

# Hand-Speaking Technology: An Expert and User-Based Assessment of the *VLibrasWidget*

Thaís Melquíades Macêdo

Centro de Informática, Universidade Federal da Paraíba  
João Pessoa, Brasil  
thaismelquiades4@gmail.com

Carolina de Melo Costa

Centro de Informática, Universidade Federal da Paraíba  
João Pessoa, Brasil  
carolina.costa@lavid.ufpb.br

Vinícius Germano

Centro de Informática, Universidade Federal da Paraíba  
João Pessoa, Brasil  
acadviniciusgermanodb@gmail.com

Yuska Paola Costa Aguiar

Centro de Informática, Universidade Federal da Paraíba  
João Pessoa, Brasil  
yuska@ci.ufpb.br

## ABSTRACT

The use of assistive technology to support the inclusion of deaf individuals in the digital environment is a valuable resource when it becomes necessary to scale the translation services typically provided by human interpreters. *VLibrasWidget* is an open-source computational solution designed to translate digital content (text, audio, and video) into Brazilian Sign Language (Libras) through animation. This tool is available on more than 120,000 websites, with an average of 100,000 translations per day. Given the widespread use and social impact of *VLibrasWidget*, it is important to consider that usability and accessibility aspects are essential for its effective use. Therefore, this experience report presents the planning, execution, and results of an evaluation of *VLibrasWidget*, combining a heuristic usability review by experts with user testing involving both deaf and hearing participants—through usage observation and subjective satisfaction probing. The complementary results enabled the identification of 28 interface and interaction *issues* with varying severity levels (1 high; 8 medium; 10 low; 9 cosmetic), indicating that the tool has room to evolve in order to better serve its users.

## KEYWORDS

Usability Test; Heuristic Evaluation; Machine Translators to Libras; Brazilian sign language (Libras); Accessibility; Deaf user; Quality Assurance

## 1 Introdução

A comunicação humana tem por base a língua como elemento fundamental, sendo esta essencial para a expressão de pensamentos, sentimentos e intenções, além de possibilitar a transmissão e preservação do conhecimento [5]. A língua, para além da comunicação, permite o acesso à informação, entretenimento, educação, saúde e direitos, sendo essencial em todas as áreas da vida.

Segundo os dados da Organização Mundial da Saúde (OMS) aproximadamente 20% da população global vive com algum grau de perda auditiva [26]. No Brasil, de acordo com dados do IBGE, aproximadamente 5% da sua população apresenta algum grau de deficiência auditiva, havendo 1,1% com surdez profunda [12].

A comunidade surda no Brasil tem a Língua Brasileira de Sinais (Libras) como língua oficial. Esta tem um papel fundamental na inclusão social e na promoção da comunicação entre surdos e ouvintes [8]. Legalmente, o uso de Libras é um direito da população

com deficiência auditiva, legitimada pela Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência) (Nº 13.146, Brasil, e 6 de Julho de 2015) [13] e a Lei de Acesso à Informação (Nº 12.527, Brasil 2011) [1].

Para promoção da acessibilidade à conteúdos digitais e não digitais, assim como para facilitar a comunicação entre pessoas surdas e ouvintes, alguns recursos de Tecnologia Assistiva (TA) vêm sendo desenvolvidos nos últimos anos. As soluções podem ser adotadas como mediadores na construção do conhecimento na promoção da autonomia de pessoas surdas, reduzindo as barreiras de acesso à informação. Estes provêm acesso a conteúdos em línguas de sinais nas Web, em dispositivos móveis, na TV, no cinema, em vídeo-chamadas, entre outros. O processo de tradução consiste em interpretar um texto em uma língua-fonte e reformular esse texto para a língua-alvo, de modo a tornar a tradução compreensível para o público ao qual se destina. A tradução pode acontecer por humanos (intérpretes) ou automaticamente. Dentre as propostas existentes de soluções de tradução para Línguas de Sinais (LS), a saber: HandTalk<sup>1</sup>, Rybená<sup>2</sup>, Suíte *VLibras*<sup>3</sup>, ICOM<sup>4</sup>.

Especificamente a *Suíte VLibras*, consiste em um conjunto de ferramentas computacionais que traduz automaticamente conteúdos digitais de texto, áudio e vídeo para Libras. Seu propósito é tornar computadores, *smartphones* e plataformas Web mais acessíveis para pessoas surdas. A *Suíte VLibras* é composta por diferentes soluções, a saber: *VLibrasWidget*, *VLibrasPlugin*, *VLibrasMóvel* (Android e iOS), *VLibras Tradutor*, *VLibras Dicionário*, *VLibras Vídeo*, *WikiLibras*, entre outros. Todos os seus produtos são softwares livres e de código aberto, e que são disponibilizados gratuitamente à sociedade. O *corpus* do dicionário do *VLibras* compreende mais de 21 mil sinais cadastrados, com 90 mil frases treinadas por inteligência artificial. As versões mobile, Android e iOS, contam com mais de 100 mil downloads. O *VLibrasWidget*, versão disponível para web, está disponível em mais de 120 mil sites, com uma média de 100 mil traduções por dia.

A *Suíte VLibras* vem sendo amplamente utilizada para prover acessibilidade para a comunidade surda em diversos sites do Poder Executivo, Legislativo e Judiciário, além de vários sites comerciais. Em 2019 o *VLibrasWidget* foi implantado em todos os sites de

<sup>1</sup><https://www.handtalk.me/br/>

<sup>2</sup><https://rybena.com.br>

<sup>3</sup><https://vlibras.gov.br/app/>

<sup>4</sup><https://www.icom.app/>

governo que implementam a barra de governo. Seu uso em sites relacionados com órgãos e serviços públicos digitais, em conjunto com uma massiva adoção da suíte em sites e serviços privados, coloca os recursos de acessibilidade da plataforma à disposição dos cidadãos de forma facilitada. Isso se reflete em cerca de 40 milhões de acessos por dia ao *VLibrasWidget* nos quase 3 milhões de traduções por mês, em média.

Dessa forma, considerando a relevância do *VLibras* para auxiliar na promoção da acessibilidade em conteúdos digitais à comunidade surda, faz-se necessário que as tecnologias desenvolvidas continuem a evoluir para melhor atender a comunidade surda, para evitar que o seu funcionamento seja interrompido, bem como para acompanhar as mudanças tecnológicas e as políticas de uso dos navegadores, e sistemas operacionais dos *smartphones* e computadores pessoais.

