabalho é apresentar uma solução que permita a inclusão de pessoas surdas em videoconferências, no contexto das reuniões de propósito geral, nas quais os interpretes e tradutores humanos encontram-se indisponíveis.

Para este propósito, será entregue uma ferramenta de transcrição de áudio em tempo real acoplada a um avatar de Libras para a realização da tradução simultânea das falas individualizadas de cada participante da conferência.

1.2.2 Objetivos específicos

Os objetivos específicos deste trabalho podem ser visualizados abaixo, respectivamente, como:

- Construir servidor *WebSocket* para a troca de mensagens entre os usuários presentes na conferência;
- Utilizar a ferramenta de reconhecimento de fala para transcrever o que for dito pelos usuários, a fim da geração das legendas e tradução para libras;
- Encapsular em um serviço *REST* a sumarização baseada em ferramentas de Inteligência Artificial (IA) previamente já desenvolvidas;
- Desenvolver uma sistema de armazenamento e exibição das transcrições captadas durante as videoconferências realizadas neste trabalho;
- Inserir o Avatar de Libras na interface do sistema de videoconferência Video For Health (V4H), sinalizando a quem pertence a fala que está sendo sinalizada no momento.

- Apresentar a legenda das falas dos usuários, através das transcrições obtidas individualmente e repassadas para os demais participantes;
- Validar a ferramenta de acessibilidade proposta com pessoas surdas, através de questionários relacionados ao entendimento mediante o Avatar de Libras e legendas.

1.3 Estrutura da monografia

A presente dissertação é composta por sete seções, a primeira delas foi utilizada para apresentar a introdução e definição do problema, primeiros conceitos importantes para o entendimento da área, hipóteses e objetivos gerais e específicos do trabalho.

No **seção 2** são discutidos conceitos gerais sobre o sistema de videoconferência utilizado, avatar de libras automático, linguagens de sinais, entre outros assuntos pertinentes ao trabalho.

No **seção 3** é contemplado um levantamento dos trabalhos relacionados, a fim de apresentar referências e realizar um comparativo com soluções existentes e que inspiraram a metodologia adotada no presente trabalho.

No **seção 4**, é tratada de forma detalhada a implementação do componente de tradução automática de linguagem oral para linguagem de sinais aplicado no sistema de videoconferência.

No **seção 5**, são apresentados os cenários de teste, configurações dos experimentos e descritivo detalhado das hipóteses e instrumentos de validação para as mesmas.

No **seção 6**, é realizado estudo dos experimentos conduzidos, descrevendo os resultados da utilização de tradução automática para Libras em videoconferências.

Por fim, no **seção 7**, são apresentadas as conclusões obtidas e um conjunto de propostas para a realização de trabalhos futuros.

2 REFERENCIAL TEÓRICO

Esta seção traz conceitos fundamentais sobre videoconferências e arquiteturas de projetos que geram novos requisitos para o sistema proposto, com o objetivo de facilitar o entendimento deste trabalho.

2.1 Libras

Libras é sigla para Língua Brasileira de Sinais que consiste em uma linguagem gesto-visual onde a comunicação é realizada através de gestos, expressões faciais e corporais. A língua de sinais é também conhecida como língua gestual (CRISTIANO, 2015). A mesma utiliza-se de gestos e sinais em substituição à língua que todos nós bem conhecemos em nossas comunicações: a língua de sons ou oral.

A língua de Libras tem um alfabeto próprio como mostrado na figura 1, assim como um repleto vocabulário de gestos e expressões faciais desenvolvidas e padronizadas durante as últimas décadas.

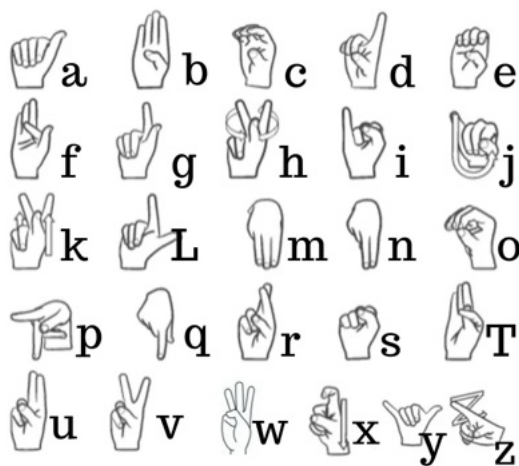


Figura 1: Alfabeto de Libras
Fonte: Brasil Escola

No contexto da acessibilidade, é importante e necessário enxergar a materialidade linguística da Libras. Reconhecida como meio legal de comunicação das comunidades surdas brasileiras pela Lei n. 10.436, de 24 de abril de 2002², é composta por morfologia, sintaxe, semântica e pragmática, características estas que a definem enquanto língua. Apesar disso, por se materializar a partir da condição de pessoas com deficit auditivo, a Libras não se apresenta na modalidade oral-auditiva, mas na gesto-visual.

²http://www.planalto.gov.br/ccivil_03/Leis/2002/L10436.htm

(Quadros und Becker Karnopp, 2007) ressaltam a riqueza e complexidade da Libras, que assim como uma língua oral, pode ser marcada pelas variações regionais e por fatores culturais e socioeconômicos. As autoras ainda exploram os recursos utilizados na língua: movimentos, espaço e expressões faciais, além dos próprios aspectos gramaticais existentes.

Os dados divulgados pelo Instituto Brasileiro de Geografia e Estatística em 2010 revelaram aproximadamente 9,7 milhões de pessoas com algum tipo de deficiência auditiva³, número que representava, à época, mais de 5% da população do país. Mantendo esse cenário em vista, é incoerente dissociar as comunidades surdas das comunidades ouvintes, ambas existem e se relacionam nas diferentes camadas da sociedade. Torna-se então oportuno reiterar a legitimidade do árduo caminho enfrentado pela parcela não ouvinte, desde o diagnóstico até o domínio da língua e a convivência com os preconceitos.

Assim como as demais, a língua de sinais é adquirida através da interação com pessoas que a dominam. Logo, tanto sob a perspectiva do ensino e aprendizagem da Libras, quanto sob o viés da inclusão e do desenvolvimento da sociedade torna-se crucial que sejam desenvolvidos instrumentos que universalizam o acesso à comunicação.

2.2 Avatar de Libras - WebAssista

O Avatar de Libras, denominado WebAssista, é um conjunto de ferramentas que traduz conteúdos digitais (texto, áudio e vídeo) em português para Libras, tornando computadores, celulares e plataformas Web mais acessíveis para as pessoas com algum tipo de deficiência auditiva.



Figura 2: Player do avatar de Libras

³http://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf

Para isso, existe um processamento do lado dos servidores, que recebem um conjunto de palavras e frases em português e em seguida transforma esse texto em glosa, a qual é repassada para o avatar desenvolvido através do motor de jogos e animações *Unity*, como mostra a figura 2.

Tabela 1: Exemplo de tradução do Português para Glosa

Texto original	Glosa
João está perdido na floresta encantada	João estar perder floresta encantar
Bom dia, como estão vocês?	Bom dia como estar vocês [Interrogação]
Na aula de hoje vamos falar sobre as tecnologias mais recentes da internet	Aula hoje ir falar tecnologia mais recente internet
Esta dissertação busca melhorar a inclusão dos surdos em videoconferências	Esta dissertação buscar melhorar inclusão surdo videoconferência

Glosas são palavras de uma determinada língua oral grafadas com letras maiúsculas que representam sinais manuais de sentido próximo. Segundo (Wilcox, 2022), glosa é uma tradução simplificada de morfemas da língua sinalizada para morfemas de uma língua oral.

Na tabela 1 é possível observar exemplos de tradução do português para a glosa. Vale salientar que o texto passa por um componente tradutor que aplica regras semânticas específicas para a utilização em língua de sinais (LS).

Segundo (Agência Brasil, 2021), existem cerca de 370 mil palavras no vocabulário ortográfico da língua portuguesa, cerca de 21 mil sinais catalogados pelo software de tradução do governo federal (Brasil, Agência, 2021) *VLibras* e cerca de 14 mil sinais no software de tradução da *WebAssista*, utilizado neste trabalho.

Com essa disparidade da quantidade de palavras presentes no vocabulário ortográfico da língua portuguesa para a quantidade de sinais presentes nos sistemas de tradução automático, muitas palavras precisam utilizar a datilologia, que consiste na soletração de letra a letra da palavra na qual o sinal não foi mapeado.

O *player* de Libras *WebAssista* disponibiliza algumas funcionalidades para o desenvolvedor e usuários:

1. **Fila de tradução:** Consiste em empilhar *strings*, palavras e textos, para que o avatar traduza em sequência;
2. **Pausar:** Pausar a animação da tradução, e retornar se desejado;

3. **Interromper:** Interromper a animação e zerar a fila de frases pendentes;
4. **Alterar velocidade do avatar:** Alterar velocidade da animação do avatar, disponibilizando três velocidades: 1x, 2x e 3x.

O avatar de Libras Assista disponibiliza um conjunto de métodos que serão utilizados no decorrer desta dissertação:

- `setLicense()`: Para ter acesso completo à API, o primeiro passo deve ser adicionar a licença de software, para isso, deve ser executada a função de definir a licença.
- `startPlay()`: Método utilizado para instanciar o player de libras em uma página web, para isso deve ser enviado como parâmetro o id, que corresponde ao ID da *div HTML* em que o player será renderizado.
- `play()`: Função que ao passar o texto como parâmetro, iniciará o processo de tradução para libras. Essa função apresenta 2 parâmetros: o primeiro é o texto que o usuário deseja traduzir e o segundo parâmetro (`append`) é responsável por definir se o player deve adicionar a mensagem para ser reproduzida ao final da reprodução atual ou iniciar imediatamente.
- `stop()`: Função que para a reprodução atual descartando o status da tradução, não existe a possibilidade de retornar para a tradução anterior após executar esse método.
- `pause()`: Função que pausa a reprodução atual, mantendo o status da tradução. Deste modo, será possível retornar para a tradução anterior.
- `repeat()`: Método responsável por repetir a tradução da última frase que foi reproduzida no player.
- `resume()`: Essa função continua a reprodução atual a partir do ponto de parada da reprodução, ou seja, após a chamada da função `pause()`, retomando a reprodução do conteúdo.

Durante o processo de tradução do português para libras, existem eventos para notificar o cliente sobre os status das ações executadas, ou seja, todo o processo de início, fim, e problemas serão reportados para o cliente com a finalidade de receberem tratamento devido pela aplicação cliente. Abaixo segue a lista dos principais eventos do Avatar da *WebAssista*:

- `onPlayerStart`: Evento quando uma nova instância do player for adicionada na página web.

- `onPlayerStartComplete`: Evento quando uma nova instância do player for adicionada e todos os recursos do player foram carregados corretamente.
- `onPlayerStartError`: Evento quando o *Unity* player não for suportado por um browser do cliente.
- `onPlayerNotSupported`: Evento quando o *Unity* player não carregar todos os recursos corretamente.
- `onPlayerStartPlay`: Evento quando uma nova reprodução for iniciada pelo avatar de libras.
- `onPlayerFinishPlay`: Evento quando uma reprodução da *string* repassada para tradução for completamente finalizada.
- `onPlayerPlayError`: Evento quando houver erro na tentativa de reprodução de uma tradução, pode ser ocasionada por problemas de conexão, ou problemas nos servidores.

2.3 Acessibilidade

A acessibilidade é um princípio fundamental, cujo objetivo é garantir que todas as pessoas tenham a possibilidade de desfrutar completamente espaços, serviços, produtos e informações, independente de quais forem suas capacidades físicas, sensoriais ou cognitivas. É coerente que esse conceito seja abrangente, visto que pretende combater justamente qualquer tipo de exclusão de pessoas com deficiências.

Ela deve representar um compromisso ético com a inclusão e a igualdade, visto que, para (Diniz u. a., 2009) ainda é recente o reconhecimento do corpo com deficiência como expressão da diversidade humana, o que configura um desafio para a sociedade.

A importância desse tema transcende a banalização, muito comum, que reduz a acessibilidade à mera adaptação física de espaços e rampas. “O conceito não se operacionaliza apenas na livre circulação espacial. Ele também convoca à arena dimensões comunicacionais, metodológicas e, sobretudo, atitudinais.” (Berselli und Magno, 2021).

A acessibilidade não se limita às deficiências visíveis, ela abrange uma ampla gama de necessidades, como a mobilidade reduzida, as deficiências auditivas, visuais, cognitivas e outras. Por isso, torna-se complexo assegurar que todas as pessoas tenham a oportunidade de participar plenamente da vida cotidiana, desde acessar edifícios até utilizar serviços digitais de forma universalizada.

Dentre as deficiências sensoriais, a deficiência auditiva merece destaque. Indivíduos com essa condição enfrentam desafios que dizem respeito à comunicação e à interação

com o ambiente. A ausência ou redução da capacidade auditiva pode levar ao isolamento social, dificuldades educacionais e profissionais, além de barreiras para acessar informações cruciais. É nesse contexto que a acessibilidade se torna vital.

A acessibilidade para pessoas com deficiência auditiva significa a adoção de medidas que permitam sua participação plena, independentemente do nível de audição. Isso pode envolver a disponibilização de legendas em vídeos, tradutores de língua de sinais em eventos, entre outras ações.

Para atingir níveis significativos de acessibilidade, é desejável que sejam adotadas abordagens multidisciplinares e colaborativas que envolvam arquitetos, designers, engenheiros, legisladores e especialistas em tecnologia assistiva ou outros profissionais, a depender do caso.

2.4 Transcrição de Áudio

A transcrição de áudio consiste no processo de transformação de um áudio em texto, podendo ser um áudio previamente gravado ou um áudio em tempo real. De forma mais abrangente, uma transcrição pode incluir no seu resultado final os ruídos, músicas e falas, isso pode ser influenciado pelos algoritmos usados e no tempo disponível para a realização da transcrição.

Muitas ferramentas já disponibilizam a transcrição do áudio, como é o caso do aplicativo de gravação de áudios da (Microsoft, 2022), da ferramenta para navegador e sistema mobile android (Web APIs — MDN, 2022), entre outros fornecedores, a exemplo do Google, Amazon, etc.

Quando se trata apenas da fala, o termo utilizado é reconhecimento de fala (*Speech Recognition*), naturalmente, esse tipo de aplicação possui limitações em seu vocabulário e são condicionadas a uma boa clareza na fala do usuário, porém com o advento das inteligências artificiais, as ferramentas estão evoluindo diariamente, sendo cada vez mais possível lidar com vários sotaques e pouca qualidade dos microfones.

Um software transforma o som do microfone para alguma linguagem escrita através de 4 etapas:

1. Análise do áudio
2. Divisão em pequenos trechos
3. Digitalização em um formato específico para computador
4. Utilização de algoritmos de reconhecimento de fala que transforma esse trecho em palavras ou frases

Nos primórdios, a transcrição automática de áudio era simplificada apenas a tentar transformar pequenos trechos palavra por palavra. Atualmente os softwares já estão sendo desenvolvidos e treinados para entender diferentes padrões de fala, sotaques, idiomas, dialetos e frases, além da realização de filtragem de ruídos e tratamentos na voz, para aumentar clareza.

Para isso, existem dois modelos separados para o reconhecimento da fala:

- **Modelo Acústico:** Representam a relação entre as unidades linguísticas (palavras, fonemas, etc) e os sinais de áudio.
- **Modelo de Linguagem:** Os resultados dos modelos acústicos são combinados em sequências de palavras e através do contexto, alterar, se necessário, para palavras que se encaixam melhor.

2.4.1 Web Speech API - Speech Recognition

A *Web Speech API* é uma ferramenta disponibilizada para navegador **Chrome** e sistemas **Android**, que oferece nativamente as funcionalidades de reconhecimento e síntese da fala.

O reconhecimento de fala consiste em receber o sinal de áudio através do microfone de um dispositivo, o qual é checado por um serviço de reconhecimento de fala (externo para *Google Chrome* e interno para sistemas *Android*) utilizando um vocabulário específico. Após essa validação, é retornado como resultado uma *string*.

Tabela 2: Web Speech API: Configurações

Configuração	Descrição
grammars	Lista de palavras que serão utilizadas como vocabulário pelo algoritmo
continuous	Controla se os resultados serão contínuos ou se será realizada apenas uma única transcrição
lang	Linguagem do reconhecimento (pt-BR, en-US)
interimResults	Define se o algoritmo vai retornar resultados parciais ou se apenas retornará frases finalizadas
maxAlternatives	Define o número de alternativas de correspondências que devem ser retornados por resultado

Na tabela 2, são apresentadas as configurações para o sistema de reconhecimento, além delas, é possível adicionar um vocabulário específico, definido pelo usuário do sistema, assim como adicionar palavras que devem ser vistas com mais atenção pelo algoritmo.

A biblioteca também disponibiliza um conjunto de métodos que são importantes para o desenvolvimento de aplicações que a utilizam, são eles:

- `start()`: Método utilizado para iniciar o reconhecimento da fala através do microfone.
- `stop()`: Método utilizado para encerrar o reconhecimento de fala.
- `onresult()`: Método de *callback* que é acionado quando um resultado é gerado pela ferramenta.
- `onspeechend()`: Método de *callback* que é acionado quando o reconhecimento é temporariamente finalizado.

2.5 Video For Health (V4H)

O projeto V4H - Vídeo for Health está sendo desenvolvido em parceria da Universidade Federal da Paraíba, a Rede Nacional de Ensino e Pesquisa e a empresa privada *Wisecare*, com a intenção de implementar uma plataforma de videoconferência que atende às demandas da área da saúde.

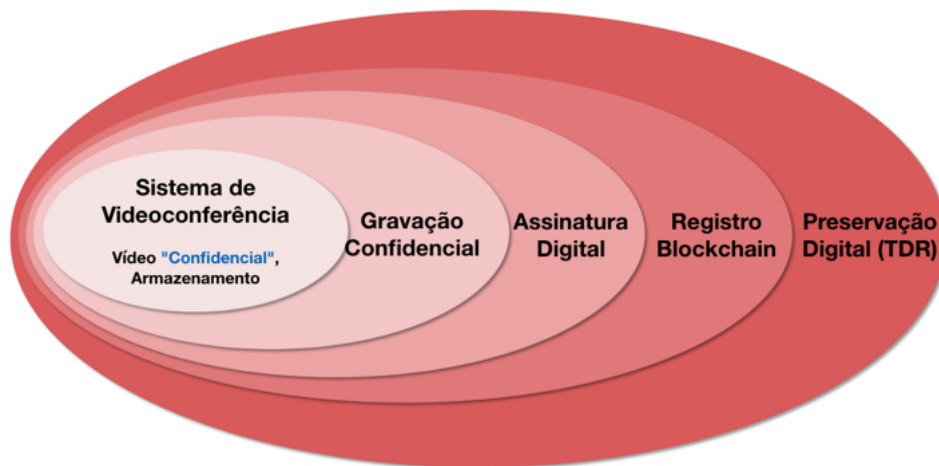


Figura 3: Camadas de Segurança adicionadas à videoconferência.

Fonte: Apresentação V4H - Video For Health

A iniciativa V4H utiliza o *Jitsi Meet* como do sistema de videoconferência base, já que o mesmo é um projeto de código aberto e possui uma larga comunidade que contribui no código. Porém, o intuito é criar uma série de camadas de segurança e de preservação

para as sessões de vídeo síncrono, que são requisitos essenciais da área médica, como mostra a Figura 3.