Adicionalmente, um dos principais desafios dos sistemas de tradução automática para língua de sinais é garantir que o conteúdo disponibilizado aos surdos chegue com a mesma consistência e qualidade do original, permitindo assim o entendimento adequado da mensagem. Essas iniciativas vêm sendo desenvolvidas e reportadas na literatura [20–22]. Além disso, a pesquisa e desenvolvimento para a melhoria da experiência de uso nas aplicações de tradução para língua de sinais desempenham um papel crucial na adoção e efetividade da ferramenta. Interfaces mais intuitivas, maior personalização e um ambiente digital acessível contribuem para uma interação mais natural, promovendo maior engajamento e conexão da comunidade surda com a tecnologia. Neste sentido, também existem referências na literatura. Sob as perspectivas de usabilidade, acessibilidade e experiência de uso, as avaliações conduzidas e reportadas na literatura são, prioritariamente, para a ferramenta *VLibrasMóvel* [2, 21, 23]. A avaliação do *VLibrasWidget* foi conduzida considerando a participação de intérpretes de Libras e pessoas surdas para o contexto específico de acessibilidade em provas teóricas para obtenção de carteira nacional de habilitação [11].

Do exposto, este artigo contempla o relato da experiência da avaliação multidimensional [9, 10] conduzida para o *VLibrasWidget* compreendendo a condução de Inspeção Heurística de Usabilidade [17], Testes de Usabilidade com Observação da Interação [4, 15] e Sondagem da Satisfação Subjetividade - considerando tanto a experiência de especialistas, quanto a percepção de uso da aplicação por parte de usuários surdos e ouvintes. O objetivo é identificar potenciais problemas de usabilidade e barreiras para o uso do *VLibrasWidget*. A partir de então, é possível conhecer quais aspectos podem ser aprimorados na ferramenta enquanto estratégias de melhoria da experiência de uso. E, neste sentido, alinhar práticas observacionais e qualitativas para o contexto da ferramenta e ampliar o alcance dos resultados para diferentes perfis de usuários do *VLibras*.

## 2 Trabalhos Relacionados

A literatura relacionada ao desenvolvimento, uso e avaliação de tecnologias assistivas é vasta [19], com uma concentração de estudos direcionados para público com deficiência visual e auditiva, face à representatividade da população concernente. Dentro do escopo de TA destinado à tradução automática de português para Libras: Colling *et al* [7] relatam a avaliação realizada com os tradutores

HandTalk, ProDeaf e Rybená em contexto interpretativo pelo viés da criança surda; Vieira *et al* [25] descrevem os resultados de um experimento no qual usuários surdos e ouvintes elaboraram 29 narrativas a partir dos aplicativos HandTalk e ProDeaf para analisar o tratamento de “Expressões Não-Manuais” dos avatares; e de Araújo e de Oliveira [14] conduziram um estudo que permitiu a identificação de quinze problemas relacionados ao tratamento de contexto no processo da tradução do Português para Libras para a Suíte *VLibras*. Especificamente para a Suíte *VLibras*, o conceito de qualidade vem sendo reportado na literatura científica de forma abrangente.

Os trabalhos de Reis *et al* [20, 22] compreendem que há aspectos gramaticais considerados relevantes em Libras cujo tratamento inadequado pode interferir na compreensão das sentenças traduzidas por parte dos usuários, levando a uma má experiência de uso. São eles: desambiguação, advérbios de intensidade, verbos direcionais e tipos de frases (interrogativas e exclamativas). Participaram do estudo 30 pessoas que atuam como intérpretes em Libras. Estes estudos evidenciam a importância de garantir que o conteúdo traduzido seja compreensível, mantendo a fidelidade e o significado original. Essas metodologias avaliam tanto a precisão linguística quanto a qualidade do sinal, essenciais para a compreensão do conteúdo pelos usuários surdo.

Reis *et al* [21] propõe e descreve uma metodologia que fornece recursos para avaliar a qualidade de tradutores automáticos considerando, de forma gradativa e complementar, a execução de avaliações técnicas baseadas em métricas computacionais, com a participação de usuários especialistas em Libras (intérpretes), assim como com usuários surdos. Um estudo de caso foi realizado utilizando o *VLibras*, envolvendo 6 intérpretes e 10 pessoas surdas. A metodologia proposta e seu ferramental apoiam a condução de avaliações passíveis de replicação para fins comparativos entre a qualidade de tradução a partir de diferentes critérios.

Aguiar *et al* [23] discute a análise de comentários espontâneos fornecidos por usuários do *VLibrasMóvel*, disponíveis nas lojas virtuais de aplicativos, para as versões Android e iOS. A análise do *feedback* foi conduzida de maneira manual e automática, de acordo com a metodologia MALTU. A análise contínua dos comentários dos usuários é sugerida como uma boa prática para garantir a satisfação do usuário a longo prazo e a qualidade da ferramenta.

Aguiar *et al* [2] contempla o planejamento, condução e os resultados de uma inspeção heurística de acessibilidade do *VLibrasMóvel* de acordo com a ABNT NBR 17060 [3]. As pessoas autoras destacam a existência de recomendações que não são aplicáveis ao contexto específico de uma aplicação voltada para surdos. O texto é um relato de experiência da avaliação realizada, seus desafios, resultados e lições aprendidas.

O estudo [11] relata a avaliação para um sistema de prova teórica para obtenção da CNH que integra o *VLibras Web*. A avaliação compreendeu tanto inspeção heurística, quanto testes exploratórios de usabilidade com usuários surdos e intérpretes de Libras, assim como a sondagem da satisfação subjetiva desses participantes na interação com o sistema. Mesmo que o *VLibras* não tenha sido o objeto de estudo principal, foi possível identificar aspectos de melhoria, sobretudo evidenciando a necessidade de sinalização de termos técnicos ao exame nacional relativo às regras de trânsito.

Este relato, por sua vez, tem como foco a aplicação de uma avaliação multidimensional da ferramenta *Widget* - mais utilizada pelo grande público por sua disponibilidade em sites do governo (mas não exclusivo a estes).

### 3 Metodologia

O objetivo deste estudo consiste em realizar uma avaliação multidimensional do *VLibrasWidget*, considerando a perspectiva de especialistas em Interação Humano-Computador e Acessibilidade, assim como de pessoas usuárias da aplicação. Para tanto, duas abordagens complementares foram adotadas de forma sequencial.

Em um primeiro momento foi realizada uma Inspeção Heurística de Usabilidade com base nas heurísticas de Nielsen [15]. A partir desse método é possível identificar quais heurísticas são violadas, em quais partes da aplicação e estabelecer potenciais barreiras de uso, com diferentes níveis de severidade. Para cada *issue* identificada, sugestões de correção e/ou melhoria, são propostas para evolução do sistema.

Em seguida, a condução de Testes com Usuários [24] foi planejada e executada com foco na observação de usuários durante a interação com o *VLibrasWidget* na realização de um conjunto de tarefas pré-definidas. Neste contexto, são identificados aspectos de interação que afetam a Experiência de Uso, positiva ou negativamente. Ao elaborar as tarefas de teste com base nas heurísticas violadas, e *issues* de usabilidade derivadas delas, pode ser realizada uma análise de impacto destas no uso do *VLibrasWidget*.

O perfil das pessoas participantes do teste com usuários e a sondagem da satisfação subjetiva no uso do *VLibrasWidget* são outras duas dimensões consideradas na avaliação. A correlação entre todas as dimensões permite uma visão abrangente dos problemas identificados e sua gravidade - informações úteis para evolução da aplicação a partir da implantação de estratégias de melhoria que aprimorem a sua qualidade geral e experiência de uso.

Para a avaliação com usuários, três perfis de uso foram considerados: **P1: Pessoa surda** e usuária de Libras (perfil específico); **P2: Pessoa iniciante no uso do *VLibrasWidget*** (perfil genérico, ouvinte); e **P3: Pessoa com conhecimento técnico** nas áreas de Interação Humano-Computador, Design e/ou Avaliação de Interface e Computação (perfil genérico com conhecimento técnico).

Deve-se destacar que, independente do perfil da pessoa participante, busca-se compreender como a interface e as interações propostas pelo *VLibrasWidget* são utilizadas e percebidas na perspectiva da experiência de uso. No entanto, considerando o público específico da pessoa surda, espera-se identificar aspectos mais específicos sobre a tradução realizada, sendo estes achados secundários a este estudo. Ainda, considerando a participação de pessoas na avaliação, o projeto para sua execução foi aprovado pelo Comitê de Ética em Pesquisa, por meio da Plataforma Brasil, na Universidade Federal da Paraíba (UFPB).

#### 3.1 Avaliação Heurística de Usabilidade

A inspeção heurística é um método de avaliação que se apoia em um conjunto predefinido de recomendações ou critérios, que devem ser verificados em um sistema, sob a ótica de um ou mais especialistas. Como resultado são indicadas as conformidades e/ou violação para cada uma das recomendações, sendo possível, a partir de análise

posterior, o agrupamento destas para definição de problemas e proposição de solução.

As 10 Heurísticas de Usabilidade propostas por Nielsen [15], são utilizadas para diferentes domínios de aplicação e interfaces interativas, por sua natureza e abrangência. São elas:

- **H1. Visibilidade do status do sistema:** O sistema deve sempre manter os usuários informados sobre o que está acontecendo, através de *feedback* visível e compreensível, com tempo adequado para a reação.
- **H2. Correspondência entre o sistema e o mundo real:** O sistema deve falar a linguagem dos usuários, utilizando termos e convenções familiares. A interação deve refletir as situações do mundo real, facilitando a compreensão.
- **H3. Controle e liberdade do usuário:** O sistema deve permitir que os usuários façam ações e as desfçam facilmente. Deve haver uma "desfazer" ou "voltar" sempre disponível, para evitar frustração.
- **H4. Consistência e padrões:** Os usuários não devem se perguntar se diferentes palavras, situações ou ações significam a mesma coisa. O sistema deve ser consistente em seus padrões de interface e comportamentos.
- **H5. Prevenção de erros:** O sistema deve ser projetado de forma que erros sejam prevenidos antes que ocorram, ou pelo menos que a chance de erro seja minimizada com escolhas claras e explicativas.
- **H6. Reconhecimento em vez de memorização:** Minimizar a carga cognitiva do usuário, tornando elementos, opções e informações visíveis, para que o usuário não precise memorizar coisas durante a interação.
- **H7. Flexibilidade e eficiência de uso:** O sistema deve ser flexível o suficiente para atender tanto a iniciantes quanto a usuários experientes. Isso pode ser alcançado com atalhos ou outras opções que acelerem o uso.
- **H8. Design estético e minimalista:** O design da interface deve ser simples e com somente informações necessárias. Menos é mais: evite elementos ou informações irrelevantes.
- **H9. Ajudar os usuários a reconhecer, diagnosticar e se recuperar de erros:** Quando um erro ocorre, o sistema deve fornecer mensagens claras e indicativas de solução para que o usuário saiba o que aconteceu e como resolver.
- **H10. Ajuda e documentação:** Embora o sistema deva ser o mais intuitivo possível, sempre que necessário, deve haver ajuda e documentação fáceis de acessar e usar, para que o usuário consiga encontrar soluções rápidas para dúvidas ou dificuldades.

Neste tipo de avaliação, especialistas percorrem a ferramenta e registram suas percepções de conformidade e/ou violação, com comentários de apoio sobre a decisão. É comum, e desejado, que a inspeção aconteça com mais de um especialista, favorecendo uma visão mais abrangente, assim como complementar dos resultados. Para tanto, em um primeiro momento, cada pessoa conduz a inspeção de forma individual e, em seguida, as classificações são comparadas, os conflitos são resolvidos, e os problemas definidos, com suas respectivas propostas de solução.

No contexto deste estudo, participaram da avaliação heurística 3 pessoas como executoras da inspeção e 1 mediadora na etapa de

resolução de conflitos (com maior nível de experiência). Todas elas possuem conhecimento prévio com as heurísticas de usabilidade e com o *VLibrasWidget*.

### 3.2 Testes com Usuários

Foi adotado o Método de Observação de Uso do sistema *VLibrasWidget* (versão em produção) com o apoio da aplicação de Questionários. Cada pessoa participante foi conduzida para a realização de um conjunto pré-definido de tarefas de teste com o uso do *VLibrasWidget*, de forma individual. As pessoas são estimuladas a expor suas interações e impressões de uso, segundo o protocolo de “verbalização de procedimentos”.

A execução das tarefas é observada por uma pessoa mediadora que toma notas sobre os acontecimentos. Esta pessoa é instruída a não ser intrusiva para não distrair nem induzir a pessoa usuária do sistema. No entanto, sendo necessário, a pessoa mediadora pode realizar esclarecimentos diante de solicitações. Para o público de pessoas surdas, a presença de uma pessoa intérprete se fez necessária, assim como a adaptação dos materiais a serem utilizados.