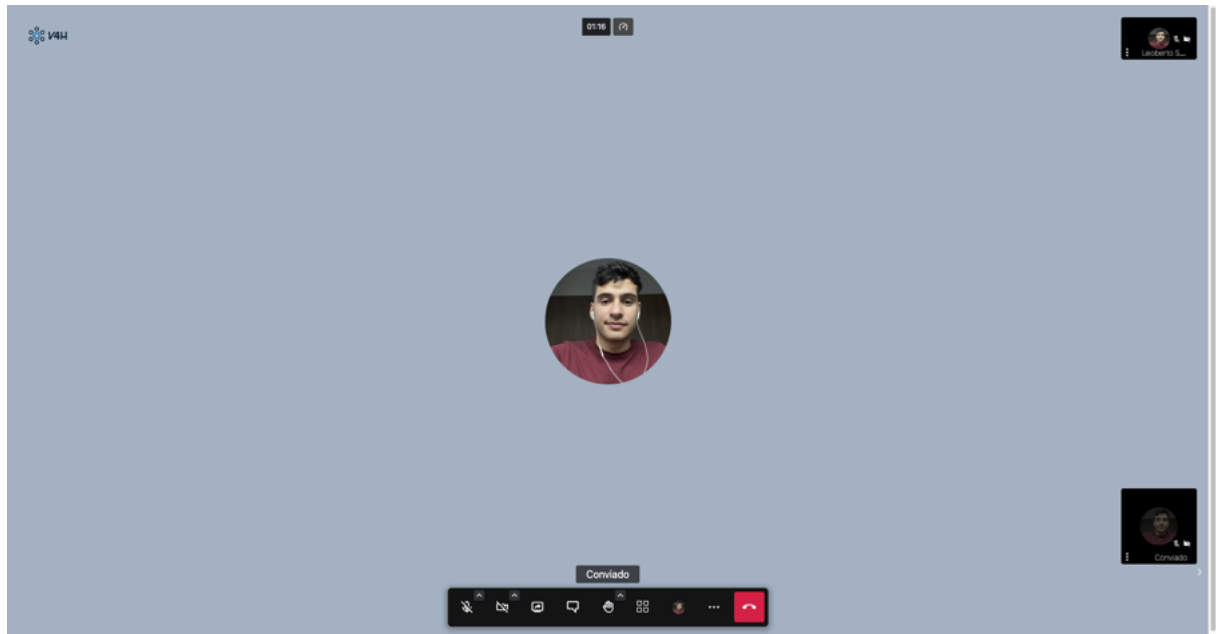


Figura 4: Interior da videoconferência

O sistema oferece aos usuários uma API com as funcionalidades de autenticação de participantes, preservação e recuperação dos vídeos com suporte a prova de existência, integridade e autenticidade utilizando-se de certificados digitais e *blockchain* e a associação da mídia gravada ao prontuário eletrônico do paciente de sistemas de terceiros.

O intuito da plataforma é que os teleconsultores e profissionais de saúde possam utilizar recursos síncronos de vídeo para atividades de supervisão, teleconsultoria, teleconsulta, telediagnóstico e preceptoria dos residentes com foco na Atenção Básica e Primária em Saúde em localidades remotas de forma segura e confiável (V4H, 2019).

O V4H disponibiliza um sistema de agendamentos, apresentado na figura 5, que possibilita a marcação de reuniões remotas de forma simples, desenvolvidos para instituições que não têm seu próprio *software* de controle de reuniões, o qual foi utilizada para a marcação dos experimentos realizados neste trabalho.

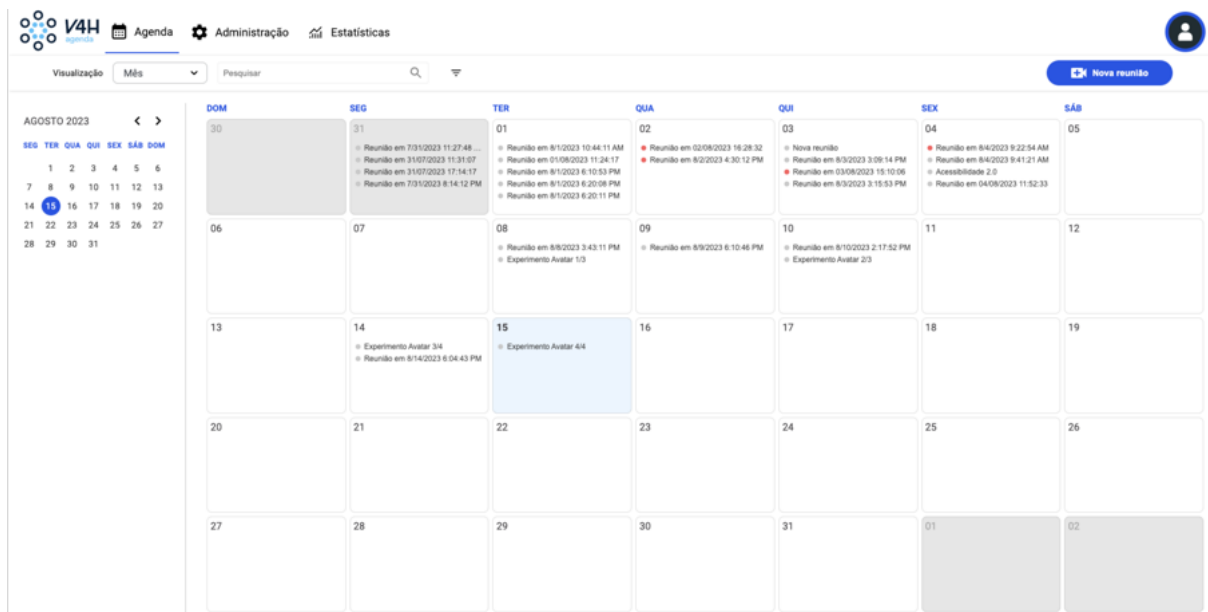


Figura 5: Sistema V4H Agenda

Após a criação de reuniões na plataforma *V4H Agenda*, o usuário recebe um *link* encurtado que poderá encaminhar para os participantes da conferência. Qualquer pessoal com acesso ao *link* conseguirá participar da chama de vídeo sem precisar realizar *login* ou criar uma conta.

As figuras [6](#) e [7](#) apresentam a tela na qual o usuário que acessa o *link* visualiza, nesta página é possível observar que o usuário deverá adicionar o seu nome no campo de texto e poderá escolher se entrará na conferência com microfone e câmera habilitados ou não.

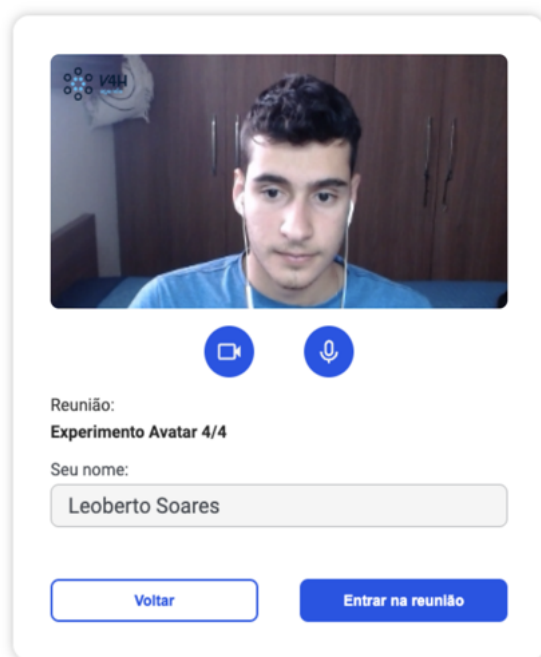


Figura 6: Pré-conferência: Câmera e microfone habilitados

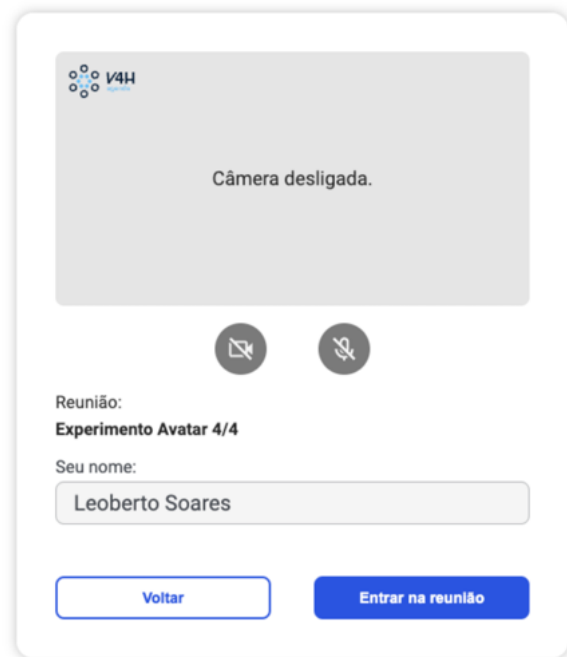


Figura 7: Pré-conferência: Câmera e microfone desabilitados

Ao clicar em **Entrar na reunião**, o usuário é redirecionado para o interior da videoconferência, como mostra a imagem 4, na qual ele pode interagir com a plataforma e com os demais usuários presentes na reunião.

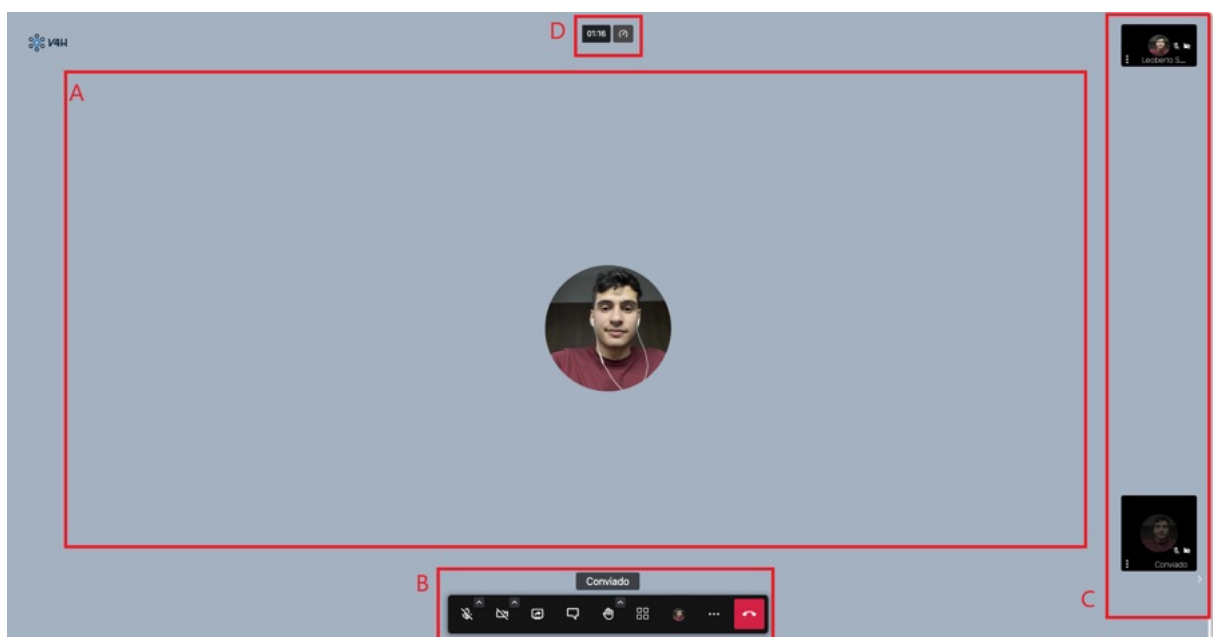


Figura 8: Áreas lógicas

O interior da videoconferência do V4H, como apresentado na figura 8, consiste em

4 áreas lógicas de primeiro nível onde estão localizadas os botões de funcionalidades e exibições de dados, vídeos, participantes, etc.

- **A:** Esta área é responsável por exibir o conteúdo da videoconferência, é onde o vídeo dos demais participantes ou do compartilhamento de tela são exibidos.
- **B:** Na área **B**, fica a barra de tarefas principal da plataforma, é nela que o usuário consegue configurar câmera e microfone, ativar *chat*, sinalizar que gostaria de falar, entre outras coisas.
- **C:** A área **C** é onde está localizado os demais participantes que não estão falando no momento, nesse espaço também é possível configurar informações de cada participante individualmente.
- **D:** A área **D** é onde está localizado o cronômetro com o tempo decorrido da videoconferência e acesso a lista de participantes que estão ou estiveram na videoconferência.

As áreas lógicas sofrem algumas mudanças quando o tipo de visualização é alterado, existem dois tipos de distribuição de informação na tela, a primeira é o formato tradicional apresentada na figura 4, a qual é a padrão do sistema.

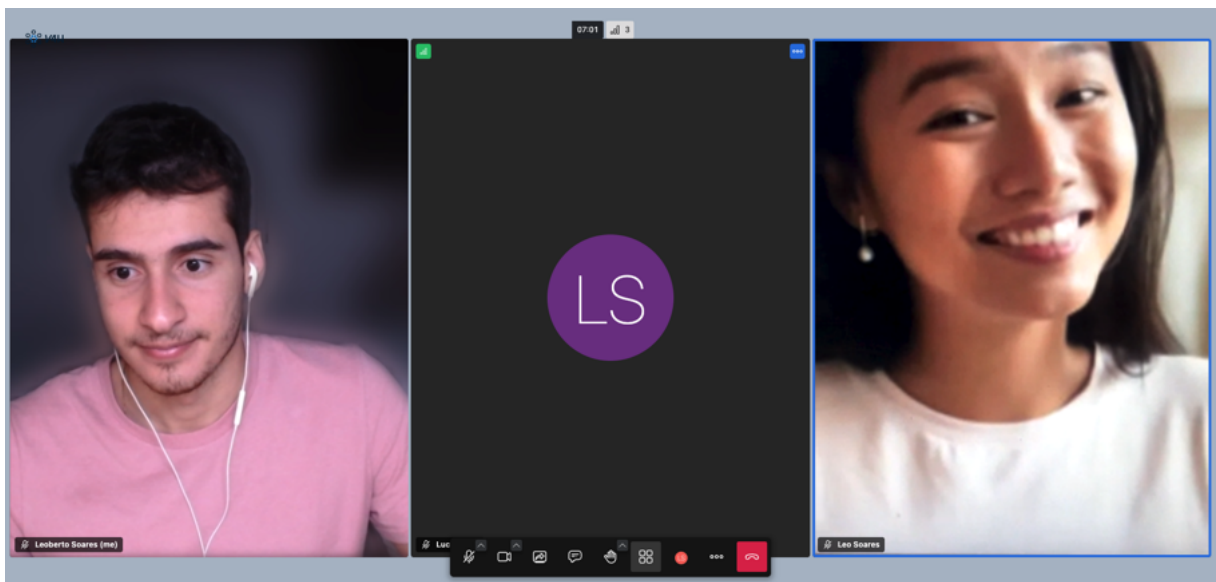


Figura 9: Visualização na disposição de blocos com três participantes

De forma opcional, o usuário pode, através da barra principal de tarefas, alterar o modo de visualização para o modo de blocos, apresentado na figura 9. Nesta disposição, os elementos que ficavam na **Área C** são redistribuídos na **Área A**, centralizando todos os vídeos em blocos menores respectivos a quantidade de pessoas presentes na reunião.

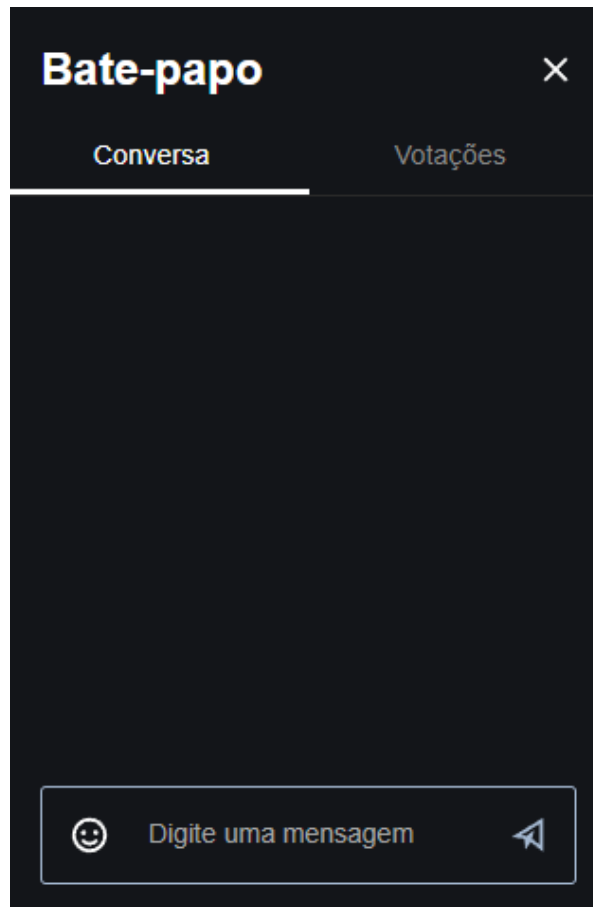


Figura 10: Modal do chat

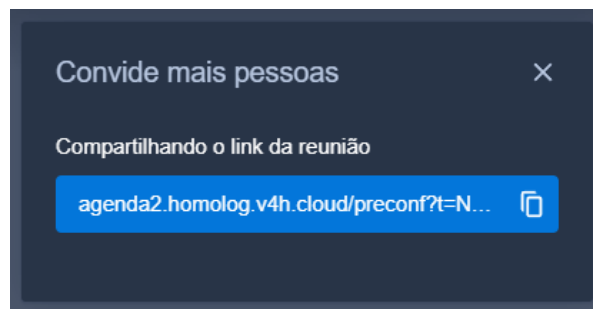


Figura 11: Modal de convite

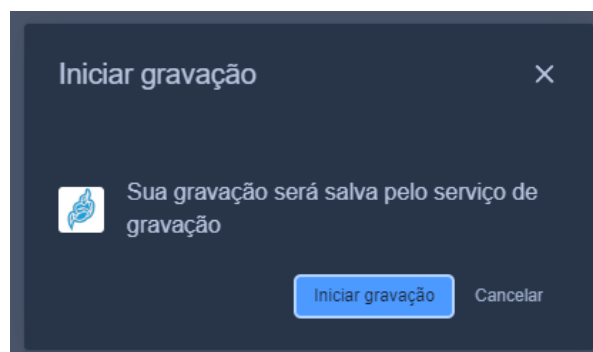


Figura 12: Modal de gravação

Vale salientar que existem outras áreas lógicas que são dinâmicas, como é o caso dos modais para iniciar/encerrar gravação, convidar participantes e chats, como é apresentado na figuras [10](#), [11](#) e [12](#).

Assim como os outros sistemas de videoconferência, o Video For Health possui um conjunto de funcionalidades visuais, na tabela [3](#) são apresentadas as principais funcionalidades e onde estão localizadas na interface.

Tabela 3: Funcionalidades na interface de videoconferência

Funcionalidade	Área Lógica	Papél
Visualização dos vídeo dos participantes	A	Todos
Tempo de duração da conferência	D	Todos
Ativar/desativar microfone	B	Todos
Ativar/desativar camera	B	Todos
Ativar/desativar compartilhamento de tela	B	Todos
Chat	A e B	Todos
Levantar mão	B	Todos
Visualização em blocos	A e B	Todos
Configuração de conta	B e Modal	Todos
Convidar pessoas	B	Moderador
Iniciar/Finalizar gravação	B	Moderador
Silenciar todos	B	Moderador
Desativar a câmera de todos	B	Moderador
Encerrar conferência	B	Todos
Visualização de participantes	C	Todos
Silência participante para todos	C	Moderador
Silência participante para si	C	Todos

2.6 WebSocket

WebSocket é uma tecnologia que torna possível abrir uma sessão de comunicação interativa entre o navegador do usuário e um servidor. A troca de mensagens entre as partes envolvidas são bidirecionais e síncronas, através de canais *full-duplex* sobre um soquete TCP.

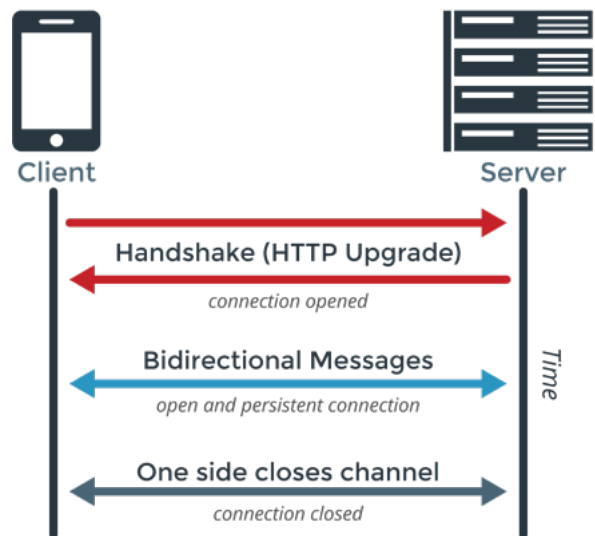


Figura 13: Fluxo de troca de mensagens via o *WebSocket*

Existem bibliotecas para servidores que disponibilizam interface que adicionam funcionalidades e facilitam a utilização de *websockets*. Como por exemplo, a biblioteca *socket.io*⁴ que disponibiliza métodos adicionais que ajudam no controle de usuários conectados.

Esta biblioteca, disponível em diversas linguagens, oferece diversas funcionalidades, a principal para o contexto do trabalho trata do conceito de salas, ou seja, um usuário conectado é adicionado a uma ou mais salas e o envio de mensagens para múltiplos participantes fica centralizado e facilitado.

2.7 OpenAI ChatGPT

O *ChatGPT* é um sistema de processamento de linguagem natural (NLP), como mostrado na figura 14, desenvolvida pela *OpenAI*. É projetada para receber entradas textuais, entender contexto e gerar respostas apropriadas para os *inputs* dos usuários (Deng und Lin, 2023).

É possível a utilização do *ChatGPT* através *APIs* e bibliotecas, sendo assim, uma solução poderosa para obter respostas e/ou gerar dados. São diversos modelos distintos que podem ser utilizados de forma gratuita ou pagos dependendo da quantidade de dados e inputs gerados. Na figura 15 é possível ter acesso a exemplos de integração via API.

⁴<https://socket.io/>

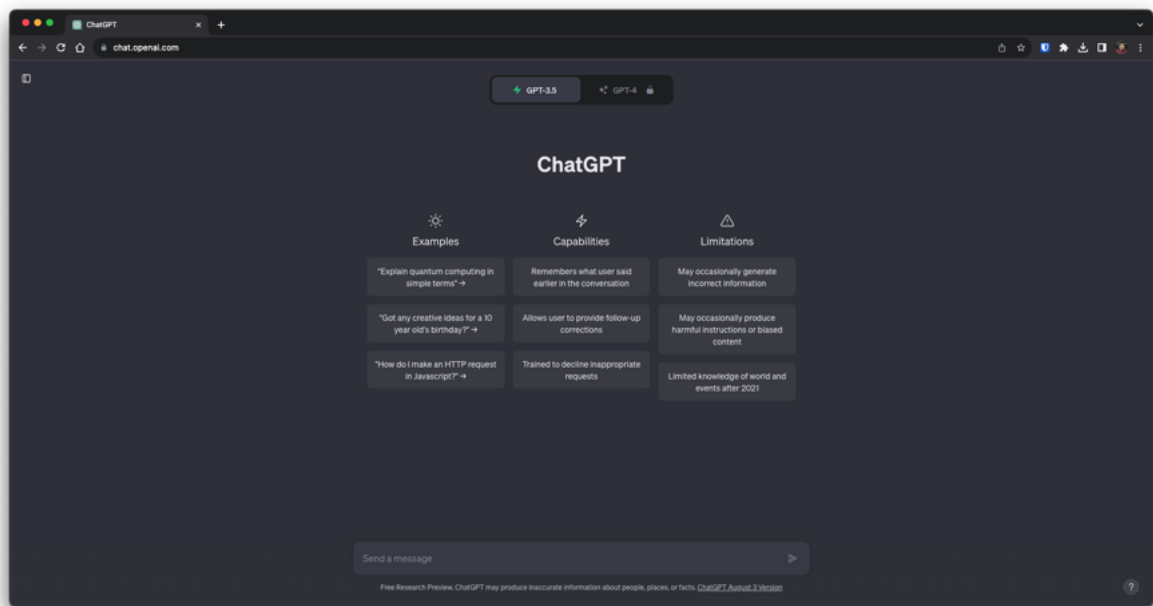


Figura 14: Tela inicial para inserção de dados no *ChatGPT*
Fonte: Imagem própria

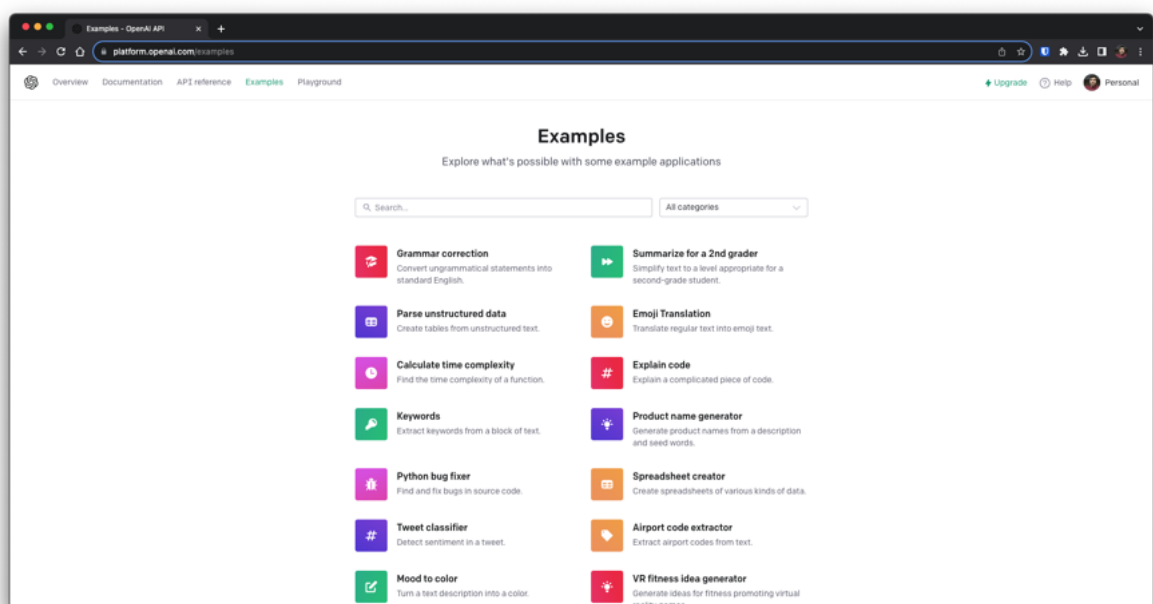


Figura 15: Exemplos de integração com o ChatGPT através de APIs
Fonte: Imagem própria

3 REVISÃO BIBLIOGRÁFICA

Nesta seção serão descritos os principais trabalhos relacionados ao tema abordado por esta dissertação. Para isso, foi utilizado as principais bases digitais de conteúdos científicos, e realizada uma busca relacionados à acessibilidade, acessibilidade em videoconferência, Libras, tradução automática de Libras e avatar de Libras.

3.1 Protocolo de Busca

Na revisão bibliográfica deste trabalho foi utilizado o método de revisão quasi-sistemática utilizando três bases de pesquisa: *IEEE Xplore*⁵, *ACM Digital Library*⁶ e *Springer*⁷.

No intuito de encontrar artigos relacionados, foi utilizado duas *strings* para a busca avançada, visando englobar a temática de Avatar de Libras, videoconferências e conversão de fala para texto. Dessa forma as seguintes *strings* foram utilizadas:

1. ("Sign Language" OR "Libras" OR "ASL") AND ("Video Conferencing" OR "avatar" OR "conference" OR "meeting")
2. ("Sign Language" OR "Libras" OR "ASL") AND ("Video Conferencing" OR "avatar" OR "conference" OR "meeting") AND ("text to Sign Language" OR "text to libras" OR "text to ASL")

Com a utilização dessas *strings* de busca foram retornados 118 itens na base *IEEE Xplore*, 16 na *ACM Digital Library* e 23 artigos no *Springer*, totalizando 157 artigos. Foi analisado o resumo de cada item, descartando os quais não se enquadravam na área da pesquisa. Após a primeira análise, restaram 30 artigos e uma dissertação de mestrado, as quais foram lidas em sua totalidade. Os trabalhos mais relevantes serão discutidos na seção seguinte, com suas respectivas descrições, metodologia e resultados, a fim de utiliza-los como referência para o presente trabalho.

3.2 Revisão Qualitativa dos Principais Trabalhos

3.2.1 Pesquisa sobre acessibilidade das ferramentas de videoconferência em plataforma web

O trabalho desenvolvido por (Ferraz u. a., 2020) tinha como objetivo avaliar as plataformas mais usadas de videoconferências, *Google Meet*, *Zoom*, *Microsoft Teams*, *Jitsi*,

⁵<https://ieeexplore.ieee.org/>

⁶<https://dl.acm.org/>

⁷<https://link.springer.com/>

WebEx e *BigBlueButton*. Esta avaliação pretendia abordar dois aspectos: apropriação de tecnologia pelos usuários e problemas na interação e na interface.

Apropriação de tecnologia pelos usuários: busca avaliar a produtividade do uso da tecnologia; identificar as necessidades e oportunidades de melhorias; compreender os efeitos da introdução dos sistemas na rotina dos usuários; entender as motivações dos usuários a negação na utilização de um sistema no cotidiano.

Problemas na interação e na interface: busca entender os problemas que prejudicam a qualidade da experiência do usuário, através da análise da interação e dos aspectos de interface que são apresentados para o usuário, com esse fim, as perguntas a seguir são levantadas:

- O usuário consegue operar o sistema?
- Ele atingiu o objetivo? Com qual eficiência? Em quanto tempo? Após cometer quantos erros?
- Que parte da interface e da interação o deixa insatisfeito?
- Ele entende o que significa e para que serve cada elemento da interface?
- Que problemas de interação dificultam ou impedem o usuário de alcançar seus objetivos?
- Em que situação esses problemas se manifestam?
- Com que frequência tendem a ocorrer?
- Qual é a gravidade desses problemas?
- Ela tem acesso a todas as informações oferecidas pelo sistema?

Com esse intuito, o trabalho utilizou múltiplos métodos de investigação tais como questionário, entrevistas e estudos de campo, com participação de usuários em contexto real de uso e por meios de métodos de inspeção com especialistas.

O trabalho dividiu os participantes convidados nos seguintes grupos:

1. **Operação somente por teclado:** para simular pessoas com deficiência motora, fazendo com que toda navegação seja feita através apenas do teclado;
2. **Usuário de tecnologia assistiva de leitor de tela, sem monitor:** para simular pessoas com deficiência visual, onde a navegação teria o *software* leitor ativado e teclado, sem a utilização visual do sistema;

3. **Operação sem áudio:** para simular pessoas com deficiência auditivas, desabilitando o áudio do computador.

Os participantes foram sorteados para utilizar uma solução de videoconferência analisada no trabalho de (Ferraz u. a., 2020), os testes foram realizados sem restrição de localização, ocorrendo no ambiente de trabalho ou na sua própria residência.

Os usuários foram encaminhados para uma videoconferência junto a um organizador que auxiliaram na realização nas tarefas, as quais são aquelas mais comuns no contexto de vídeo síncrono de diversos contextos, consistindo nas seguintes atividades:

- Entrar na reunião
- Habilitar/desabilitar microfone
- Habilitar/desabilitar câmera
- Enviar mensagem por chat
- Compartilhar tela
- Sair da reunião

Participaram 32 pessoas como avaliadores da pesquisa, resultando em 64 avaliações ao todo. No que diz respeito ao tipo de deficiência dos participantes, foi declarado que:

- 42,19% não têm deficiência
- 34,38% com deficiência visual total (cegueira)
- 3,13% com deficiência visual parcial (baixa visão)
- 15,63% com deficiência auditiva total (surdez)
- 4,69% com deficiência auditiva parcial (baixa audição)

Acerca ao modo de navegação, ou seja, navegação somente por teclado, navegação sem monitor e navegação sem áudio, as 64 avaliações foram distribuídas da seguinte forma:

- 46,88% avaliaram com o perfil de navegação somente por teclado
- 37,50% avaliaram com o perfil de navegação por leitor de tela
- 15,63% avaliaram com o perfil sem áudio

Ao fim do experimento, ou seja, quando eles entraram na videoconferência e realizaram as atividades dentro da plataforma previamente sorteadas, um questionário foi respondido. O autor do trabalho dividiu a análise dos resultados em 4 tópicos: acesso a ferramenta; interagindo na plataforma; compreensão da apresentação; saindo da reunião.

Acesso a ferramenta: foi questionado para os avaliadores se eles encontraram problemas para acessar a ferramenta, os quais todos que afirmaram ter tido algum tipo de dificuldade, foi no momento de autenticação na plataforma.

Percentual de respondentes que marcaram “sim” em suas respostas	TOTAL (64)	Perfil de navegação por leitor de tela (24)	Perfil de navegação sem áudio (10)	Perfil de navegação somente por teclado (30)
Você conseguiu fazer o login na ferramenta?	84%	71%	60%	83%
Ao entrar na reunião, sua câmera já estava habilitada?	47%	17%	90%	53%
Você consegue perceber se seu microfone e/ou câmera estão habilitados?	88%	71%	70%	93%
Ao entrar na reunião, seu microfone já estava habilitado?	64%	46%	70%	67%
Você consegue habilitar e desabilitar o microfone?	91%	71%	80%	97%
Você consegue habilitar e desabilitar a câmera?	86%	71%	80%	87%

Figura 16: Resultados do questionário para o tópico de *Acesso a Ferramenta*
 Fonte: (Ferraz u. a., 2020)

Em relação a câmera e microfone apresentado na figura 16, 47% dos avaliadores afirmaram que suas câmeras já estavam habilitadas ao entrar na reunião e 86% conseguiam habilita-la ou desabilita-la durante a conferência, de forma semelhante, 64% dos microfones já estavam com o microfone ativo no início da videoconferência e 91% conseguiram habilita-lo ou desabilita-lo a qualquer momento.

De modo geral é possível verificar que as pessoas que utilizam apenas teclados tiveram menos dificuldades para acessar as plataformas, em contra ponto, os avaliadores que navegavam com leitor de tela foram os que mais apresentaram mais barreira no uso dos sistemas.

Interagindo na plataforma: neste tópico foi questionado aos usuários sobre as interações com as plataformas, portanto, segundo a figura 17, no total 94% responderam que conseguiram interagir com áudio ou vídeo com os demais participantes, porém quando o recorte é feito para o grupo de deficiência visual, 71% não conseguem interagir com os demais participantes.

Percentual de respondentes que marcaram “sim” em suas respostas	TOTAL	Perfil de navegação por leitor de tela	Perfil de navegação sem áudio	Perfil de navegação somente por teclado
Você consegue interagir por áudio e/ou vídeo com o apresentador e os demais participantes?	94%	71%	90%	100%
Você consegue enviar mensagens por chat?	97%	75%	90%	100%
Você consegue sinalizar de alguma forma que quer se manifestar durante a reunião (por exemplo, ações como “levantar a mão” durante a reunião)?	61%	38%	60%	77%
Você consegue acessar a lista de participantes da reunião?	83%	67%	90%	77%
Você consegue acessar os links enviados pelo chat?	67%	50%	90%	60%
Você consegue compartilhar sua tela?	86%	67%	80%	90%

Figura 17: Resultados do questionário para o tópico de *Interagindo na Plataforma*

Fonte: (Ferraz u. a., 2020)

Vale salientar que em relação ao *chat*, 97% conseguiram enviar mensagem por *chat*,

porém apenas 67% dos avaliadores conseguiram acessar um *link* que tinha sido enviado durante a videoconferência.

Uma atividade comum nas reuniões é expressar que deseja se manifestar ou falar algo, porém apenas 61% dos avaliadores conseguiram sinalizar, no retrato da deficiência visual, apenas 38% obtiveram êxito nesta atividade.

Compreensão da apresentação: na seção de compreensão, como mostra a figura 18 apenas 19% dos participantes conseguiram habilitar as legendas, independente do idioma que a plataforma disponibilizava.

Após as exibições dos vídeos e imagens durante a apresentação, foi questionado para os avaliadores se conseguiram compreender tudo o que foi exibido na tela, no total apenas 38% dos participantes informaram que tinha conseguido entender o que tinha sido descrito, no recorte dos usuários que navegaram por leitor de tela e que navegaram sem áudio, o resultado foi respectivamente 4% e 60%.

Percentual de respondentes que marcaram “sim” em suas respostas	TOTAL	Perfil de navegação por leitor de tela	Perfil de navegação o sem áudio	Perfil de navegação somente por teclado
Você consegue habilitar/desabilitar legendas?	19%	13%	40%	17%
Você consegue compreender tudo o que está sendo exibido na tela pelo apresentador?	38%	4%	60%	57%

Figura 18: Resultados do questionário para o tópico de *Compreensão da Apresentação*

Fonte: (Ferraz u. a., 2020)

Saindo da reunião: Por fim, foi questionado se os participantes conseguiam sair da reunião por conta própria, sem a interferência do apresentador. O resultado obtido foi que 97% afirmaram que conseguiam sair, como apresentado na figura 19.

Percentual de respondentes que marcaram “sim” em suas respostas	TOTAL	Perfil de navegação por leitor de tela	Perfil de navegação o sem áudio	Perfil de navegação somente por teclado
Você consegue sair da reunião por conta própria?	97%	75%	90%	100%

Figura 19: Resultados do questionário para o tópico de *Saindo da Reunião*
 Fonte: (Ferraz u. a., 2020)

Nas figuras 20 e 21 é possível observar as respostas em relação as ferramentas utilizadas, observa-se que todas as ferramentas tiveram resultados negativos em algumas das questões, mostrando que todas tem pontos de melhorias importantes.

Resultado total por ferramenta (resposta sim)	Zoom	Google Meet	Jitsi	Microsoft Teams	WebEx	BigBlue Button
Você conseguiu fazer o login na ferramenta?	100%	78%	75%	80%	83%	92%
Ao entrar na reunião, sua câmera já estava habilitada?	13%	44%	83%	40%	67%	23%
Você consegue perceber se seu microfone e/ou câmera estão habilitados?	63%	100%	83%	100%	92%	85%
Ao entrar na reunião, seu microfone já estava habilitado?	50%	67%	75%	80%	58%	54%
Você consegue habilitar e desabilitar seu microfone?	88%	100%	83%	100%	92%	85%

Figura 20: Resultados do questionário por ferramenta
 Fonte: (Ferraz u. a., 2020)

Resultado total por ferramenta (resposta sim)	Zoom	Google Meet	Jitsi	Microsoft Teams	WebEx	BigBlue Button
Você consegue habilitar e desabilitar sua câmera?	88%	89%	92%	100%	100%	54%
Você consegue interagir por áudio e/ou vídeo com o apresentador e os demais participantes?	100%	100%	92%	100%	100%	77%
Você consegue habilitar/desabilitar legendas?	0%	67%	8%	30%	8%	8%
Você consegue compreender tudo o que está sendo exibido na tela pelo apresentador?	38%	44%	42%	50%	50%	8%
Você consegue enviar mensagens por chat?	88%	100%	100%	100%	100%	92%
Você consegue sinalizar, de alguma forma, que quer se manifestar durante a reunião (por exemplo, ações como “levantar a mão” durante a reunião)?	50%	22%	92%	100%	67%	31%
Você consegue acessar a lista de participantes da reunião?	75%	100%	50%	90%	92%	92%
Você consegue acessar os links enviados pelo chat?	13%	78%	83%	80%	50%	85%
Você consegue compartilhar sua tela?	75%	89%	92%	100%	83%	77%
Você consegue sair da reunião por conta própria?	88%	100%	100%	100%	100%	92%

Figura 21: Resultados do questionário por ferramenta

Fonte: (Ferraz u. a., 2020)

3.2.2 Uma proposta de interface de videoconferência acessível para pessoas com deficiência auditiva

Na dissertação de (Araújo, 2021) é apresentada uma proposta de interface para a ferramenta V4H (*Video For Health*) de videoconferência, apresentado na figura 22, no intuito de gerar uma conferência acessível para pessoas com deficiência auditiva, para isso foi adicionado o suporte para intérprete humano de libras e geração de legendas automáticas.