A observação é precedida pela aplicação de um questionário para levantamento do perfil da pessoa participante (pré-teste), e seguida pela aplicação do questionário para levantamento da satisfação subjetiva de uso (pós-teste). Se necessário, ao final da sessão de teste, algumas perguntas podem ser feitas para auxiliar a pessoa mediadora na compreensão de acontecimentos específicos ocorridos durante a observação. Foram conduzidas sessões de teste piloto para validar o procedimento e artefatos do estudo, para o público específico e genérico. Cada sessão de teste foi planejada para duração máxima de 50 minutos, incluindo:

- **Antes do início do teste:** Aplicação do questionário pré-teste (perfil do usuário) e Aplicação do Checklist de setup;
- **Ao início do teste (10’):** Acolhida da pessoa participante; Explicação sobre os objetivos do teste; Assinatura do Termo de Consentimento Livre e Esclarecido (TCLE); Configuração do ambiente (compartilhamento de tela);
- **Durante a realização das tarefas de teste (30’):** Realização das atividades de teste (roteiro de tarefas de teste) e preenchimento da ficha de observação (pela equipe de teste);
- **Após a realização das tarefas de teste (10’):** Aplicação do questionário pós-teste (satisfação subjetiva) e perguntas finais relacionadas a observações feitas durante a sessão;
- **Após a finalização do teste:** Aplicação do Checklist de setup.

A seguir são apresentadas as estruturas para os artefatos utilizados: questionário pré-teste, roteiro de tarefas de teste e questionário pós-teste. Além deste, uma ficha de observação foi elaborada, com base no roteiro de tarefas de teste, para que a pessoa mediadora pudesse fazer os registros necessários durante cada sessão de teste. Todos os artefatos estão disponíveis na íntegra para consulta, assim como na versão em Libras elaborada por uma pessoa intérprete em formato de vídeo, a partir do link.

**3.2.1 Questionário Pré-teste (Perfil).** O questionário pré-teste tem por objetivo identificar o perfil das pessoas participantes do teste, contemplando: idade, gênero, nível de escolaridade, informações sobre surdez e conhecimento em Libras e Português, assim como

familiaridade com as tecnologias envolvidas no teste. Estas informações são úteis quando correlacionadas com os demais dados coletados durante o teste.

**3.2.2 Roteiro de Tarefas de Teste.** O roteiro de tarefas de teste é um guia a ser seguido por todas as pessoas participantes no teste de forma a permitir uma análise homogênea sobre o uso do sistema. As tarefas foram estruturadas para que o *VLibrasWidget* pudesse ser explorado na íntegra, favorecendo a navegação entre diferentes telas e uso das funcionalidades disponíveis. Cada tarefa de teste foi composta por um identificador, um título, um breve descritivo e uma lista de instruções. As tarefas de teste são independentes entre si e devem ser executadas uma por vez, de forma sequencial, contemplando todas as instruções disponíveis.

O roteiro de tarefa de teste compreende: (i) informações sobre a sessão de teste; (ii) uma definição de contexto/motivação para o uso do *VLibrasWidget* e (iii) a descrição de 6 tarefas de teste com suas respectivas instruções, a saber:

- **T01: Inicialização do VLibras:** Você entrou no Portal do MEC para acessar as últimas notícias sobre educação usando o *VLibras*;
- **T02: Exploração das traduções com o VLibras:** Após assistir ao tutorial, você deve explorar a tradução das notícias utilizando o *VLibras* usando a barra de velocidade, legenda e repetição;
- **T03: Alteração do avatar do VLibras e do regionalismo utilizado na tradução:** É possível escolher um personagem dentre as opções disponíveis: Ícaro (homem), Hosana (mulher) e Guga (criança). Ainda, é possível indicar o uso dos sinais específicos de um Estado, a partir do recurso de regionalismo;
- **T04: Mudança de página, tradução por digitação de texto e envio de avaliação:** Após a exploração na tela inicial do portal, você clica em uma outra aba. Nela, a tradução pode ser realizada pela seleção de textos no site, mas também a partir da digitação de um texto fornecido por você. Em ambos os casos é possível enviar uma avaliação positiva ou negativa sobre a tradução oferecida;
- **T05: Uso do dicionário de sinais e consulta ao histórico de traduções:** É possível fazer tradução a partir do dicionário de sinais disponível no *VLibras*. As traduções realizadas podem ser consultadas no histórico, na ordem que foram feitas; e
- **T06: Alteração da posição, tamanho e opacidade do VLibras na tela:** Depois do primeiro contato com o *VLibras*, você vai poder configurar algumas características às suas preferências de uso.

**3.2.3 Ficha de Avaliação.** Este artefato consiste em uma planilha, na qual são registrados a data, o perfil da pessoa participante (P1: pessoa surda, P2: pessoa ouvinte sem experiência técnica, P3: pessoa ouvinte com experiência técnica) e o identificador da sessão de teste. Para cada tarefa de teste, são registrados os horários de início e final e para cada instrução registra-se se esta foi realizada sem dificuldade, com dificuldade ou se a instrução não foi executada pela pessoa usuária. Ainda, a tarefa recebe um estado de execução final:

- **Concluída com sucesso:** o objetivo final da tarefa foi alcançado pela pessoa usuária sem dificuldades, sem consultar ou demandar ajuda, mesmo quando algumas instruções não são executadas (não essenciais para alcançar o objetivo);
- **Concluída com falha:** o objetivo final da tarefa foi alcançado pela pessoa usuária, no entanto, para sua conclusão a pessoa cometeu algum erro (buscou pela funcionalidade explorando a ferramenta por tentativa e erro, por exemplo) ou consultou/demandou ajuda;
- **Não concluída:** o objetivo final da tarefa não foi alcançado pela pessoa usuária por (i) extrapolar o tempo médio definido para a realização da tarefa; (ii) cometer erros que impediram a execução da tarefa ou (iii) por não executar as instruções essenciais da mesma;
- **Abandonada:** quando a pessoa usuária passa para a tarefa seguinte sem executar a tarefa em questão, ou interrompe a sessão de teste.

3.2.4 *Questionário Pós-teste (Sondagem da Satisfação Subjetiva).* O questionário pós-teste tem por objetivo identificar a satisfação subjetiva declarada pela pessoa participante do teste, considerando a Usabilidade do VLibras (seção 1, com base no SUS, *System Usability Scale* [6]), a Percepção sobre a Naturalidade da Sinalização no VLibras (seção 2) e a opinião geral sobre o VLibras (seção 3). As duas primeiras seções são compostas por afirmativas para as quais espera-se uma resposta de nível de concordância em uma escala de 5 pontos. A seção final é composta por duas questões abertas sobre (i) O que eu mais gostei da minha experiência usando o VLibras foi... e (ii) O que poderia melhorar na experiência usando o VLibras é...