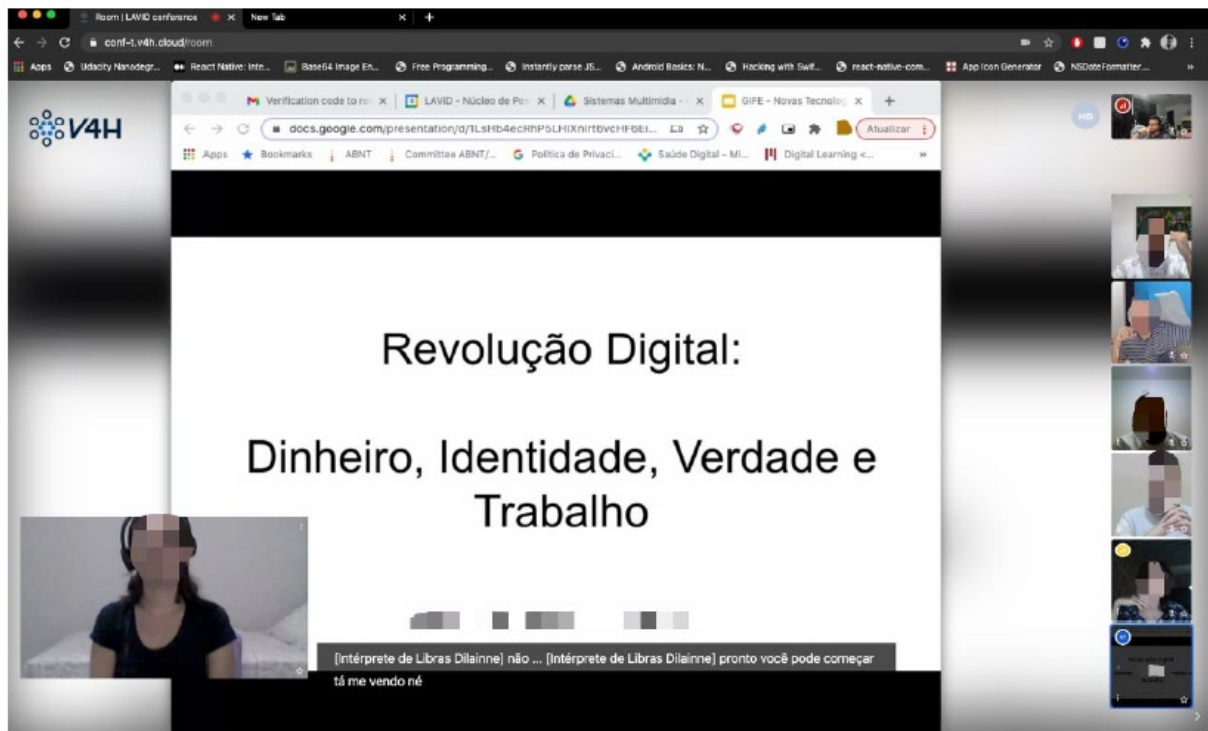


Figura 22: Proposta de inserção do intérprete humano na videoconferência (Araújo, 2021)

Neste trabalho, o autor realizou uma pesquisa no âmbito da Interação Homem-Máquina na qual buscou propor a inserção do interprete humano na tela da videoconferência. Como mostra a figura a 23, a interface tem a finalidade de entregar uma experiência satisfatória na escolha, na visualização do intérprete, legendas automáticas e no conteúdo que esteja sendo apresentado em uma videoconferência tradicional.

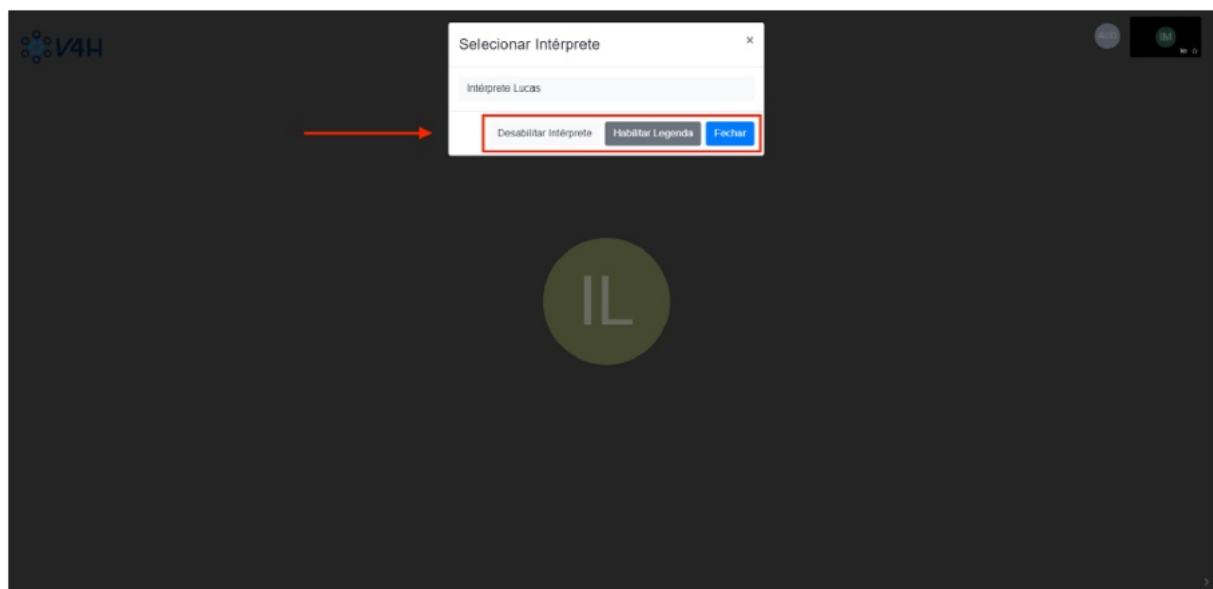


Figura 23: Proposta do menu de acessibilidade (Araújo, 2021)

A proposta do autor consiste em um menu de acessibilidade, no qual é possível habilitar ou desabilitar o intérprete, assim como ativar e desativar as legendas automáticas. Para abertura do menu apresentado, foi adicionado um botão de acessibilidade ao lado dos demais botões presentes na interface do V4H - Video For Health.

A arquitetura desenvolvida, exibida na figura 24, consiste na captura da fala no cliente através da biblioteca *Web Speech API*, que utiliza do microfone do usuário para transcrever o que é falado por ele.

Essa transcrição é encaminhada para os demais participantes através de um servidor *WebSocket* e armazenada em uma banco de dados. No lado do cliente, o texto recebido é adicionado como legenda, como apresentado na figura 22.

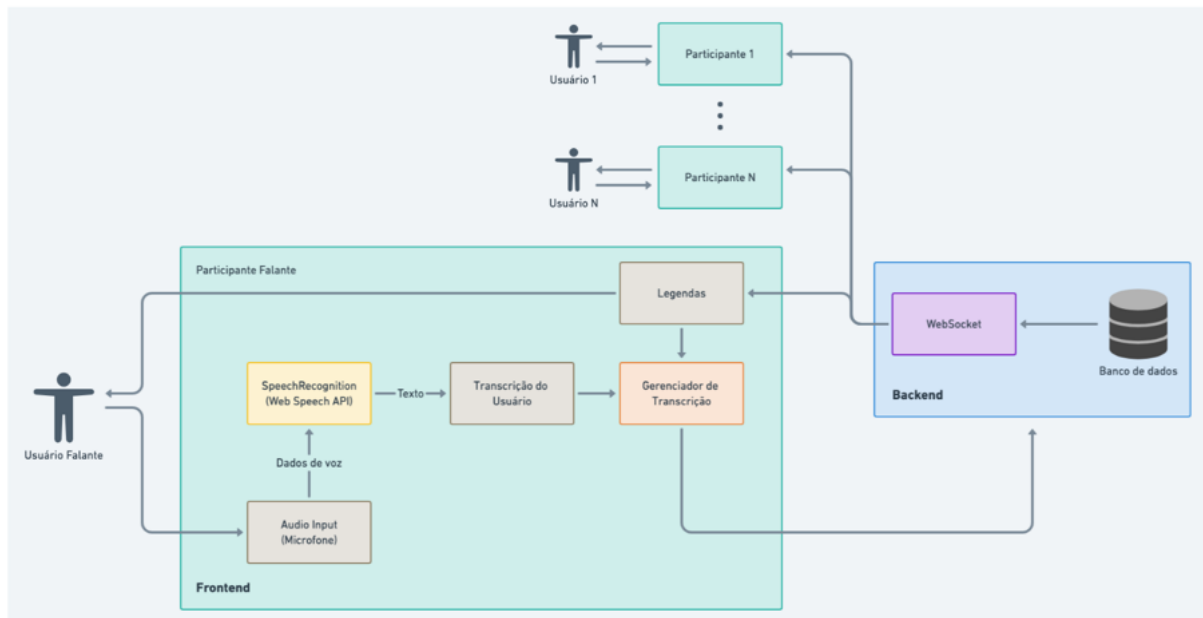


Figura 24: Arquitetura de captura da fala e troca de mensagens (Araújo, 2021)

O experimento por ele realizado consistiu na realização de uma palestra de tema livre com duração média de 15 minutos, onde cada participante, de forma sortida, recebe um conjunto de tarefas pré definidas para serem realizadas pelos participantes durante a conferência.

A fim de obter informações sobre a qualidade da interface desenvolvida, foram utilizadas tarefas, questionários e metodologia do trabalho de (Ferraz u. a., 2020). O experimento conduzido pela dissertação descreve um teste de usabilidade do formato remoto, sem qualquer tipo de explicação inicial, e com as tarefas aplicadas para acessibilidade.

Pergunta	Outras Ferramentas (SIP)	Solução Proposta (CIP)
1. Você conseguiu fazer o login na ferramenta?	60%	90%
3. Ao entrar na reunião sua câmera já estava habilitada?	90%	80%
4. Você consegue perceber se seu microfone e/ou câmera estão habilitados?	70%	80%
5. Ao entrar na reunião seu microfone já estava habilitado?	70%	90%
6. Você consegue habilitar e desabilitar o microfone?	80%	80%
7. Você consegue habilitar e desabilitar sua câmera?	80%	90%
8. Você consegue interagir por áudio e/ou vídeo com o apresentador e os demais participantes?	90%	90%
9. Você consegue habilitar/desabilitar legendas?	40%	90%
10. Você consegue compreender tudo o que está sendo exibido na tela pelo apresentador?	60%	90%
11. Você consegue enviar mensagens por chat?	90%	90%
12. Você consegue sinalizar, de alguma forma, que quer se manifestar durante a reunião (por exemplo, ações como “levantar a mão” durante a reunião)?	60%	90%
14. Você consegue acessar a lista de participantes da reunião?	90%	80%
15. Você consegue acessar os links enviados pelo chat?	90%	100%
16. Você consegue compartilhar sua tela?	80%	90%
18. Você consegue sair da reunião por conta própria?	90%	90%

Figura 25: Tabela comparativa entre a solução proposta e demais ferramentas de videoconferência (Araújo, 2021)

Como apresentado na figura 25, é possível visualizar de modo geral um melhor desempenho com a utilização da solução proposta pelo autor, quando observadas de maneira isolada as tarefas que envolvem acessibilidade, essa disparidade positiva aumenta consideravelmente.

3.2.3 Avaliação de Tecnologias de Tradução Português-Libras Visando o Uso no Ensino de Crianças Surdas

No artigo de (Colling und Boscarioli), é realizado um levantamento das tecnologias e ferramentas que colaboram no âmbito educacional do ensino de crianças surdas. Essas ferramentas utilizam Avatar de Libras ou intérpretes humanos em conjunto com a metodologia de ensino tradicional para contribuir com o aprendizado dessas crianças.

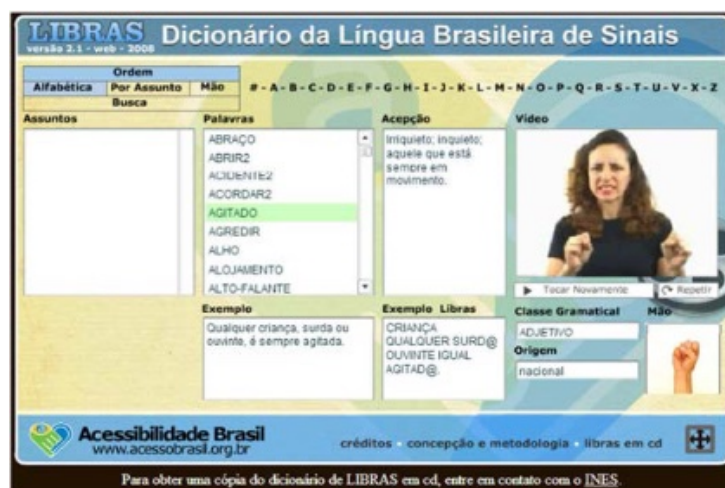


Figura 26: Tela principal do Dicionário da Língua Brasileira de Sinais (Colling und Boscarioli)



Figura 27: Tela principal do Jogo Multi-Trilhas (Colling und Boscarioli)

Como tecnologia no apoio pedagógico, o Dicionário Da Língua Brasileira de Sinais apresenta um conjunto vasto de vocabulário em português, com traduções pré-gravados em Libras, exemplo de aplicações em frases, classe gramatical, entre outras informações importantes para o aprendizado da linguagem de sinais, como apresentado na figura 26. Outra abordagem para apoio pedagógico, exemplificada na figura 27, consiste no jogo que visa a aquisição da segunda língua através da utilização prática de classes gramaticais, com apoio da linguagem escrita em português.

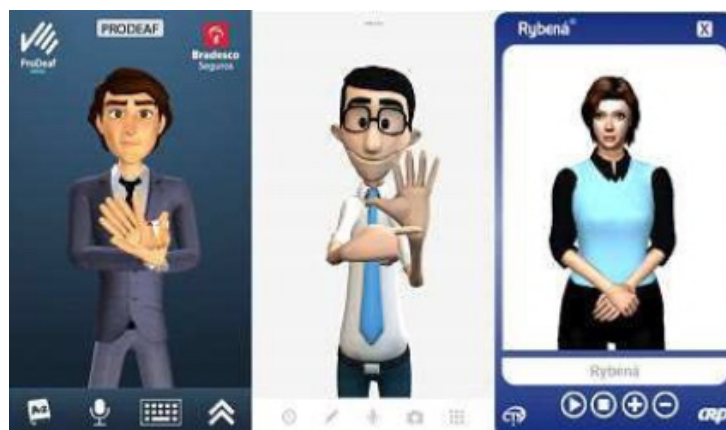


Figura 28: Tradutores Automáticos Português-Libras (Colling und Boscarioli)

No intuito de facilitar a criação de softwares de apoio, os avatares de Libras têm papel essencial nesse contexto, pois eliminam a necessidade do intérprete humano, muitas vezes indisponível por escassez no mercado. Visando isso, o autor fez uma comparação entre três tradutores automáticos de Português-Libras: **Rybená**, **Handtalk** e **ProfDeaf**. A finalidade da comparação entre as ferramentas visa validar a qualidade das traduções e a acurácia no entendimento pelas crianças.

Os experimentos consistem na inserção de frases em português para que os avatares individualmente fizessem as traduções para Libras, verificando se a tradução estava adequada seguindo parâmetros pré-estabelecidos. Por outro lado se busca investigar se as crianças conseguiram captar corretamente o conteúdo das frases.

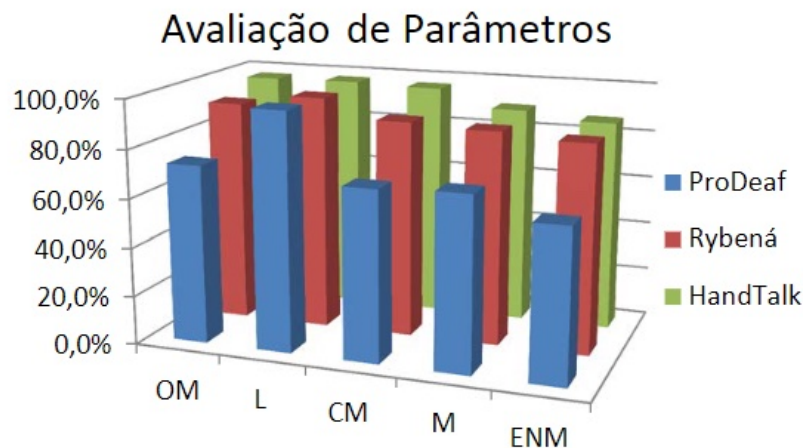


Figura 29: Resultado da Corretude dos Parâmetros da Libra Para os Tradutores Automáticos de Libras (Colling und Boscarioli)

O gráfico da figura 29 reflete a tradução incorreta em múltiplos casos, caracterizados principalmente pela regionalização em sinalizações e a falta de expressão nos avatares como principal falha observada.



Figura 30: Proposta de Processo Para Tradução Português-Libras (Colling und Boscarioli)

Por fim, os autores apresentam uma proposta de processo de tradução apresentada na figura 30, visando a melhoria dos resultados obtidos pelas traduções. Para isso, os seguintes pontos principais devem ser modificados: eliminação de artigos, preposições e conectivos; expressões contextualizadas; público alvo da tradução; entre outros pontos.

3.2.4 ES2ISL: An advancement in speech to sign language translation using 3D avatar animator

O artigo (Patel u. a., 2020) apresenta um sistema chamado **ES2ISL**, que consiste em uma plataforma *Web* de tradução da fala em inglês para língua de sinais indiana (ISL). O sistema proposto, busca melhorar a acurácia, tempo de resposta, utilização de memória e reconhecimento de fala em comparação a um sistema já existente, para isso, foi desenvolvido uma proposta de plataforma com três estágios, apresentado na figura 31

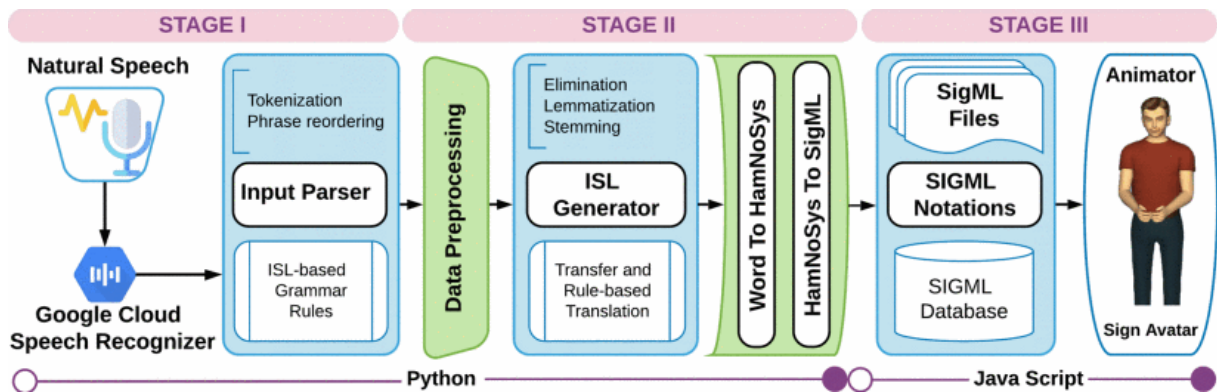


Figura 31: Estágios do conversor de fala para língua de sinais (Patel u. a., 2020)

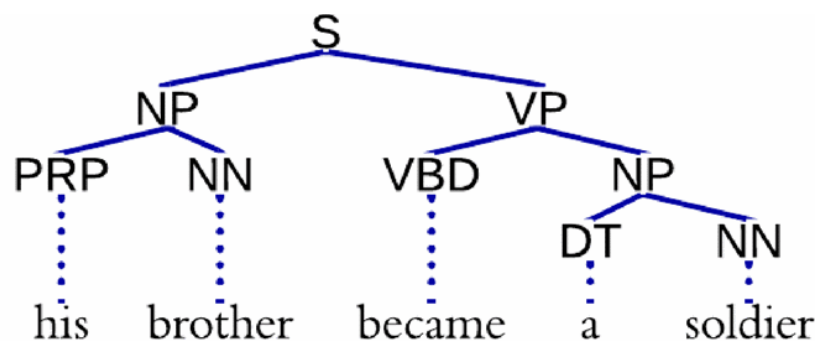
Estágio 1 - Conversão de fala para texto: Este módulo utiliza o microfone do usuário para capturar a fala, com essa entrada de informação é utilizado o *Google Cloud Speech API* para converter o sinal áudio em texto. O texto transcrito é analisado e dividido em sentenças, e para cada frase, as palavras presentes são extraídas e separadas, como apresentado na figura 32.

Input (text): His brother became a soldier.

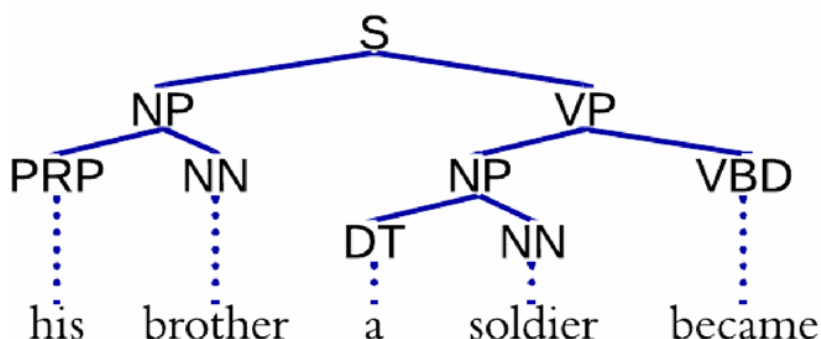
Output (tokens): 'His', 'brother', 'became', 'a', 'soldier', '.'