Quanto à **percepção sobre usabilidade**, as sentenças foram:

- Eu acho que gostaria de **usar** o VLibras com frequência.
- Eu acho o VLibras desnecessariamente **complexo**.
- Eu achei o VLibras **fácil** de usar.
- Eu acho que precisaria de **ajuda** de uma pessoa com conhecimentos técnicos para usar o VLibras.
- Eu acho que as várias **funções** do VLibras estão muito bem integradas.
- Eu acho que o VLibras apresenta muita **inconsistência**.
- Eu imagino que as pessoas **aprenderão** como usar o VLibras **rapidamente**.
- Eu achei o VLibras **confuso** de usar.
- Eu me senti **confiante** ao usar o VLibras.
- Eu **precisei aprender** várias **coisas novas** antes de conseguir usar o VLibras.

Em relação à **percepção sobre naturalidade** da sinalização, as sentenças foram:

- A **animação** (movimentação) durante a tradução realizada no VLibras foi natural.
- Eu percebi que o avatar **expressou emoções** positivas, negativas ou neutras.
- As **expressões faciais** do avatar foram adequadas durante a tradução.
- Os **movimentos corporais** do avatar foram adequados durante a tradução.

- Os **gestos corporais e faciais** estão em **sintonia temporal** e/ou aparentam **naturalidade**.
- A **amplitude** dos gestos corporais e faciais é adequada
- Eu percebi que os **gestos** corporais e faciais são **fluidos** (sem travamento).
- A **intensidade** dos gestos corporais e faciais do avatar é adequada.
- A **velocidade** dos gestos corporais e faciais do avatar é natural
- Eu percebi **pausas excessivas** durante a interação com o avatar.

## 4 VLibrasWidget: Avaliação

### 4.1 Violação das Heurísticas

A análise das 8 telas do VLibrasWidget resultou na identificação de 30 violações, distribuídas entre as telas/funcionalidades da aplicação conforme mencionado a seguir: *Tela Inicial*: 8; *Configurações*: 6; *Dicionário*: 5; *Feedback*: 5; *Ajuda*: 3. Para outras 3 (*Tradutor*, *Quem somos* e *Troca de avatar*) apenas uma heurística foi violada.

A *Tela Inicial* além de ter a maior quantidade de violações, tais inconformidades se deram para quase todas as heurísticas de usabilidade (H1, H2, H3, H4, H6, H7, H9 e H10). Esta constatação é crítica pois é o primeiro contato que se tem com a ferramenta e que, se proporcionar uma experiência frustrante ou desanimadora para o usuário, pode desencorajá-lo para continuar navegando, pois o induz a crer que as demais funcionalidades podem corresponder aos meus problemas, então destaca a importância de efetuar melhorias que possam propor uma melhor usabilidade.

Para o *Dicionário*, as violações concentraram-se nas heurísticas H2, H4, H5, H6 e H9. Enquanto nas funcionalidades relativas às Configurações as heurísticas inconformes foram: H1, H2, H3, H7, H9 e H10.

Outras telas apresentaram problemas mais pontuais. A troca de avatar violou apenas a heurística H4, o *Tradutor* a H5, e a tela *Quem somos* a H8. Já na tela de *Ajuda* foram identificadas violações das heurísticas H6, H7 e H10. A tela de *Feedback* apresentou cinco violações (H2, H3, H5, H8 e H10), que foram agrupadas e tratadas como um único problema de necessidade de refatoração da interface.

Durante a análise, foi identificado apenas um problema transversal: a violação da heurística H2, comum às telas inicial e dicionário, indicando um desafio recorrente na representação de elementos compreensíveis para o público usuário, que posteriormente foi confirmado nas sessões do teste de usabilidade.

### 4.2 Testes de Usabilidade

A realização dos Testes de Usabilidade para o VLibrasWidget aconteceu nos meses de abril e maio de 2025, nas dependências do Lavid (Laboratório de Aplicações de Vídeo Digital). Foram conduzidas 18 sessões de testes, das quais 3 foram realizadas enquanto teste piloto, com objetivo de validar o procedimento e artefatos de teste elaborados pela equipe de avaliação. Para cada perfil de participante foram realizadas 5 sessões de teste, totalizando 15 sessões válidas (5 pessoas de cada perfil). A decisão pela quantidade de participantes por perfil está alinhada com recomendações teóricas e práticas [18] na área de testes de usabilidade, com o pressuposto de que testes qualitativos com 5 participantes são suficientes para identificar cerca de

85% dos principais problemas de usabilidade. A partir de então, o aumento no quantitativo de participantes não implica na identificação de novos problemas. Ao envolver 15 participantes segmentados em 3 perfis distintos, o percentual de problemas descobertos tende a contemplar o universo de problemas mais relevantes para a versão atual da ferramenta.

Em todas as sessões de teste estiveram presentes pelo menos duas pessoas da equipe de avaliação - com experiência sobre usabilidade, interação humano computador e com conhecimento sobre o *VLibrasWidget*. Para as sessões de teste conduzidas com participantes do perfil 1, os documentos utilizados foram disponibilizados em vídeo com a sinalização realizada por uma intérprete. Ainda, durante toda a sessão de teste, uma pessoa ouvinte e intérprete em Libras atuou como mediadora na comunicação entre participante e equipe de avaliação.

**4.2.1 Perfil de Usuários Participantes.** Os participantes do teste apresentam um **perfil diverso** em termos de idade, gênero, escolaridade e familiaridade com tecnologia. A maioria, composta por jovens adultos, concentra-se na faixa etária entre 18 a 34 anos, sendo predominantemente do gênero masculino (61,1%), seguidos por 33,3% de mulheres e 5,6% que optaram por não informar. A **escolaridade** dos participantes também se mostrou variada: 77,8% possuem ensino superior incompleto, 11,1% têm ensino superior completo, e os demais distribuem-se entre ensino médio completo e pós-graduação incompleta. Quanto à **deficiência auditiva**, 66,7% não apresentam perda auditiva, 27,8% têm surdez profunda ou anacusia, e 5,6% possuem surdez severa.

No que diz respeito à **compreensão de Libras**, 33,3% não compreendem a língua de sinais, enquanto outros 33,3% declararam ter alto nível de compreensão; os restantes variam entre compreensão parcial e razoável. A **proficiência em português** também é alta, com 72,2% compreendendo muito bem o idioma, enquanto os demais relatam compreensão razoável ou limitada. Em relação à tecnologia, 66,7% se consideram muito familiarizados, porém metade dos participantes se declarou nada familiarizada com o *VLibrasWidget* e o *VLibras mobile*.