Figura 32: Exemplo de análise e quebra de cada frase (Patel u. a., 2020)

A quebra das sentenças em palavras é necessário para estruturar gramaticalmente essas mesmas frases seguindo as regras gramaticais *Hindi*. Essas alterações são notórias na figura 33, onde é possível observar através de um exemplo a organização das palavras na frase.



(a) English Language Phrase Tree



(b) Hindi (Indian) Language Phrase Tree

Figura 33: Transição do Inglês para o Hindi (Patel u. a., 2020)

Estágio 2 - Processamento dos dados: No segundo estágio, as frases convertidas passam por um processamento para se adequar corretamente a língua de sinais, transformando o texto em glosa. Esse processamento é dividido em três etapas ou regras: eliminação (*Eliminator*), radicalização (*Stemming*) e normalização (*Lemmatization*).

Essas regras buscam remover palavras de ligação, ou seja, são removidos verbos de ligação (*am, is, are, was, were*) e artigos (*a, an, some, the*), entre outras. Também é necessário a remoção dos sufixos e gerúndios, pois a língua de sinais (ISL) apenas utilizam a palavra na sua forma radicalizada, além de eliminar as palavras flexionadas para garantir que a palavra raiz faça parte do idioma.

Estágio 3 - Notação SIGML: Nesta etapa é necessário converter as palavras para sinais presentes no dicionário para que o avatar consiga interpreta-la. Caso a palavra não esteja presente no dicionário, a mesma deverá ser soletrada, ambos os casos são apresentados na figura 34.

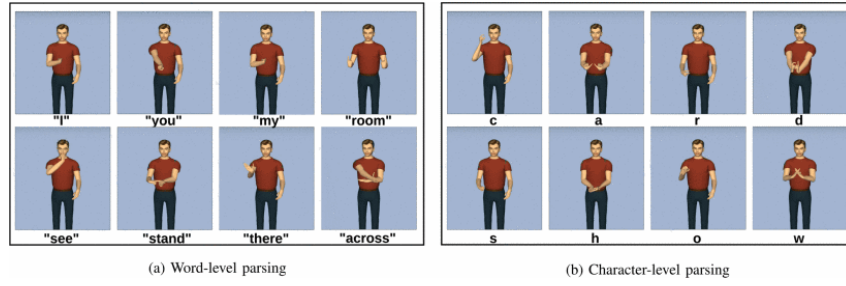


Figura 34: Dicionário de animações para *ISL* (Patel u. a., 2020)

O experimento utilizou 101 falas como entradas, todas as falas foram produzidas por pessoas em tempo real. As frases faladas contemplam textos curtos e grandes para avaliar a solução. A acurácia dos sistemas foram avaliadas seguindo três métricas: reconhecimento de voz (V_a), análise gramatical (P_a), e ação interpretada pelo avatar (A_a). Os resultados foram obtidos em porcentagem comparada com a quantidade total inseridas. A figura 35 mostra os resultados obtidos.

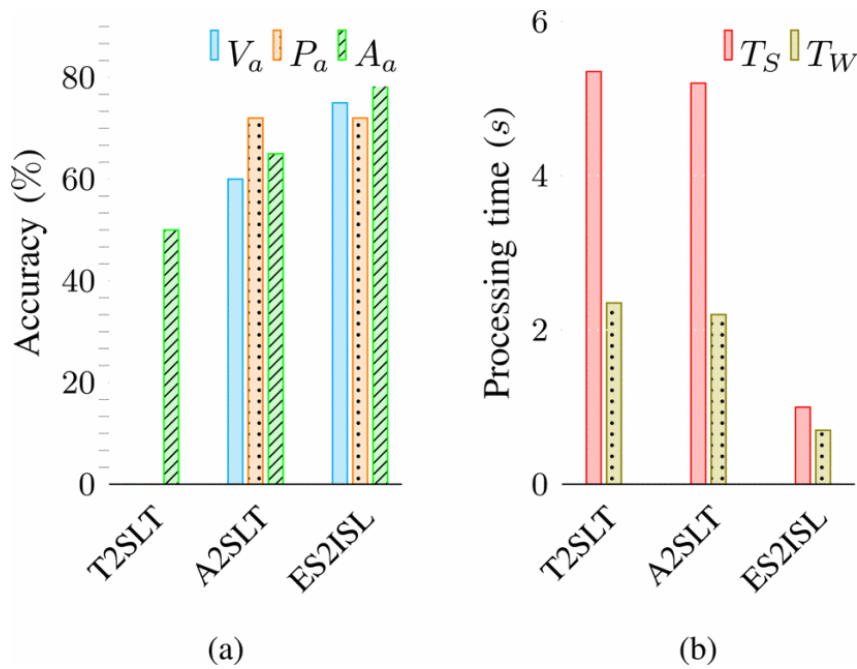


Figura 35: Resultados da acurácia e tempo de processamento em comparação com sistemas presentes na literatura (Patel u. a., 2020)

4 Descrição da Solução Proposta

Nesta seção será apresentada a solução proposta para esta dissertação, a qual consiste na adição do componente visual de acessibilidade para surdos, através do *player* de avatar de libras e legendas automáticas. Para este propósito, serão adicionados os componentes de reconhecimento de fala em tempo real e avatar de libras no servidor de videoconferência Video For Health (V4H), além da utilização de um servidor *websocket* próprio para as trocas de mensagens.

4.1 Arquitetura

A figura [36](#) contém o diagrama da arquitetura proposta neste trabalho, a qual pode ser dividida em quatro estágios: captura, processamento, transmissão e apresentação.

1. Captura: O áudio do usuário é captado através do microfone;
2. Processamento: O reconhecimento de fala é aplicado e o sinal de áudio é transformado em texto;
3. Transmissão: Os textos gerados são encaminhados para o servidor que envia para os demais participantes;
4. Apresentação: Os textos recebidos são enviados para o avatar que transformará em sinais de Libras.

Ao entrar na conferência os usuários são conectados a um servidor de mensagens via *WebSocket*, a fim de enviar e receber as futuras transcrições geradas pelos usuários presentes na reunião. A captura do áudio do participante acontece quando o mesmo habilita o seu microfone através do botão na interface, esse sinal é encaminhado para a biblioteca que fará o reconhecimento da fala e retornará o texto transcrito.

Após a transcrição ser realizada é averiguado o tamanho do texto, se o texto possuir mais que 20 palavras, esse texto é enviado para um sumariador, que consiste em uma *REST API* que recebe como entrada o texto original e devolve como resposta um texto sumariado.

O texto final é enviado para o servidor de mensagens que encaminhará para os demais participantes da videoconferência que o usuário faz parte. Nesta etapa, o servidor de mensagens envia as transcrições para um microserviço responsável por armazenar as transcrições geradas durante a reunião, denominado de **TSS** (*Transcription Storage Service*).

As mensagens recebidas pelos demais participantes são ignoradas caso o avatar de Libras e legenda estejam desabilitados. Com contrário, as mensagens são exibidas no espaço destinado para a legenda e empilhadas em uma fila para serem sinalizadas pelo avatar de Libras.

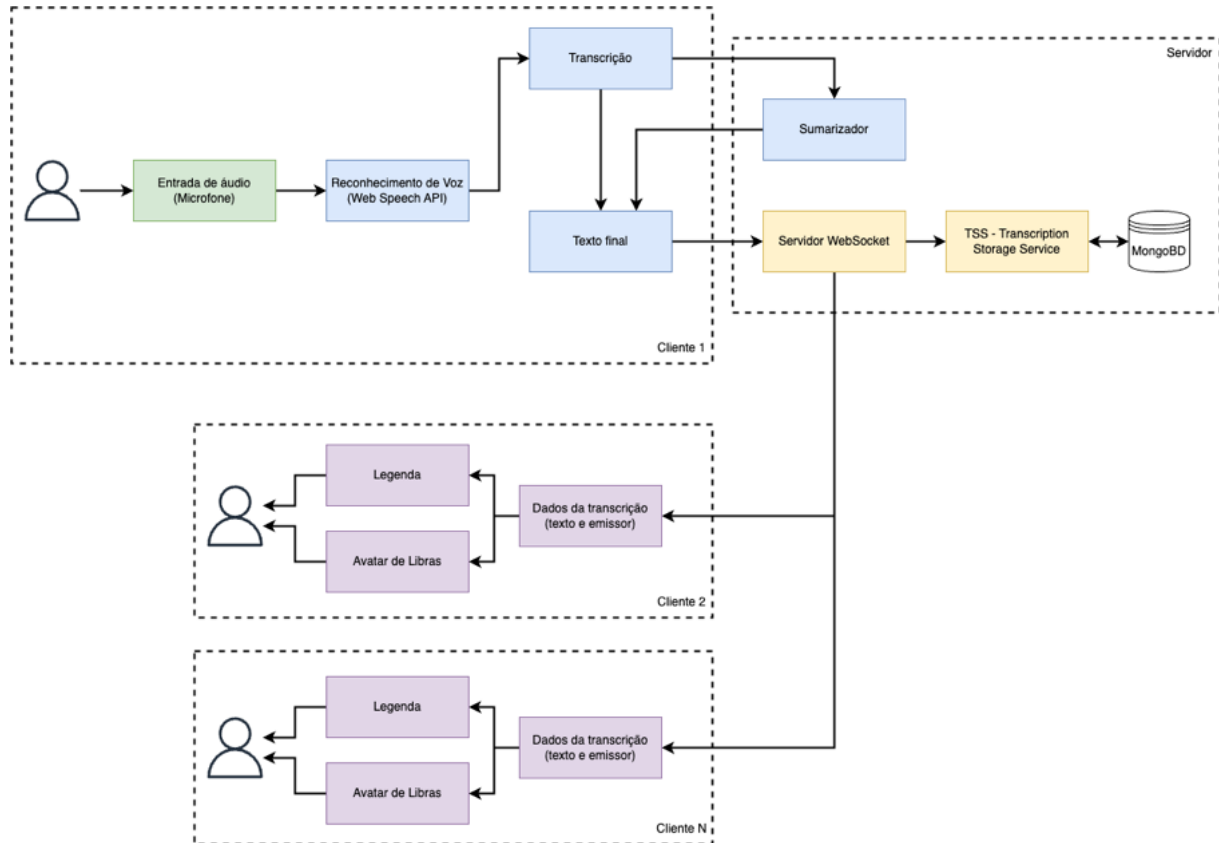


Figura 36: Diagrama do funcionamento das legendas e tradução para avatar em tempo real (Fonte: Imagem própria)

4.2 Proposta de Interface Gráfica

A proposta inicial para interface gráfica consiste na criação de dois componentes de acessibilidade: o avatar de Libras e a legenda. Ambos os itens devem ser adicionados na tela principal de conferência, figura 4, de forma condicional, acionados por um botão.

4.2.1 Componente do Avatar de Libras

Este componente é responsável pela exibição da tradução do português para Libras através do avatar da *WebAssista*. Como o avatar não tem por padrão a informação de quem está falando, foi adicionado o nome do usuário logo acima, para deixar claro para os usuários de quem corresponde aquela sinalização, como apresentado na figura 37.



Figura 37: Proposta do componente do Avatar de Libras (Fonte: Imagem própria)

Abaixo do avatar, foram adicionados dois botões auxiliares: **Zerar** e **Visualizar texto**. A opção de zerar tem a funcionalidade de limpar a fila de transcrições pendentes de sinalização, esta opção foi pensada para o caso do avatar fique para trás em relação ao que está sendo falado e apresentado pelos participantes. Esse atraso pode ser causado pela alta velocidade que os participantes se utilizam durante a videoconferência, e/ou caso a velocidade do avatar não seja compatível.

O segundo botão presente consiste na exibição do texto que o avatar estar sinalizando no momento, apresentado na figura [38](#). Esta opção é utilizada principalmente para depuração ou para avaliar a tradução e caso o avatar não esteja sinalizando nada, não será exibido nada ao ser clicado.

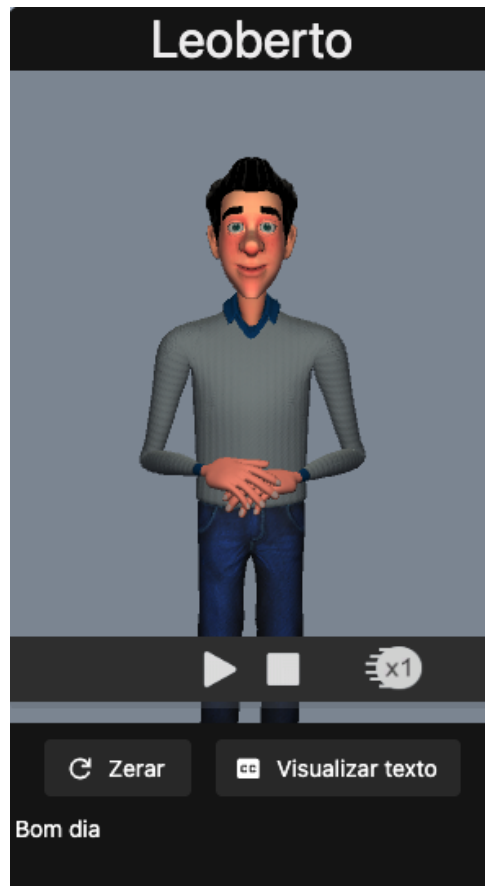


Figura 38: Proposta do componente do Avatar de Libras quando o botão de "Visualizar texto" estiver ativado (Fonte: Imagem própria)

Vale salientar que as áreas que estão os botões auxiliares e o nome do responsável pela transcrição estão com fundo opaco, para não confundir com o que está sendo apresentado na videoconferência. Já o avatar se encontra com fundo transparente, pois como é uma parte maior do componente de Libras poderia esconder dados importantes que fossem apresentado durante a conferência.

4.2.2 Componente de legenda

A figura 39 refere-se ao componente da legenda, o qual é responsável por exibir as transcrições de cada participante da conferência com o indicativo do usuário responsável pela fala. Esse componente consiste em uma *div* com rolagem ativa que comporta até 3 linhas de texto, sendo assim, é possível visualizar todo o histórico de legendas.



Figura 39: Proposta do componente de legenda (Fonte: Imagem própria)

Cada nova transcrição é inserida em uma nova linha a fim de separar momentos diferentes de falas e facilitar a visualização e identificação pelo usuário. Um pouco diferente do componente de Libras, o fundo escolhido tem um grau de transparência, fazendo com que os dados atrás não fiquem escondidos por completo, mas ao mesmo tempo possibilitando uma boa legibilidade do texto.

4.2.3 Barra de tarefas

Os componentes descritos nas seções anteriores serão ativados através de um botão de acessibilidade na barra de ferramentas na parte inferior central da tela, onde se encontram os demais botões. O botão de acessibilidade é do tipo *toggle*, ele assumirá uma cor de fundo diferente quando estiver ativado como apresentado nas figuras 40 e 41.



Figura 40: Proposta da barra de ações (Fonte: Imagem própria)



Figura 41: Proposta da barra de ações quando acessibilidade automática estiver ativada (Fonte: Imagem própria)

4.2.4 Proposta final

Na figura 42 é possível visualizar a proposta completa da interface com as inserções dos componentes, é válido salientar que os componentes de avatar e legenda serão móveis por toda a tela, a fim de não atrapalhar o compartilhamento de tela ou visualização da câmera do participante em destaque, deixando a responsabilidade para o usuário mover-se assim achar necessário.

As posições iniciais destinadas para o componente do avatar e o componente de legenda foram definidas baseadas no trabalho de (Araújo, 2021), diferindo apenas no botão de acessibilidade, o qual adicionou no canto inferior direito da tela do usuário. Essa mudança tem como justificativa atualizações recentes da plataforma de videoconferência do V4H, centralizando todos os botões em uma única barra de tarefas.

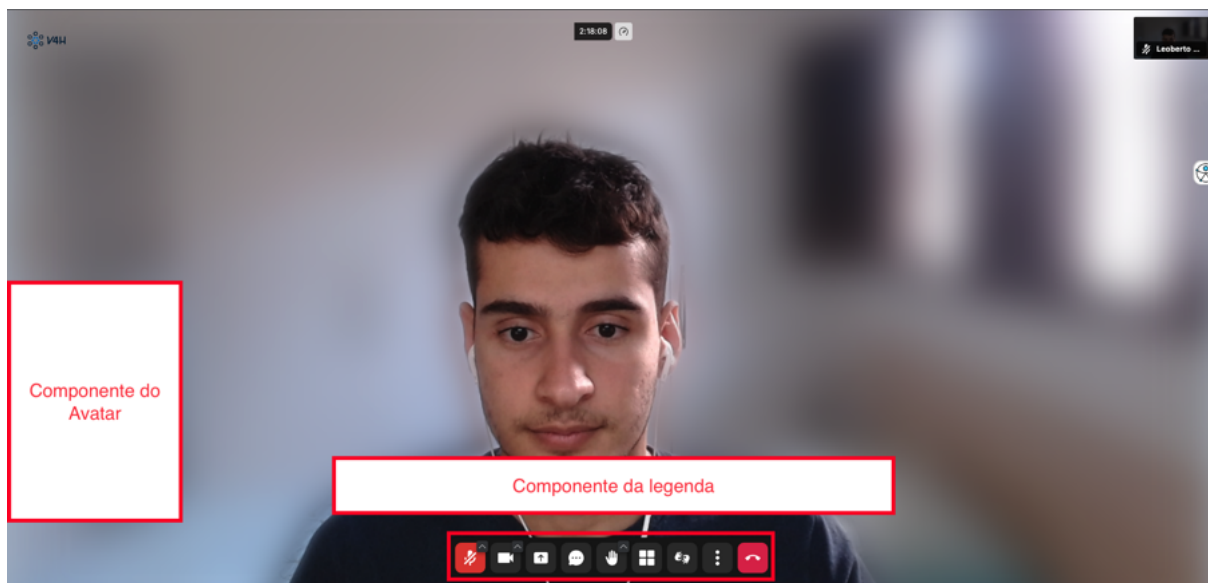


Figura 42: Proposta da interface completa (Fonte: Imagem própria)

Baseado nas análises realizadas e propostas dos componentes mostrados anteriormente, a figura 43 apresenta a implementação da interface gráfica, com o adiç o do componente de Avatar de Libras e legendas, assim como a inserç o do bot o de acessibilidade na barra principal.

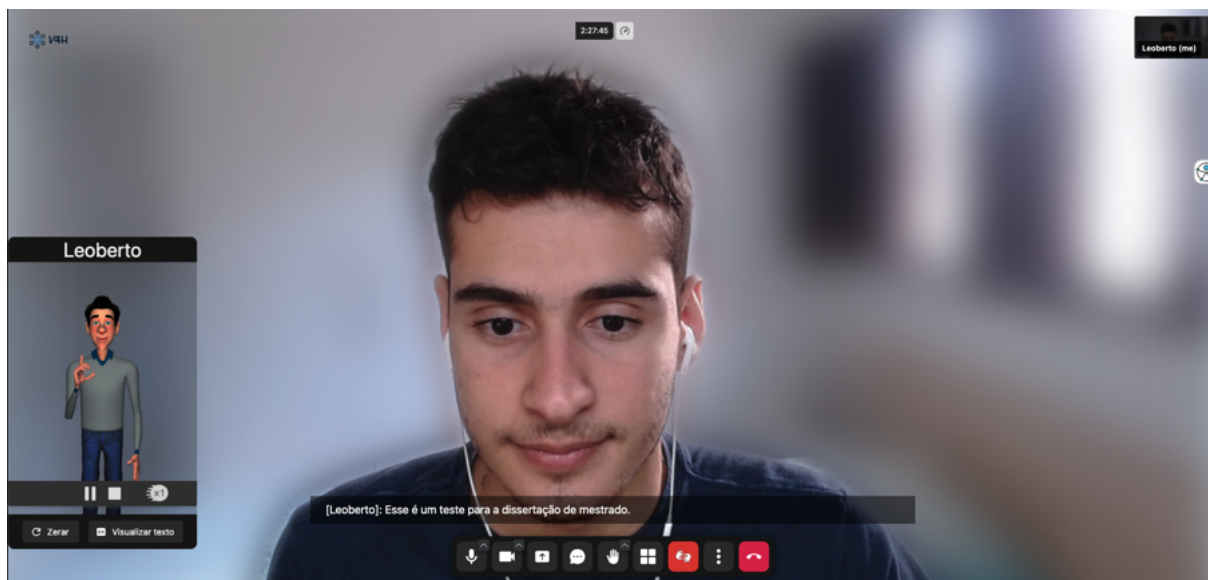


Figura 43: Resultado da implementa o baseada na proposta da interface (Fonte: Imagem pr pria)

4.3 Avatar de Libras

O Avatar de Libras   respons vel pela tradu o da linguagem escrita no idioma portugu s para a linguagem gestual de Libras, o texto transcrito e o seu respons vel s o recebidos de forma s ncrona atrav s do servidor do *Websocket* no formato de *json*.