No grupo dos participantes de perfil específico (P1) houve predominância masculina (80%) e faixa etária concentrada entre 25 e 34 anos (80%), com apenas 20% na faixa de 35 a 44 anos. A escolaridade do grupo mostra diversidade, com a maioria (60%) tendo ensino superior incompleto, seguido por 20% com ensino superior completo e 20% com pós-graduação incompleta. Todos os participantes deste perfil possuem surdez profunda ou anacusia, e demonstram um domínio elevado da Libras, com 100% indicando compreensão total. E, apesar da fluência em Libras, a compreensão do português apresenta uma variação: 60% compreendem razoavelmente, 20% pouco e apenas 20% muito bem. A familiaridade com tecnologia é significativa, com 60% relatando muita familiaridade, enquanto os demais se dividem entre razoável (20%) e pouca (20%). No entanto, o conhecimento sobre as ferramentas do *VLibras* é limitado, com 80% relatando pouca familiaridade com o *widget* e 20% razoável; já no *mobile*, a tendência se inverte, com 80% indicando razoável familiaridade e 20% pouca. Esses dados sugerem a necessidade maior de divulgação do *VLibrasWidget* entre o público surdo.

No conjunto dos participantes de perfil genérico e ouvinte (P2), houve distribuição equilibrada de gênero, com 40% de mulheres,

40% de homens e 20% que preferiram não informar. A maioria se concentra na faixa etária de 18 a 24 anos (60%), e o restante entre 25 e 34 anos. No que se refere à escolaridade, 80% têm ensino superior incompleto e 20% superior completo, e todos não possuem perda auditiva. A compreensão de Libras é limitada, com 40% compreendendo pouco, 40% razoavelmente e 20% sem compreensão, mas todos os participantes dominam muito bem o português. Em relação à tecnologia, a maioria (60%) se considera razoavelmente familiarizada, enquanto 20% indicam pouca familiaridade e outros 20% muita familiaridade. O nível de conhecimento sobre o *VLibras* é restrito: 80% relataram não ter familiaridade com o *widget* e 20% pouca familiaridade; com o *mobile*, 60% declararam não ter familiaridade e 40% relataram pouca.

Já entre os participantes de perfil genérico com conhecimento técnico (P3), o perfil é predominantemente jovem (100% entre 18 e 25 anos), com 60% do gênero masculino e 40% do feminino. A escolaridade se concentra em níveis mais elevados, com 80% dos participantes com ensino superior incompleto e 20% já com superior completo. Nenhum dos participantes apresenta perda auditiva, e a compreensão de Libras é limitada: 60% não compreendem a língua e 40% compreendem pouco. Todos têm uma compreensão elevada do português. Neste perfil, destaca-se a alta familiaridade com tecnologia, com 100% se considerando muito familiarizados. Entretanto, a familiaridade com as ferramentas do *VLibras* é reduzida: 60% não estão familiarizados com o *widget* e 40% têm pouca familiaridade; para o *mobile*, 80% indicam não ter familiaridade e 20% se consideram razoavelmente familiarizados.

**4.2.2 Tarefas de Teste.** Para o contexto do *VLibrasWidget* o tempo de realização das tarefas de teste depende, significativamente, das sentenças selecionadas para tradução. Portanto, não é adequado fazer uma análise comparativa entre os tempos previstos e reais da execução das tarefas devido a esta variabilidade. Sendo assim, para esta métrica, optou-se por registrar apenas a duração das sessões de teste, em minutos, para fins de comparação entre os três perfis de usuários participantes do teste (Tabela 1).

**Table 1: Tempo de realização das sessões e teste por perfil**

Sessão	P1	P2	P3
S1	25	17	16
S2	26	27	16
S3	27	28	17
S4	20	29	16
S5	35	18	18
Média	26	23	16

Vale destacar que: (i) o tempo médio dos usuários P1 é ligeiramente superior aos demais participantes pela necessidade de mediação com intérprete de Libras na comunicação e pelo consumo dos artefatos adaptados (tempo de sinalização superior ao tempo de leitura); a S5 para o usuário P1 foi a sessão mais longa, pois a pessoa participante foi muito expressiva sobre suas impressões no uso do sistema durante a execução dos testes, fornecendo muito

*feedback*; e as sessões para P3 foram as de menor duração, o que foi associado a dois fatores, familiaridade com tecnologia e interesse reduzido pela tradução em si, por serem ouvintes.

Quanto aos estados de realização das tarefas, cada tarefa pode assumir 4 estados durante a sessão de teste: Concluída com sucesso (CS), Concluída com falha (CF), Não concluída (NC) ou Abandonada (AB). Na Tabela 2 é possível consultar, por tarefa de teste (T01 a T06) e perfil de usuário (P1, P2 e P3) o quantitativo de ocorrências relativas ao estado de realização da tarefa.

Percebe-se que nenhum participante, independente do perfil, abandonou qualquer tarefa. Assim como que, para o P3, todas as pessoas participantes concluíram todas as tarefas de teste. Do exposto, é possível inferir que, mesmo que o VLibrasWidget possa impor algumas barreiras no uso da aplicação, estas não são impeditivas para realização das tarefas de teste, ou desmotivantes para seu abandono.

Ainda, é possível perceber uma similaridade no estado de finalização das tarefas para os Perfis 1 e 2, pessoas surdas e pessoas sem conhecimento prévio com o VLibrasWidget. Para todas as tarefas de teste, tem-se o mesmo quantitativo de ocorrências para cada estado de finalização da tarefa. Esta informação sugere que o VLibrasWidget não impõe, necessariamente, barreiras distintas para pessoas usuárias que sejam ou não surdas.

Duas tarefas foram concluídas com sucesso para 4 dos 5 participantes: T01 (inicialização do VLibras) para P1 e P2 e T04 (Mudança de página, tradução por digitação de texto e envio de avaliação) para P3. Estas tarefas são essenciais para o uso da aplicação. Em contrapartida, as tarefas T03 (Alteração do avatar do VLibras e do regionalismo utilizado na tradução) e T06 (Alteração da posição, tamanho e opacidade do VLibras na tela) impuseram mais dificuldades para os participantes do teste, P1 e P2; P3 respectivamente. As funcionalidades exploradas nessas tarefas são, essencialmente, de configuração da aplicação.

De uma forma geral, a maioria das tarefas (25/30) foram concluídas para os perfis P1 e P2 e todas (30) para o perfil 3, levando à reflexão sobre a existência de barreiras para o uso do VLibrasWidget independente do perfil de usuários, no entanto, tais barreiras não são obstruções severas.