Algorithm 1 Transcrições para o Avatar de Libras

Require: Conexão com o websocket

```
fila ← array
while Receber transcrição do
    fila ← fila + transcricao
end while
while tamanho(fila) > 0 do
    enviar_para_avatar(transcrição)
    atualizar_nome_do_orador(nome)
    remover_transcrição_da_fila()
end while
```

Como mostra no algoritmo [1](#), quando uma nova mensagem é recebida a transcrição é adicionada ao final da fila de tradução, e quando o Avatar de Libras estiver disponível, será retirado a primeira frase da fila, utilizando a lógica FIFO (*First In First Out*). Com o texto disponível, é encaminhado para o *player* finalizar e o nome do responsável pela aquela transcrição será atualizado e exibido.

O *player* de libras tem três velocidades disponíveis, como a linguagem oral é mais rápida que a linguagem gestual, é indicado que a velocidade escolhida seja a mais alta, na tentativa de compensar essa diferença, caso a velocidade seja superior à que o usuário estiver habituado ele poderá alterar para a que se sentir mais confortável.

4.4 Legendas

As legendas, de forma semelhante ao componente de avatar de libras, recebem a transcrição e o responsável por ela, com a finalidade de adicionar tais dados no fim de uma *string*, contendo todas as falas dos participantes da conferência.

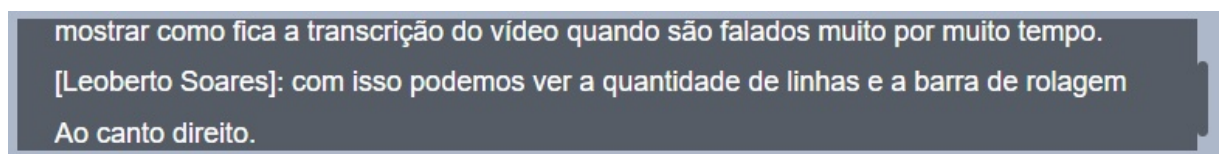


Figura 44: Legenda com no máximo três linhas e rolagem (Fonte: Imagem própria)

Na figura [44](#), o componente das legendas mostrará até três linhas contendo as últimas transcrições recebidas ou geradas localmente, porém existe a possibilidade de visualizar todo o histórico de legenda através da rolagem vertical presente no componente.

Caso o usuário utilize da rolagem para visualizar falas anteriores, as transcrições atuais vão continuar sendo concatenadas no final da texto, para facilitar a sincronização do usuário com as novas adições, o componente de legendas sempre mostrarão as mais

recentes.

4.5 Servidor WebSocket

O servidor *WebSocket* é responsável por armazenar as informações sobre os usuários conectados na conferência e repassar as mensagens no formato de *broadcast* para todos dentro da mesma conferência.

A tecnologia de *WebSocket* foi escolhida em oposição do protocolo HTTP pois a troca de mensagens é no formato bidirecional, já que no momento do recebimento de mensagem originada pelo cliente é necessário enviá-la para os demais do mesmo grupo.

Deste maneira, o servidor foi desenvolvido baseado em quatro eventos:

- *join*: Este evento tem a funcionalidade de adicionar o usuário no grupo da conferência que ele faz parte, ou seja, após acionar o evento, todas as mensagens subsequentes destinadas ao grupo da conferência serão recebidas por esse usuário.
- *left*: Ao acionar o evento de *left*, o usuário será removido do grupo da conferência, fazendo que as mensagens subsequentes relacionadas a conferência atual não sejam recebidas.
- *transcription*: Este evento é utilizado para enviar as transcrições realizadas localmente para os demais participantes da conferência.
- *notification*: Para repassar notificações extras é utilizado o método de *notification*, por exemplo, quando um usuário abre o avatar de libras, é enviado uma notificação para o falante ter ciência que existe uma cliente com o avatar de libras habilitado.

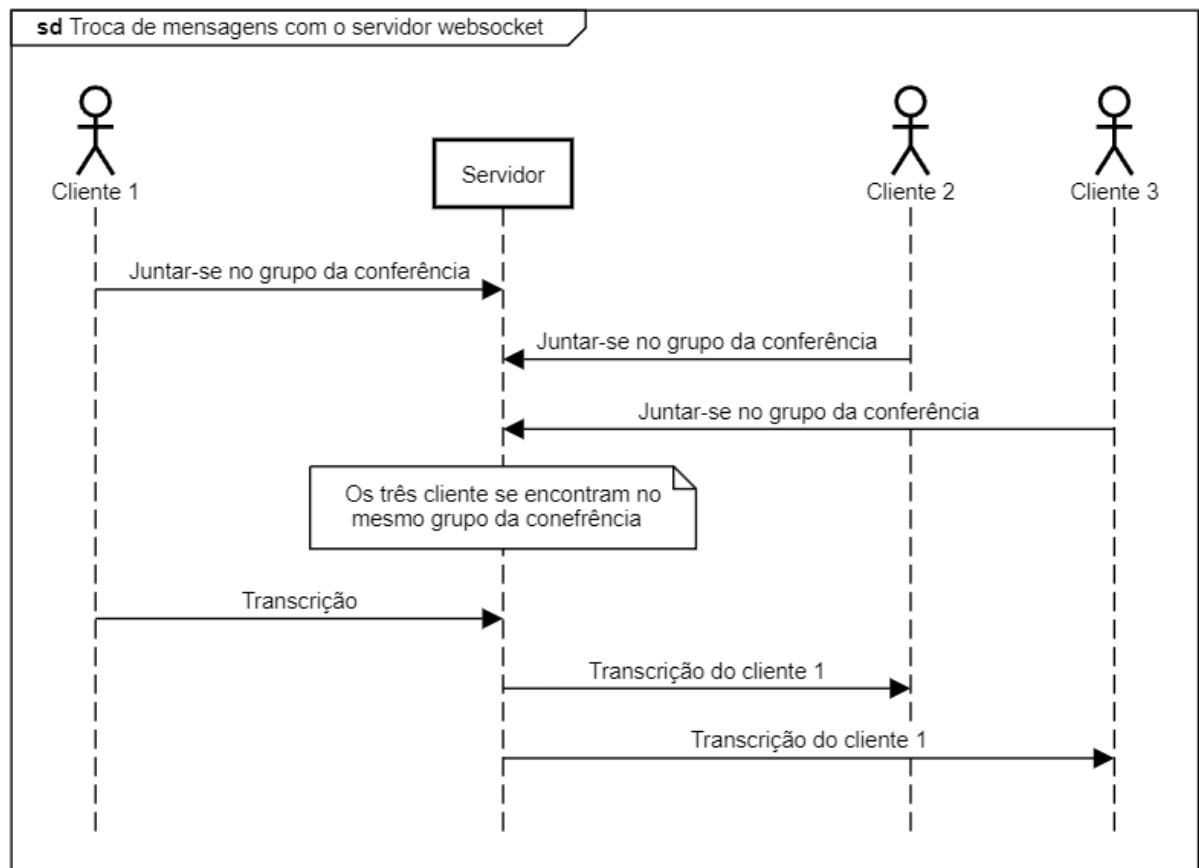


Figura 45: Troca de mensagens entre clientes e servidor (Fonte: Imagem própria)

Na figura 45 é possível visualizar as trocas de mensagens entre os clientes e o servidor, quando todos estão conectados ao *backend* e um deles envia a sua transcrição, os demais recebem e tratam para exibir na legenda e sinalizar no Avatar de Libras.

4.6 Serviço de Sumarização de texto

O serviço de sumarização de texto consiste em uma *REST API* com apenas uma rota, a qual é responsável por sumarizar o texto recebido via JSON, como mostra a documentação *Swagger* na figura 46.

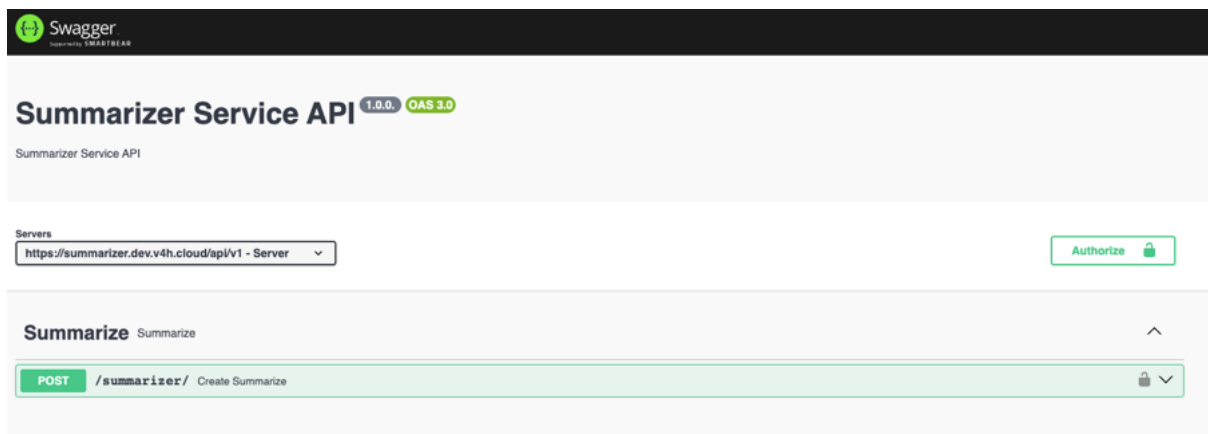


Figura 46: Documentação *Swagger* da API do sumariador (Fonte: Imagem própria)

Para sumarizar o texto é utilizado a *API* do *ChatGPT*, como o *ChatGPT* espera uma entrada textual, é necessário realizar uma solicitação escrita para a inteligência artificial. Na tabela 4, é possível averiguar os parâmetros de configuração utilizados na requisição para *OpenAI*.

Tabela 4: Configuração utilizada da OpenAI

Parâmetro	Descrição	Valor
model	Modelo utilizado da IA	text-davinci-003
prompt	Entrada textual	Simplifique e reduza o texto a seguir:{text}
n	Quantidade de respostas	1
max_tokens	Número máximo de caracteres gerados	256
temperature	Nível de randomização da resposta 0-2	0

4.7 Serviço de Armazenamento de Transcrições

O Serviço de Armazenamento de Transcrição (*TSS - Transcription Storage Service*) é responsável por guardar todas as transcrições geradas durante a reunião no banco de dados. No momento que a transcrição é recebida pelo Servidor *WebSocket*, a mesma é enviada para o microserviço de *TSS* para armazenamento.

O *TSS* é um *REST API* com duas rotas disponibilizadas, uma rota para inserir as transcrições e uma rota para listar as transcrições de uma videoconferência, apresentado na figura 47.

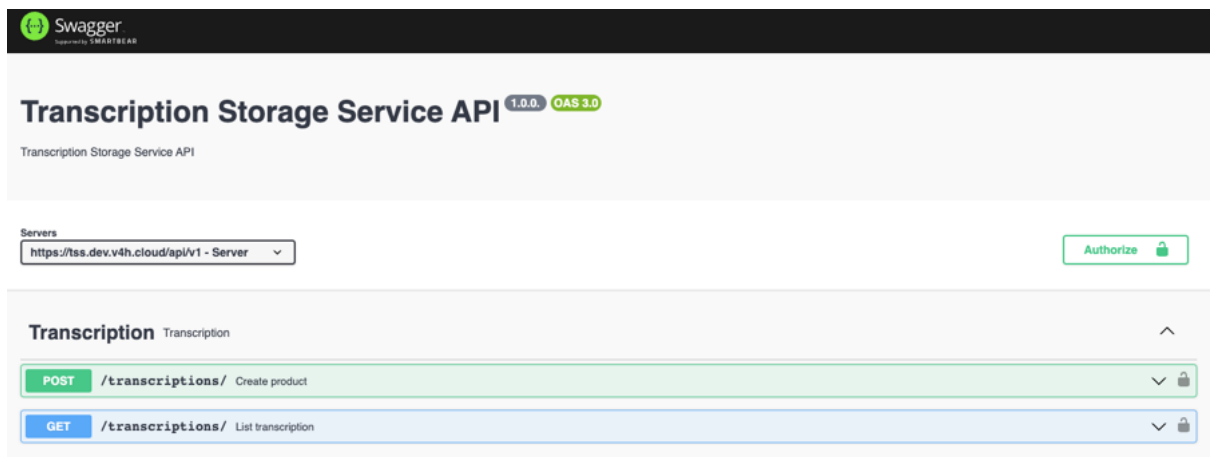


Figura 47: Documentação *Swagger* do *TSS* (Fonte: Imagem própria)

As transcrições são armazenadas no banco de dados não-relacional *MongoDB* com apenas uma coleção de dados chamada *transcription*. Na figura 48 é possível visualizar os dados salvos em um dos testes realizados durante o desenvolvimento.

	_id	compressed	issuer	session	timestamp	transcription
1	64dfb8ac25a56d2efb7f08aa	falhou o do Ricardo foi	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:54.587Z	falhou o do Ricardo foi
2	64dfb89925a56d2efb7f08a9	tem alguma bronca do Lion que tu resolveu que estava pendente	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:57.495Z	tem alguma bronca do Lion que tu resolveu que estava pendente
3	64dfb89425a56d2efb7f08a8	Mas beleza é só para finalizar é Manuel	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:52.883Z	Mas beleza é só para finalizar é Manuel
4	64dfb89825a56d2efb7f08a7	não é isso mesmo	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:28.711Z	não é isso mesmo
5	64dfb88b25a56d2efb7f08a6	pronto certo	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:23.695Z	pronto certo
6	64dfb88425a56d2efb7f08a5	Ah tá eu vejo também eu vejo também	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:16.419Z	Ah tá eu vejo também eu vejo também
7	64dfb87d25a56d2efb7f08a4	que bichão	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:09.951Z	que bichão
8	64dfb87425a56d2efb7f08a3	pronto para o que você tiver feito pode deixar mas não mas não...	Leoberto	rfm297gw92otcqx	2023-08-18T17:55:08.875Z	pronto para o que você tiver feito pode deixar mas não mas não...
9	64dfb86c25a56d2efb7f08a2	deixa deixa esse deixa esse start e os top comigo não precisa...	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:52.707Z	deixa deixa esse deixa esse start e os top comigo não precisa...
10	64dfb86425a56d2efb7f08a1	é faz o seguinte Gustavo bota lá para cima para o zeny Points	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:44.702Z	é faz o seguinte Gustavo bota lá para cima para o zeny Points
11	64dfb85b25a56d2efb7f08a0	Ah tá beleza aí voltou já falei para ele entrar	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:35.576Z	Ah tá beleza aí voltou já falei para ele entrar
12	64dfb85525a56d2efb7f089f	parou de compartilhar a tua tela aqui	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:29.679Z	parou de compartilhar a tua tela aqui
13	64dfb85125a56d2efb7f089e	é emitida quando o novo usuário aqui você entra ou sai da fila	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:25.394Z	é emitida quando o novo usuário aqui você entra ou sai da fila
14	64dfb84a25a56d2efb7f089d	ai o carro pode me atestado você metida quando o atendimento o...	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:18.688Z	ai o carro pode me atestado você metida quando o atendimento o...
15	64dfb84225a56d2efb7f089c	é aí para quem Para os planos deles solicitantes daquela espec...	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:09.907Z	é aí para quem Para os planos deles solicitantes daquela espec...
16	64dfb83b25a56d2efb7f089b	lá para frente no status e metida quando o novo flantonista fi...	Leoberto	rfm297gw92otcqx	2023-08-18T17:54:03.899Z	lá para frente no status e metida quando o novo flantonista fi...
17	64dfb83525a56d2efb7f089a	ai o ai o o carro	Leoberto	rfm297gw92otcqx	2023-08-18T17:53:57.471Z	ai o ai o o carro
18	64dfb82f25a56d2efb7f0899	não é a última mesmo	Leoberto	rfm297gw92otcqx	2023-08-18T17:53:51.649Z	não é a última mesmo
19	64dfb82a25a56d2efb7f0898	é que tá bem no final mesmo	Leoberto	rfm297gw92otcqx	2023-08-18T17:53:46.757Z	é que tá bem no final mesmo
20	64dfb82025a56d2efb7f0897	e	Leoberto	rfm297gw92otcqx	2023-08-18T17:53:36.887Z	e

Figura 48: Visualização dos dados do *TSS* (Fonte: Imagem própria)

Os dados salvos são: o texto original (*transcription*) capturado pelo usuário, quem é o responsável pela fala (*issuer*), identificador da videoconferência (*session*), o *timestamp* do momento da captura da transcrição (*timestamp*) e o texto sumarizado (*compressed*).

5 CENÁRIO E CONFIGURAÇÕES DOS EXPERIMENTOS

Esta seção tem como finalidade apresentar o cenário de experimentos que tem como objetivo avaliar a solução proposta. Para isso, será aplicado o experimento em um conjunto de pessoas surdas afim de comprovar e analisar as hipóteses apresentadas na [Subseção 1.2](#) e complementadas nesta seção.

Segundo [\(Wohlin u. a., 2000\)](#), um experimentos é uma investigação formal, controlada e rigorosa, na qual os sujeitos do experimento são atribuídos a diferentes tratamentos de forma aleatória, manipulando uma ou mais variáveis enquanto outras são fixadas. A fim de testar teses e hipóteses, os resultados entre os diferentes abordagens são comparados.

5.1 Definição do experimento

Para uma boa qualidade dos resultados, é importante definir adequadamente o que se busca entender através dos experimentos. Segundo [\(Wohlin u. a., 2000\)](#) e ilustrado na figura [49](#), os objetivos devem ser processados para que não ocorra retrabalho, ou no pior caso, o experimento não ser utilizado no estudo pretendido.

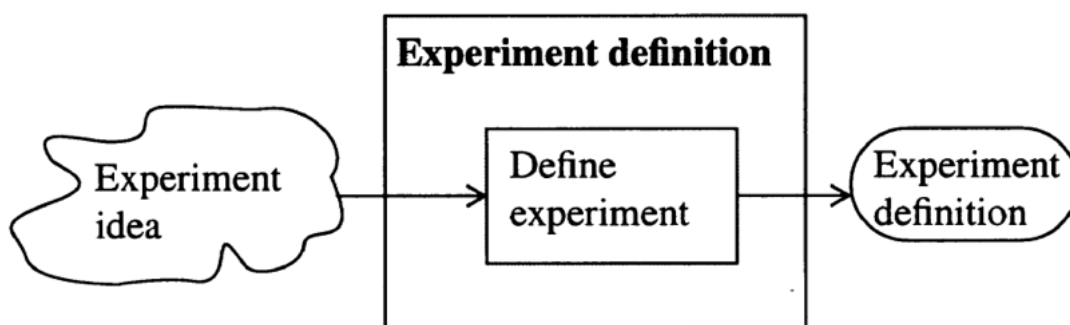


Figura 49: Visão geral do estágio de definição do experimento [\(Wohlin u. a., 2000\)](#)

5.1.1 Objetivo do estudo

O objetivo de estudo é a entidade que o experimento estudará, para este experimento, consiste na interface proposta com componentes de acessibilidade voltados para surdos, com a inclusão de legendas e avatar de Libras.

5.1.2 Propósito

O propósito define qual é a intenção do experimento, portanto, o propósito do experimento desta dissertação compreende na avaliação da interface proposta de videocon-

ferência com componentes de acessibilidades por pessoas surdas, no âmbito da experiência do usuário em videoconferências de propósito geral.

5.1.3 Foco de Qualidade

O foco de qualidade consiste no principal efeito em estudo no experimento, aplicando para o presente experimento, compreende a qualidade da experiência do usuário surdo utilizando a ferramenta proposta com componentes de acessibilidade.

5.1.4 Perspectiva

A perspectiva indica o ponto de vista no qual os resultados serão interpretados, para este experimento os resultados serão interpretados no ponto de vista dos usuários surdos de videoconferência com propósito geral.

5.1.5 Contexto

O contexto é o ambiente no qual o experimento é realizado, definindo os envolvidos e os artefatos de software utilizados, portanto, podemos definir o contexto como usuários surdos utilizando e analisando a proposta de ferramenta de videoconferência com componentes de acessibilidade.

5.2 Planejamento do experimento

Segundo a metodologia de (Wohlin u. a., 2000), após a definição do experimento deve-se planejar a execução do mesmo. Como apresenta a figure 50, a fase do planejamento pode ser dividida em seis etapas.

A primeira consiste na **Seleção de Contexto** que busca selecionar o ambiente no qual o experimento deverá ser executado. Em seguida, as etapas **Formulação de Hipóteses** e **Seleção de Variáveis** são desenvolvidas. A quarta etapa corresponde a **Seleção de Sujeitos**, definindo os participantes do experimento, por fim **Design do Experimento** e **Instrumentação** preparam para implementação prática do experimento.

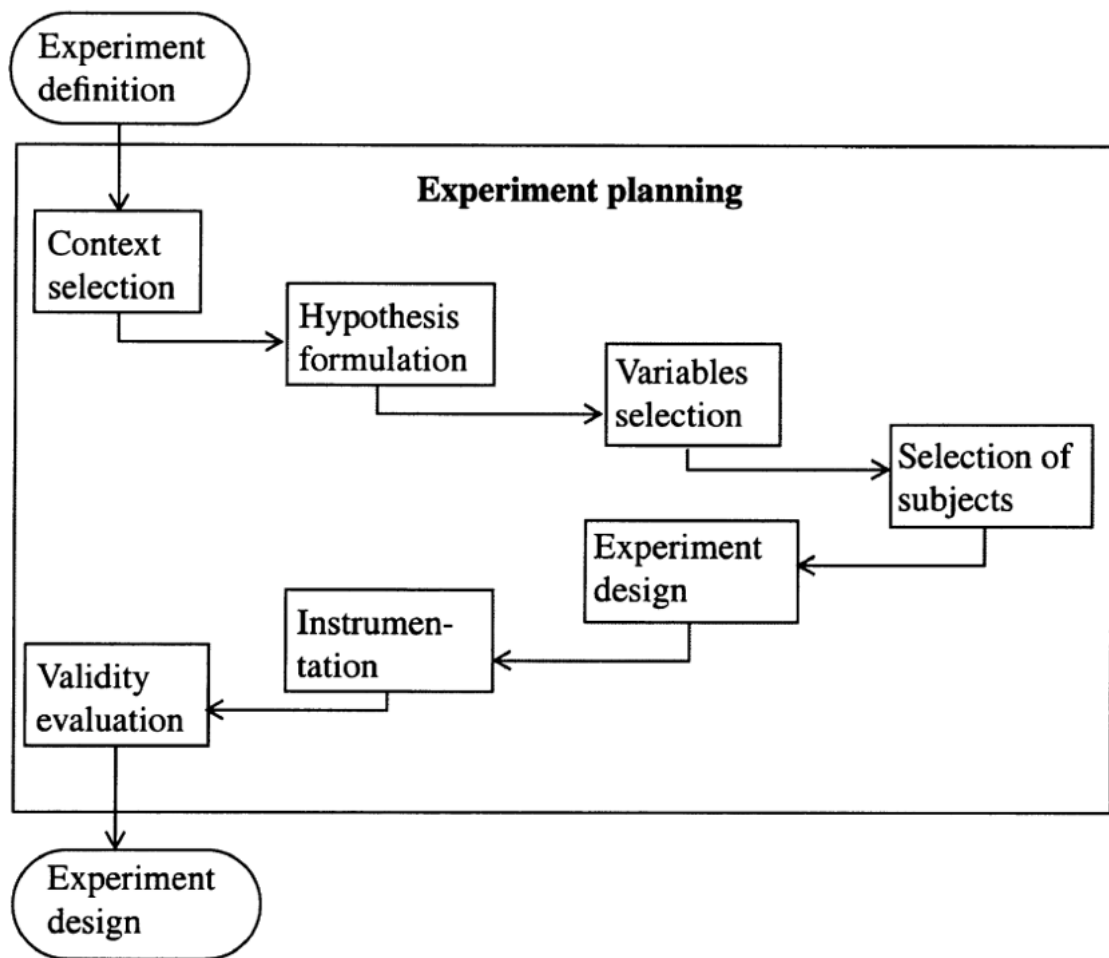


Figura 50: Visão geral do estágio de planejamento do experimento (Wohlin u. a., 2000)

5.2.1 Seleção de Contexto

O ambiente selecionado é de uma reunião remota síncrona, com um conjunto de pessoas surdas e ao menos um apresentador, para simular uma reunião na qual um usuário deseja transmitir algum tipo de conhecimento ou informação.

5.2.2 Formulação de Hipótese

Baseado na hipótese desta dissertação desenvolvida na seção 1.2, é possível apresentá-las da seguinte forma:

1. A implementação proposta ajudará na inclusão e compreensão do conteúdo pelos de surdos em videoconferências através de um avatar de Libras automático.
2. A implementação proposta irá melhorar na inclusão e compreensão do conteúdo

pelos surdos utilizando o avatar de Libras automático em comparação com videoconferências com intérprete humano.

5.2.3 Seleção de Variáveis

Para o experimento proposto, a **variável independente** é a plataforma de videoconferência utilizada durante uma reunião de propósito geral, assumindo as valores de (**ICA**) como a plataforma proposta, (**ITV**) plataformas tradicionais e (**ICI**) com o intérprete humano.

As **variáveis dependentes** consiste nas respostas dos avaliadores perante a variável independente, ou seja, perante a plataforma de videoconferência utilizada por cada um, fazendo com que a qualidade da experiência obtida pelo usuário se configure como a variável dependente.

5.2.4 Seleção de Sujeitos

Os sujeitos deste experimento serão dez voluntários surdos na faixa etária de 20 a 30 anos. Todos os participantes foram escolhidos por conveniência, que ajudam ou ajudaram no desenvolvimento de ferramentas no *Lavid Núcleo de Pesquisa e Extensão*.

5.2.5 Design do Experimento

O design do experimento foi baseado no tipo de um fator com três tratamentos, ou seja, um grupo de pessoas utilizarão a plataforma proposta (**ICA**), um grupo as plataformas tradicionais (**ITV**) e outro grupo com a interface com intérprete (**ICI**). Seguindo as normas estabelecidas por (Wohlin u. a., 2000), os participantes são alocados de forma aleatória para cada tipo de tratamento e um usuário avaliará apenas uma versão para evitar influências de outras versões, além da mesma quantidade de usuários utilizados em ambos os tratamentos a fim de existir um balanceamento dos resultados.

5.2.6 Instrumentação

O experimento consiste na realização de uma videoconferência de propósito geral com duração de 10 minutos, com um ou mais locutor pré-definido. Por questões de disponibilidade, as aplicações do experimento aconteceram em múltiplas datas.

Foi utilizado o trabalho de (Ferraz u. a., 2020) descrito na seção 3.2.1 como o resultado do grupo do **ITV**, no qual foram testados outras seis ferramentas de videoconferência mais tradicionais: Zoom, Google Meet, Microsoft Teams, Jitsi, WebEx e BigBlueButton.

De forma semelhante, os resultados para o grupo **ICI** foram utilizados os dados apresentado na dissertação de (Araújo, 2021).

Para os experimento na plataforma proposta, os participantes surdos participarão no ambiente de casa ou no trabalho, utilizando obrigatoriamente um *desktop* ou *notebook* com o navegador *Google Chrome*, como descrito na seção 4.2.2.

6 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Nesta seção será descrito a execução do experimento, a apresentação os resultados obtidos pelos questionários e a análise desses dados. Os detalhes da definição e planejamento do experimento estão presentes na seção 5, o qual tem a finalidade de provar a hipótese apresentada na seção 1.2.

Como já abordado, o experimento possui três grupos distintos que passaram pelos seguintes tratamentos: (**ICA**) interface com avatar, (**ICI**) interface com intérprete e (**ITV**) interface tradicional de videoconferência. Dentre os três grupos desse experimento, o presente trabalho avaliou a solução **ICA**, proposta desenvolvida nesta dissertação, a qual dispõe do avatar de Libras automático e legendas em tempo real.

Os resultados do grupo **ICI** foram avaliados por (Araújo, 2021), da mesma forma, o grupo **ITV** foram avaliados por (Ferraz u. a., 2020). Os dados de ambos os grupos serão reutilizados para a análise e comparação com a solução **ICA**.

6.1 Execução do Experimento

O experimento conta com 30 pessoas surdas, divididas igualmente em grupos de 10 pessoas para as soluções **ICA**, **ICI** e **ITV**. Do total dos avaliadores, como apresentada na figura 51, 13,3% declararam que possuem deficiência auditiva parcial, e os outros 86,7% afirmaram que possuem deficiência auditiva total.

Nível de deficiência auditiva dos avaliadores

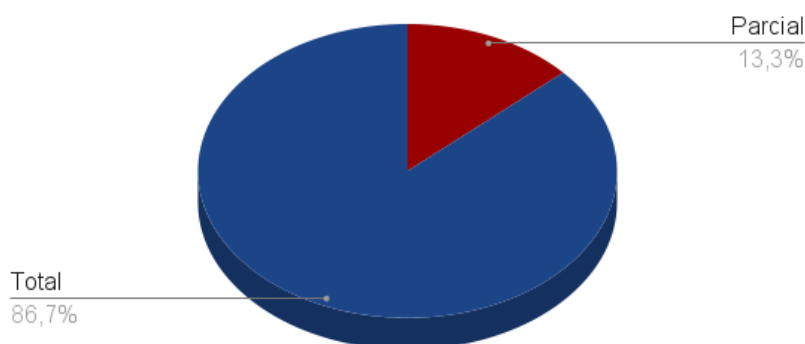


Figura 51: Nível de deficiência auditiva dos avaliadores

Todos os avaliadores do experimento receberam um conjunto de tarefas, visando os critérios de acessibilidade e usabilidade, para realizar durante a videoconferência, citadas anteriormente na seção 3.2.1.

O grupo **ICI**, segundo (Araújo, 2021), foi dividido em dois grupos de 5 pessoas devido a disponibilidade de horários dos avaliadores para a realização do mesmo experimento. O experimento foi aplicado nos dias 4 de março de 2021 e 15 de março de 2021, sem nenhuma metodologia diferente entre eles.

Segundo (Ferraz u. a., 2020), o experimento para o grupo **ITV** aconteceu em dias distintos, e utilizou seis ferramentas de videoconferência diferentes, de forma semelhante ao **ICI**, para cada dia foram utilizados os mesmos passo-a-passo estabelecido nas etapas de definição e planejamento do experimento. A seguir é apresentado as datas da aplicação do experimento, a respectiva plataforma utilizadas e quantidade de avaliadores.

- Zoom; 1 participante; 11 de novembro de 2020.
- Google Meet; 1 participante; 13 de novembro de 2020.
- Jitsi; 2 participantes; 18 de novembro de 2020.
- Microsoft Teams; 1 participante; 19 de novembro de 2020.
- WebEx; 3 participantes; 25 de novembro de 2020.
- BigBlueButton; 2 participantes; 26 de novembro de 2020.

Para o grupo **ICA**, por questões de conflito de horários houve a necessidade de dividir em 5 datas, todas as aplicações do experimento foram conduzidas pela mesma pessoa e a mesma apresentação foi utilizada. A seguir encontra-se as datas e a respectiva quantidade de pessoas presentes:

- 8 de agosto de 2023; 4 participantes.
- 10 de agosto de 2023; 2 participantes.
- 14 de agosto de 2023; 1 participante.
- 16 de agosto de 2023; 2 participantes.
- 17 de agosto de 2023; 1 participante.

As figuras 52a e 52b ilustram a aplicação do experimento para a solução **ICA**, na qual os usuários têm a disposição o avatar de Libras e a legenda. Logo após a palestra e a execução das tarefas pelos participantes, os mesmos respondiam um questionário avaliando a interface em relação a acessibilidade e usabilidade.

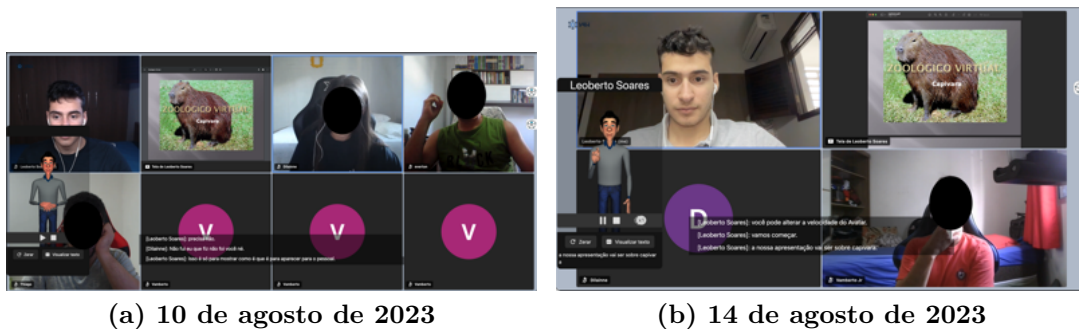


Figura 52: Aplicação do experimento para o solução ICA

As respostas foram obtidos a partir do questionário apresentado por (Ferraz u. a., 2020), adicionando algumas questões referentes ao avatar de Libras, como apresentado no apêndice A. A aplicação do mesmo foi realizada 100% remota utilizando a plataforma de formulários *Google Forms*⁸.

6.2 Análise dos dados

Na tabela 5, o percentual para cada pergunta indica a quantidade de avaliadores que marcaram "sim" em suas respostas, em outras palavras, é o percentual de participantes que obtiveram sucesso na tarefa correspondente a pergunta.

No contexto do acesso a ferramenta, que corresponde as perguntas de 1 a 6 da tabela 5, é possível averiguar que todos os usuários do grupo **ICA** conseguiram adentrar a ferramenta sem problema, isto é justificado pela não necessidade de autenticação na plataforma proposta, da mesma maneira, isso é refletido no grupo **ICI** que também não existe barreiras no momento de entrar na videoconferência.

As questões que diz respeito a entrada na conferência com microfone habilitado e a possibilidade de habilitar ou desabilitar o microfone durante a reunião, a solução **ICA** obteve valores bem abaixo em comparativos as demais soluções, onde 50% e 60%, respectivamente, afirmaram que obtiveram sucesso nesses dois quesitos. Em contra ponto, 90% afirmaram que a câmera estava habilitada ao entrar na videoconferência, 90% conseguiam habilitar e/ou desabilitar suas câmeras, além que 100% dos avaliadores conseguiram perceber se seu microfone e/ou câmera já estavam habilitados, consequentemente superando ou empatando com as soluções **ICI** e **ITV**.

As questões de 7 à 12 da tabela 5, são respectivas a interação do usuário com a plataforma, as quais mostraram um resultado abaixo do esperado para a solução proposta **ICA**. A sinalização de levantar a mão e a lista de participantes foram os quesitos que

⁸<https://docs.google.com/forms>

tiveram o pior desempenho dentre as questões de interação com a plataforma, com respectivamente 60% e 50% respostas afirmativas. A indagação sobre o compartilhamento de tela obteve um total de 70%, conseqüentemente, abaixo das demais soluções apresentadas.

Tabela 5: Resultados do total de participantes que marcaram "sim" para cada questão na respectiva solução

Pergunta	ITV	ICI	ICA
1. Você conseguiu fazer o login na ferramenta?	60%	90%	100%
2. Ao entrar na reunião sua câmera já estava habilitada?	90%	80%	90%
3. Você consegue perceber se seu microfone e/ou câmera estão habilitados?	70%	80%	100%
4. Ao entrar na reunião seu microfone já estava habilitado?	70%	90%	50%
5. Você consegue habilitar e desabilitar o microfone?	80%	80%	60%
6. Você consegue habilitar e desabilitar sua câmera?	80%	90%	90%
7. Você consegue interagir por áudio e/ou vídeo com o apresentador e os demais participantes?	90%	90%	80%
8. Você consegue enviar mensagens por chat?	90%	90%	90%
9. Você consegue sinalizar, de alguma forma, que quer se manifestar durante a reunião (por exemplo, ações como "levantar a mão" durante a reunião)?	60%	90%	60%
10. Você consegue acessar a lista de participantes da reunião?	90%	80%	50%
11. Você consegue acessar os links enviados pelo chat?	90%	100%	100%
12. Você consegue compartilhar sua tela?	80%	90%	70%
13. Você consegue habilitar/desabilitar legendas?	40%	90%	90%
14. Você consegue compreender tudo o que está sendo exibido na tela pelo apresentador?	60%	90%	70%
15. Você consegue sair da reunião por conta própria?	90%	90%	100%

Acerca da compreensão da apresentação pelos participantes, 70% afirmaram que compreenderam tudo que foi apresentado na solução **ICA**, na solução **ITV** e **ICI**, esses valores foram de 60% e 90% respectivamente. O resultado é melhor do que apresentado pelo solução **ITV**, portanto podemos aferir que houve uma melhora na compreensão, mas não de forma satisfatória como apresentado pela solução **ICI**.

Como discutido por (Ferraz u. a., 2020), algumas soluções não disponibilizavam

legendas e outras que disponibilizavam a legenda apenas em inglês, dificultando o entendimento pelos avaliadores. Ilustrado na tabela 5, apenas 40% dos usuários da solução **ITV** obtiveram sucesso, enquanto as soluções **ICA** e **ICI** alcançaram 90% de taxa de sucesso.

No contexto da compreensão do conteúdo, o fato da solução **ICI** ser expressivamente melhor que a solução **ICA** pode ser explicado pelo fato do avatar de Libras ainda não corresponde as expectativas para situações onde a tradução deve ser feita em tempo real, corroborando com o fato do dicionário utilizado pelo avatar ser relativamente pequeno, com apenas 13 mil palavras. Além que os intérpretes conseguem sumarizar adequadamente, cadenciar ou acelerar a sinalização quando necessário e entender contextos que o avatar de Libras ainda não pode.

De forma complementar, na solução **ICA** foi questionado para os avaliadores se conseguiram habilitar o avatar de Libras, como ilustrado na figura 53. 90% dos usuários conseguiram abrir o avatar. A solução **ICI** fez uma pergunta equivalente para o intérprete humano, segundo (Araújo, 2021), obtendo a mesma taxa de sucesso.



Figura 53: Quantidade de avaliadores que afirmaram conseguir habilitar o avatar dentro da solução proposta

A fim de averiguar a importância do avatar, foi questionado ao usuário qual foi a contribuição do avatar de Libras para a compreensão da apresentação, os avaliadores responderam utilizando a escala de *Lickert*, com as seguintes opções: **não contribuiu**, **contribuiu pouco**, **contribuiu razoavelmente**, **contribuiu bem** e **contribuiu muito**.

Qual foi a contribuição do Avatar de Libras para a compreensão da apresentação?

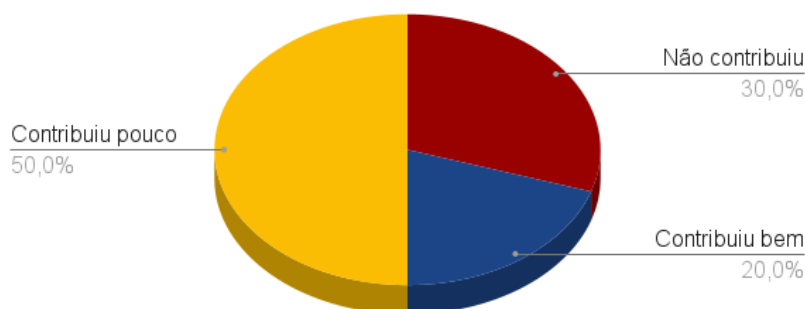


Figura 54: Contribuição do Avatar de Libras para a compreensão da apresentação

Como apresentado na figura 54, apenas 20% afirmaram que o avatar contribuiu bem na compreensão, enquanto 50% indicaram que contribuiu pouco e outros 30% declararam que não contribuiu. Dentre os motivos, o número limitado de palavras do dicionário utilizado e o atraso entre a fala do locutor e a sinalização pelo avatar foram os mais críticos.

O atraso pode ser reflexo da velocidade de sinalização incompatível realizada pelo avatar de Libras, pensando nisso, foi questionado ao usuário se a velocidade de sinalização estava adequada, e como ilustrado na figura 55 apenas 50% dos avaliadores afirmaram que sim.

A velocidade do avatar estava adequado?