**Table 2: Estado de realização das tarefas de teste por perfil**

	P1			P2			P3		
	CS	CF	NC	CS	CF	NC	CS	CF	NC
<b>T01</b>	4	1		4	1		3	2	
<b>T02</b>	2	3		2	3		2	3	
<b>T03</b>		4	1		4	1	3	2	
<b>T04</b>	3	1	1	3	1	1	4	1	
<b>T05</b>		3	2		3	2	2	3	
<b>T06</b>	2	2	1	2	2	1	1	4	
<b>Total</b>	11	14	5	11	14	5	15	15	0

As tarefas de teste agrupam funcionalidades do VLibrasWidget em subtarefas. Durante a observação da interação e dos dados

coletados, foi possível identificar que algumas funcionalidades se destacaram por imporem barreiras significativas à interação dos usuários. Essas barreiras se manifestaram tanto por dificuldades de compreensão sobre o significado de ícones, ou fluxo de navegação, quanto por problemas de usabilidade, interferindo na experiência de uso. A seguir, são destacadas as funcionalidades a partir das quais foi possível identificar dificuldades enfrentadas pelos usuários, nas respectivas tarefas:

**T01:** As ocorrências de falha concentraram-se na subtarefa “**Assista ao tutorial**”, que foi ignorada (não executada) por alguns participantes.

**T02:** As falhas dessa tarefa concentraram-se na subtarefa “**Escolha um texto para traduzir**” Durante a seleção de textos para tradução observou-se que muitos participantes não perceberam que o cursor do mouse permite a tradução por meio de cliques. Como alternativa, alguns recorreram ao método de copiar e colar o conteúdo no Tradutor, enquanto outros tentaram selecionar imagens diretamente.

**T03:** Essa tarefa também explorava a funcionalidade de **tradução por escolha de texto**, e foi comum que as falhas da tarefa anterior fossem reproduzidas de forma repetida. Porém, a funcionalidade que mais impôs obstáculos foi a “**Escolha uma imagem para traduzir**”, sendo comum que os usuários tivessem dúvida, fizessem perguntas, voltassem ao tutorial ou simplesmente não executassem a subtarefa. Algumas tentativas de colar imagens no Tradutor foram observadas, mas sem êxito. Ainda, alguns usuários demonstraram dificuldade para **encontrar e alterar o avatar**, assim como, **ativar o regionalismo**.

**T04:** Na subtarefa “**Avalie a tradução como positiva ou negativa**”. Alguns participantes deram *feedback* sem uso da ferramenta (direcionando a opinião aos mediadores), enquanto outros tiveram dúvida e perguntaram onde e como fazê-lo.

**T05:** Na subtarefa “**Selecione o dicionário do VLibras**”, diversos participantes relataram dificuldades para localizar o dicionário, apontando que o ícone associado à funcionalidade não remete claramente à sua finalidade.

**T06:** Por fim, as **subtarefas de configurações** “Mude a posição do VLibras na tela” e “Altere a opacidade para que possa ver o texto através do VLibras” causaram muitas dúvidas entre os participantes. A principal fonte de confusão foi o reposicionamento da aplicação na tela, pois muitos usuários presumiam que a ação de modificar a localização do *widget* poderia ser realizada por meio de arraste, o que não é suportado pela interface. Em relação à opacidade, alguns demonstraram confusão quanto a compreensão de como ela funciona, uma vez que a alteração só pode ser visualizada após sair da tela de configurações.

**4.2.3 Sondagem da Satisfação Subjetiva.** O questionário pós-teste abrangeu três aspectos principais: usabilidade do VLibrasWidget, naturalidade da sinalização e opiniões gerais sobre a ferramenta. As respostas foram analisadas para cada perfil de participante e estão descritas a seguir.

Quanto aos aspectos de usabilidade, o SUS permite avaliar, a partir de dez itens, aspectos como a facilidade de uso, a confiança no sistema e a percepção geral dos participantes sobre sua utilidade. A pontuação obtida a partir de sua aplicação pode assumir valores entre 0 e 100, onde pontuações **abaixo de 68 indicam problemas**

**de usabilidade.** As médias SUS obtidas por perfil foram: **P1 = 54; P2 = 84 e P3 = 91.**

A partir das pontuações obtidas é possível identificar que o público alvo do *VLibrasWidget* (P1) corresponde àquele que mais declara, subjetivamente, enfrentar problemas de usabilidade no uso da aplicação. Esta realidade pode ser consequência da necessidade de melhorar aspectos de acessibilidade na aplicação, assim como pelo fato desse perfil ser mais crítico por serem usuários que utilizam Libras em seu dia a dia, percebendo nuances específicas. Os demais perfis obtiveram pontuações que indicam satisfação com a usabilidade do *VLibrasWidget*, no entanto, apresentando melhor percepção de usabilidade para usuários experientes no uso de tecnologia.

No que se refere aos aspectos de naturalidade da sinalização no *VLibrasWidget*, a avaliação foi realizada com base na percepção dos usuários sobre a animação, a expressão facial e a sincronia dos gestos corporais e faciais dos personagens (Ícaro, Hosana e Guga). Das 10 afirmativas colocadas, 9 eram positivas e uma negativa. Neste sentido, o cálculo para pontuação tomou como base o valor médio por participante, com uma análise dos valores mínimo e máximo por perfil. As pontuações obtidas indicam, em consonância com a percepção sobre a usabilidade, menor satisfação sobre a naturalidade da sinalização por parte de usuários surdos (P1) com menor média mínima de 2,4 e máxima de 3,6. Em sequência, P2 com médias 3 e 4,4, e P3 com médias 4 e 4,8, mínimas e máximas, respectivamente.

É importante destacar que, o público ouvinte (P2 e P3) avaliou de forma mais negativa aspectos relacionados à fluidez nos movimentos e expressões. Enquanto que, as pessoas surdas indicaram a necessidade de melhorar a naturalidade sobre a fluidez e ritmo das animações, assim como a sincronização dos gestos corporais e faciais, indicando relevante, ainda, as expressões faciais. Os elementos apontados são essenciais na compreensão de Libras e apenas este perfil de participantes tem o conhecimento necessário para avaliá-las. De uma forma geral, os aspectos de naturalidade da sinalização que foram bem avaliados foram: movimentos corporais, intensidade e amplitude dos gestos adequados. Em contrapartida, existe uma expectativa de melhoria para a naturalidade da animação de uma forma geral, com destaque para representação das emoções primárias.

Após essa avaliação, as pessoas participantes poderiam fornecer suas opiniões opcionalmente. As respostas abertas foram agrupadas, por perfil, em: pontos positivos e sugestões de melhorias. Exemplos de *feedback* também são fornecidos a seguir.