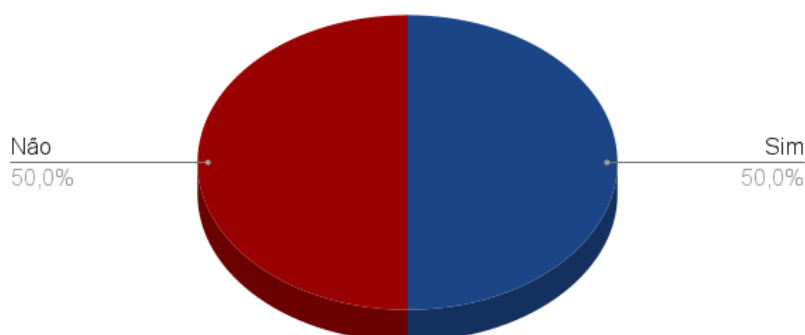


Figura 55: Velocidade do Avatar de Libras

Por fim, no tópico de sair da reunião, todos os usuários da solução **ICA** obtiveram

êxito nesta tarefa, superando as soluções **ICI** e **ITV**, onde ambas obtiveram 90% de taxa de sucesso.

6.3 Avaliando hipóteses

Para avaliar as hipóteses proposta na seção 1.2 e complementada na seção 5 iremos analisar e fazer um comparativo entre as soluções, no gráfico de barras da figura 56 é apresentado todos os resultados obtidos neste trabalho, na dissertação de (Araújo, 2021) e no artigo de (Ferraz u. a., 2020).

Comparativo entre as soluções

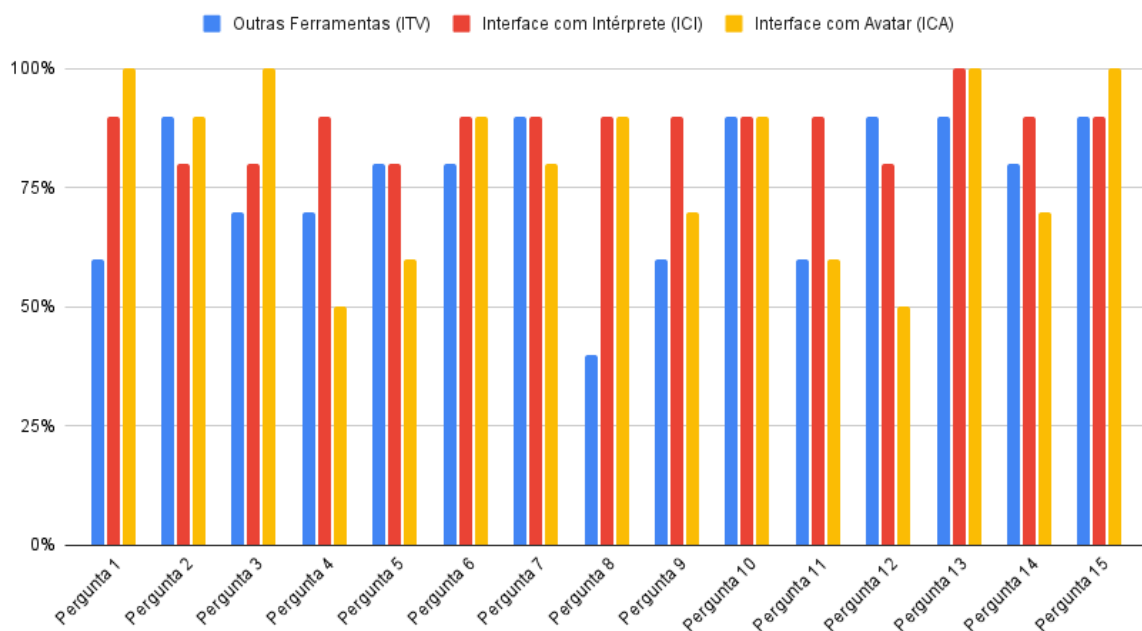


Figura 56: Velocidade do Avatar de Libras

Seguindo a metodologia apresentada o trabalho conduzido por (Araújo, 2021), foi gerado um *box plot* agora comparando as três soluções, a fim de analisar a dispersão dos resultados do experimento. Esta ferramenta gráfica permite a visualização dos dados em relação a cinco estatísticas: o mínimo, o primeiro quartil (Q1), a mediana, o terceiro quartil (Q3) e o máximo.

A partir do gráfico da figura 57, é possível observar que para a demais ferramentas (**ITV**) os valores de mediana, primeiro e terceiro quartil são respectivamente 80%, 65% e 80%. Isso reflete que 50% dos avaliadores obtiveram taxa de sucesso maior que 80%, porém 25% dos avaliadores obtiveram taxa de sucesso de até 65%.

Quando comparado com a solução **ICI**, na qual tinha como principal diferencial a legendas em português e intérprete humano em tempo real, houve uma melhoria con-

siderável, os valores de mediana, primeiro e terceiro quartil respectivamente 90%, 85% e 90%. Consequentemente, isso implica que metade dos participantes obtiveram taxa de sucesso igual ou superior a 90%. O principal aspecto na comparação com as demais ferramentas (**ITV**) é a dispersão dos resultados, no qual a solução **ICI** obteve 80% como o menor valor e 100% como maior valor, em contraste as plataformas tradicionais **ITV** obtiveram respectivamente 40% e 90%.

A solução proposta neste trabalho **ICA**, que contava com legendas e avatar de Libras em tempo real, obteve valores de mediana, primeiro e terceiro quartil respectivamente 90%, 65% e 95%. Assim como a solução **ICI**, metade dos avaliadores obtiveram taxa de sucesso maior ou igual a 90%, porém 25% dos participantes obtiveram taxa de sucesso maior ou igual que 95%, como apresentado pelo valor do terceiro quartil.

Todavia, assim como as plataformas tradicionais (**ITV**) o valor do primeiro quartil da solução proposta (**ICA**) foi de 65%, isso mostra uma dispersão maior dos dados obtidos pelo experimento, no qual obteve valores entre 50% e 100%, em comparação a solução com intérprete (**ICI**).

Em resumo, é possível observar que a solução proposta (**ICA**) e as plataformas tradicionais (**ITV**) obtiveram uma dispersão elevada, e para algumas tarefas os resultados foram negativos, atingindo valores de 50% e 40% respectivamente. Porém, a solução proposta (**ICA**) obteve um desempenho superior em mais tarefas, fazendo com que 25% dos avaliadores obtivessem resultados igual ou maior que 95%, enquanto a solução com intérprete (**ICI**) obteve valores igual ou superior a 90%.

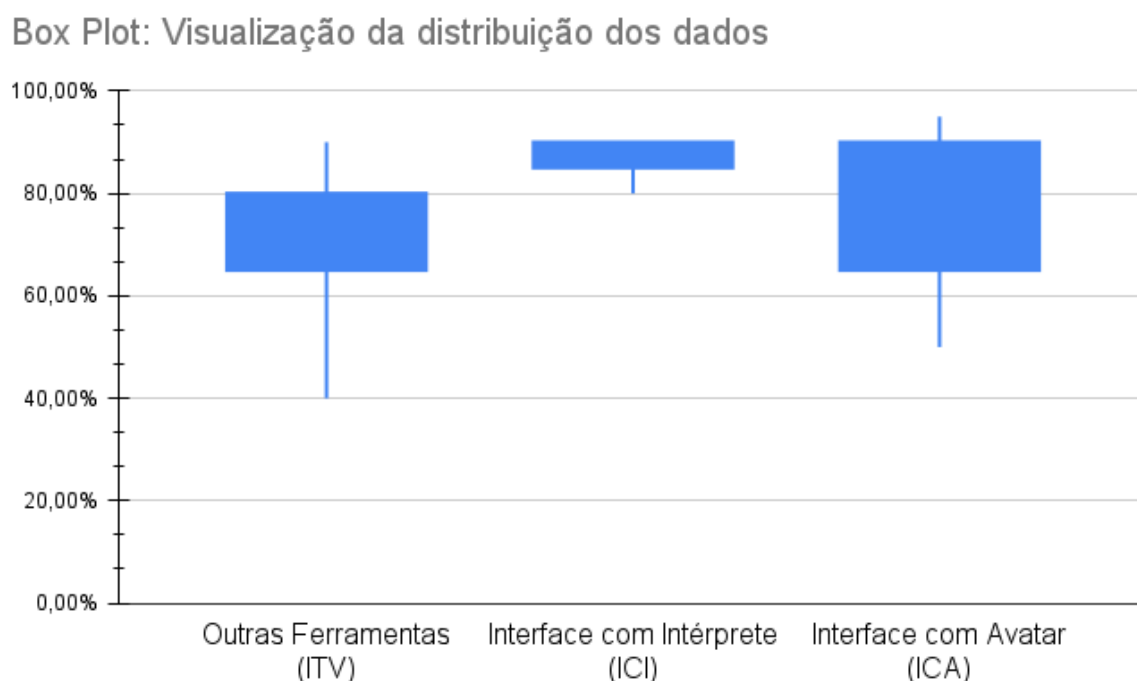


Figura 57: Gráfico *box plot* para visualização de dispersão dos resultados dos experimentos

Após a avaliação dos dados, pode-se dizer, levando em consideração as questões de entendimento do usuário, que a implementação proposta irá melhorar na inclusão e compreensão do conteúdo apresentado em comparação as ferramentas tradicionais de videoconferência. Porém, a mesma conclusão não é verdadeira ao comparar com videoconferências com a presença de intérprete humano.

7 CONCLUSÕES E TRABALHOS FUTUROS

Em uma realidade pós-pandemia, onde as videoconferências se tornaram comuns no dia a dia de boa parte da população mundial, a necessidade de inclusão de pessoas com deficiência nesse meio se mostraram de suma importância.

Pensando nisso, o atual trabalho teve como objetivo o desenvolvimento de uma plataforma de videoconferência acessível para o público surdo, levando em consideração que os intérpretes humanos não estão sempre disponíveis para auxiliar. A solução proposta conta com legendas em tempo real e o componente de Avatar de Libras para a sinalização do que está sendo falado.

O desenvolvimento baseou-se na confecção dos componentes de acessibilidade e aplica-los na plataforma já existente de videoconferência *V4H*, para que tais componentes funcionarem da forma desejada, foi necessário desenvolver *APIs* para sumarização do texto e armazenamento das transcrições, além do servidor *WebSocket* para a troca de mensagens entre os participantes.

Para a validação das hipóteses levantadas nesta dissertação, foi utilizado o método experimental para comparar três grupos que utilizaram soluções distintas de videoconferência, com esse propósito foi aproveitado os resultados obtidos por (Ferraz u. a., 2020), no qual diversas plataformas de videoconferências são testadas a respeito da acessibilidade, de forma semelhante foi aproveitado os dados capturados pela dissertação de (Araújo, 2021), na qual adicionava um intérprete humano na videoconferência.

Ao comparar os dados, é possível observar uma melhora em relações as plataformas presentes no mercado, principalmente no que se diz respeito a compreensão do que foi apresentado nas conferências. Quando comparado a solução com o intérprete humano, a solução do presente trabalho não obteve sucesso, porém se encontra em valores próximos.

Essa distinção pode ser explicada pelo fato do avatar de Libras ainda não estar suficientemente preparado para traduções de texto em tempo real. O tamanho do dicionário, sumarização aplicada e a velocidade do avatar indicam serem os pontos principais desse resultado.

Em resumo, é possível verificar que os resultados foram satisfatórios com uma melhora significativa na compreensão do texto em relação as plataformas tradicionais (*ITV*), e valores semelhantes a proposta com o intérprete humano (*ICI*).

Para trabalhos futuros, a sumarização e a melhoria da tradução do português para Libras devem ser os pontos chaves para que a inclusão de surdos seja aprimorada, além do aumento do dicionário de sinais e a velocidade que os mesmos são apresentados.

Referências

- [Agência Brasil 2021] AGÊNCIA BRASIL: *ABL finaliza 6ª edição do Vocabulário Ortográfico da Língua Portuguesa*. <https://agenciabrasil.ebc.com.br/educacao/noticia/2021-06/abl-finaliza-6a-edicao-do-vocabulario-ortografico-da-lingua-portuguesa>. Juni 2021. – Accessed: 2023-8-11
- [Araújo 2021] ARAÚJO, Matheus Lima Moura d.: *Uma proposta de interface de videoconferência acessível Para Pessoas Com deficiência auditiva*. May 2021. – URL <https://repositorio.ufpb.br/jspui/handle/123456789/21282>
- [Berselli und Magno 2021] BERSELLI, Marcia ; MAGNO, Mariane: *Pensamento Acessível – Cena Aberta – O que tenho a ver com isso?* Editora UFSM, 2021
- [Brasil, Agência 2021] BRASIL, AGÊNCIA: *Vlibras amplia em mais de 30% o número de sinais em 2 anos*. <https://agenciabrasil.ebc.com.br/educacao/noticia/2021-12/vlibras-amplia-em-mais-de-30-o-numero-de-sinais-em-2-anos>. De- zember 2021. – Accessed: 2023-8-11
- [Capovilla u. a. 2005] CAPOVILLA, Fernando ; CAPOVILLA, Alessandra Gotuzo S. ; VIGGIANO, Keila ; MAURICIO, Aline ; BIDÁ, Márcia: Processos logográficos, alfabéticos e lexicais na leitura silenciosa por surdos e ouvintes. In: *Estudos de Psicologia (Natal)* 10 (2005), Apr, Nr. 1, S. 15–23
- [Colling und Boscarioli] COLLING, João P. ; BOSCARIOLI, Clodis: *Avaliação de Tec- nologias de Tradução Português-Libras visando o Uso no Ensino de Crianças Surdas*. – URL <https://doi.org/10.22456/1679-1916.53550>
- [CRISTIANO 2015] CRISTIANO, ALMIR: *O que é Libras?* April 2015. – URL <https://www.libras.com.br/>. – Zugriffsdatum: 2022-07-25
- [Deng und Lin 2023] DENG, Jianyang ; LIN, Yijia: The benefits and challenges of ChatGPT: An overview. In: *Frontiers in Computing and Intelligent Systems* 2 (2023), Nr. 2, S. 81–83
- [Diniz u. a. 2009] DINIZ, Debora ; BARBOSA, Livia ; SANTOS, Wederson Rufino d.: Deficiência, direitos humanos e justiça. In: *Sur Rev. Int. Direitos Hum.* 6 (2009), Nr. 11, S. 64–77
- [Emilio 2020] EMILIO, P.: *O que a pandemia nos ensinou sobre a telemedicina?* <https://medicinas.com.br/pandemia-telemedicina/>. 2020. – Acesso em: dia 24 de Nov de 2020.

- [Felipe 2020] FELIPE, D.: *IBGE pessoas com deficiência auditivas*. <https://www.ibge.gov.br/apps/snig/v1/?loc=0cat=-1,-2,-3,128ind=4643>. 2020. – Acesso em: dia 22 de Jul de 2022.
- [Ferraz u. a. 2020] FERRAZ, Reinaldo ; CUENTRO, Ana ; PEREIRA, Bruno W. ; NASCIMENTO, Cláudia M. ; CAMPOS, Thiago P. de: *Pesquisa sobre acessibilidade das ferramentas de videoconferência em plataforma web*. Dec 2020. – URL <https://www.nic.br/media/docs/publicacoes/16/20201211144102/Pesquisa%20-%20Ferramentas%20de%20videoconfere%CC%82ncia%20FINAL.pdf>
- [de Góes 2020] GÓES, Maria Cecília R. de: *Linguagem, surdez e educação*. Autores Associados, 2020
- [Jitsi Meet 2020] JITSI MEET: *Jitsi Meet*. <https://meet.jit.si/>. 2020. – Acesso em: dia 24 de Nov de 2020.
- [Karine 2020] KARINE, M.: *CNJ comemora sucesso de audiências e sessões por videoconferência, Agência Brasil*. <https://agenciabrasil.ebc.com.br/justica/noticia/2020-08/cnj-comemora-sucesso-de-audiencias-e-sessoes-por-videoconferencia>. 2020. – Acesso em: dia 24 de Nov de 2020.
- [Khattar u. a. 2020] KHATTAR, A. ; JAIN, P. R. ; QUADRI, S. M. K.: Effects of the Disastrous Pandemic COVID 19 on Learning Styles, Activities and Mental Health of Young Indian Students - A Machine Learning Approach. In: *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, S. 1190–1195
- [Microsoft 2022] MICROSOFT: *Record and transcribe audio*. 2022. – URL <https://support.microsoft.com/en-us/office/transcribe-your-recordings-7fc2efec-245e-45f0-b053-2a97531ecf57>
- [Organization 2021] ORGANIZATION, World H.: *Deafness and hearing loss*. 2021
- [Patel u. a. 2020] PATEL, Bhavinkumar D. ; PATEL, Harshit B. ; KHANVILKAR, Manthan A. ; PATEL, Nidhi R. ; AKILAN, Thangarajah: ES2ISL: An advancement in speech to sign language translation using 3D avatar animator. In: *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, IEEE, 2020
- [Quadros und Becker Karnopp 2007] QUADROS, Ronice Müller de ; BECKER KARNOPP, Lodenir: *Língua de sinais brasileira: estudos lingüísticos*. Artmed, 2007
- [Ribeiro 2019] RIBEIRO, Giovanna: *Especialistas pedem mais intérpretes de Libras em Universidades; MEC Aponta dificuldades -*

- Notícias*. 2019. – URL <https://www.camara.leg.br/noticias/577880-especialistas-pedem-mais-interpretres-de-libras-em-universidades-mec-apont>
- [Richard 2020] RICHARD, S.: *The Distributed Workforce: A QA With Zoom's Global CIO*. <https://www.forbes.com/sites/insights-zoom/2020/10/26/the-distributed-workforce-a-qa-with-zooms-global-cio/?sh=63ee83d06c35>. 2020. – Acesso em: dia 24 de Nov de 2020.
- [V4H 2019] V4H: *Video For Health*. <https://v4h.cloud>. 2019. – Acesso em: dia 24 de Nov de 2020.
- [Web APIs — MDN 2022] WEB APIs — MDN: *Using the web speech API - web apis: MDN*. 2022. – URL https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API/Using_the_Web_Speech_API
- [Wilcox 2022] WILCOX, Sherman: *Aprender a ver*. 08 2022. – ISBN 9788589002103
- [Wohlin u. a. 2000] WOHLIN, Claes ; RUNESON, Per ; HÖST, Martin ; OHLSSON, Magnus C. ; REGNELL, Björn ; WESSLÉN, Anders: *Experimentation in Software Engineering*. Boston, MA : Springer US, 2000

APÊNDICE A

QUESTIONÁRIO

1. Qual é o seu nome?
R: _____
2. Qual é a sua experiência com acessibilidade digital? Por acessibilidade digital, entenda o uso de tecnologias assistivas para superar barreiras de acesso em sites e aplicativos
☐ Inciante
☐ Intermediário
☐ Avançado
3. Qual é o seu grau de deficiência auditiva?
☐ Parcial / baixa audição
☐ Total / surdez
4. Você conseguiu fazer o login na ferramenta?
☐ Sim
☐ Não
5. Ao entrar na reunião, sua câmera já estava habilitada?
☐ Sim
☐ Não
6. Você consegue perceber se seu microfone e/ou câmera estão habilitados?
☐ Sim
☐ Não
7. Ao entrar na reunião, seu microfone já estava habilitado?
☐ Sim
☐ Não
8. Você consegue habilitar e desabilitar o microfone?
☐ Sim
☐ Não
9. Você consegue habilitar e desabilitar a câmera?
☐ Sim
☐ Não
10. Você consegue interagir por áudio e/ou vídeo com o apresentador e os demais participantes?

- ☐ Sim
☐ Não
11. Você consegue enviar mensagens por chat?
☐ Sim
☐ Não
12. Você consegue sinalizar de alguma forma que quer se manifestar durante a reunião (por exemplo, ações como “levantar a mão” durante a reunião)?
☐ Sim
☐ Não
13. Você consegue acessar a lista de participantes da reunião?
☐ Sim
☐ Não
14. Você consegue acessar os links enviados pelo chat?
☐ Sim
☐ Não
15. Você consegue compartilhar sua tela?
☐ Sim
☐ Não
16. Você consegue habilitar/desabilitar legendas?
☐ Sim
☐ Não
17. Você consegue habilitar/desabilitar o avatar de libras?
☐ Sim
☐ Não
18. Você consegue compreender tudo o que está sendo exibido na tela pelo apresentador?
☐ Sim
☐ Não
19. Em que grau o avatar de Libras contribuiu para a compreensão do que foi apresentado na videoconferência?
☐ Não contribuiu
☐ Contribuiu pouco
☐ Contribuiu razoavelmente
☐ Contribuiu bem
☐ Contribuiu muito

20. A velocidade do avatar estava adequada?

☐ Sim

☐ Não

21. Você consegue sair da reunião por conta própria?

☐ Sim

☐ Não