**P1 - Aspectos mais apreciados:** a facilidade de uso e a contribuição para a inclusão digital; assim como a função de velocidade ajustável para a sinalização da animação; **Sugestões de melhoria:** naturalidade da animação dos personagens, com ajustes no ritmo e velocidade, além de melhorias na expressão facial.

- "A velocidade 1,5x foi mais fluida que a padrão. Alguns sinais precisam melhorar, mas a ferramenta é muito útil para a comunidade surda."

**P2 - Aspectos mais apreciados:** facilidade de uso e a integração com o conteúdo da página; **Sugestões de melhoria:** o ícone para a funcionalidade "Dicionário" é pouco intuitivo, e a função de configurar opacidade deve ser ajustada em tempo real.

- "O dicionário é útil, mas o ícone não é intuitivo. Gostei muito da função de velocidade, mas a tradução de textos longos apresentou algumas falhas."

**P3 - Aspectos mais apreciados:** funções de ativar "Regionalismo" e "Legenda", assim como acessar o "Dicionário"; Sugestões de melhoria: na forma de apresentação e seleção para a velocidade da tradução (lista de todas as opções disponíveis) e maior consistência nas expressões faciais.

- "A ferramenta é prática e fácil de usar. No entanto, a tradução de termos técnicos e a velocidade do avatar precisam de ajustes."

É interessante perceber que, por mais que a percepção de satisfação subjetiva (SUS) tenha sido a menor para o P1, o que indica uma "insatisfação" com a usabilidade do *VLibrasWidget*; e que as pessoas participantes desse perfil tenham executado tarefas que foram concluídas com falhas, em declaração livre e opcional, indicaram facilidade de uso e reforçaram a importância social da aplicação no tocante à inclusão digital.

## 5 Resultados da Avaliação

As informações obtidas a partir da inspeção heurística e dos testes de usabilidade foram analisadas e organizadas enquanto *issues* e registradas no repositório do projeto *VLibras*, o GitLab. Esta estratégia permite a rastreabilidade dos problemas encontrados em seu ciclo de correção e verificação, possibilitando acompanhar a evolução da aplicação. As descrições das *issues* são padronizadas, contendo: identificador único, título, descrição, passo a passo para reprodução do problema, solução de mudança proposta, tela/funcionalidade alvo e severidade da *issue*. É possível acrescentar mídias de imagem e áudio para dar suporte ao registro, quando necessário. Caso a *issue* se aplique a mais de uma ferramenta da Suíte do *VLibras*, uma *tag* identificadora é utilizada. Em alguns casos, as *tags* também são utilizadas para indicar se a *issue* é referente a um *bug* ou a um *upgrade*.

A severidade de uma *issue* está relacionada com o impacto que o problema pode impor ao uso do sistema, podendo ser classificada, segundo Nielsen [16], em: **S1 - Alta** - Catástrofe de usabilidade impedido o uso: imperativo consertar isso antes que o produto possa ser lançado; **S2 - Média** - Problema maior de usabilidade, impondo obstruções importantes de uso: importante consertar, por isso deve receber alta prioridade; **S3 - Baixa** - Problema menor de usabilidade, que podem gerar frustração, mas não são barreiras para o uso da aplicação, com baixa prioridade para correção e **S4 - Neutra** - apenas estético ou sugestões de melhoria, cuja implantação deve ser realizada apenas quando existe tempo extra disponível no projeto.

Na abordagem de avaliação multidimensional, as *issues* oriundas das heurísticas e dos testes de usabilidade são complementares. Ainda, considerando que o *VLibrasWidget* e o *VLibrasMóvel* fazem parte da mesma suíte de aplicações, e que as interfaces e interações sejam semelhantes, em análise posterior, foi possível mapear algumas *issues* identificadas para a versão *widget* à aplicação móvel - ampliando o impacto dos resultados para além do planejado inicialmente.

A contribuição da avaliação dos participantes surdos orientou a priorização de ajustes e melhorias, garantindo que a experiência do público principal do *VLibras* esteja no centro das decisões e ações



promovidas a partir dos resultados do estudo. As *issues* documentadas foram encaminhadas para a equipe de desenvolvimento da ferramenta, e as correções devem ser realizadas e disponibilizadas nas próximas atualizações desta, assegurando que o retorno dos usuários seja efetivamente incorporado ao ciclo de melhorias contínua no *VLibras*.

### 5.1 Issues da Avaliação Heurística

A análise das heurísticas violadas com inspeção permitiu a definição de um conjunto de 23 *issues*.

No que se refere a severidade das *issues*, o quantitativo por nível de severidade obtido foi: **S1 Alta = 1; S2 - Média = 5; S3 - Baixa = 7 e S4 Neutra = 9**. Segue a lista das *issues* reportadas para os níveis de Alta, Média e Baixa.

**Severidade Alta [S1]:** #94: Nome técnico das palavras no dicionário pode torná-las pouco compreensíveis.

**Severidade Média [S2]:** #89: Informação do texto escolhido para tradução não é clara; #93: Necessidade de abrir o *widget* novamente em cada mudança de página dentro de um mesmo site; #97: Botão de velocidade de tradução é pouco intuitivo em relação às velocidades disponíveis; #101: Estabelecer limites para a opacidade; #106: O sistema não possui nenhum tipo de ajuda, além do tutorial.

**Severidade Baixa [S3]:** #87: Existem itens no dicionário sem sinal associado; #63: A aba recentes no dicionário registra apenas os sinais tocados, e não as palavras cujos sinais foram procurados; #100: A opacidade não é ajustada em tempo real; #102: Implementar uma pré visualização de opacidade e restaurar para os valores padrões; #103: Modificação da posição do *widget* na tela não é intuitiva; #104: As configurações não oferecem ajuda contextual ou explicações sobre o que cada configuração faz, e #108: Mover o regionalismo para o dicionário.

No que se refere às telas/funcionalidades implicadas, 9 *issues* estão associadas à tela inicial do *VLibrasWidget* (#88, #89, #90, #91, #92, #93, #95, #96, #97), indicando a necessidade de melhoria para apresentar ícones mais representativos, uma navegação mais fluida, apresentar ao usuário facilidade de identificar o que cada funcionalidade faz de forma intuitiva e ter uma linguagem de entendimento compreensível. Sete *issues* referem-se às possibilidades de configurações disponíveis na aplicação (#99, #100, #101, #102, #104, #108). Quatro estão relacionadas ao Dicionário (#87, #88, #94, #98), sobretudo para que a linguagem textual adotada seja compreensível por um público mais abrangente - favorecendo o uso da aplicação como instrumento de aprendizado para pessoas ouvintes com interesse em Libras, por exemplo. Duas *issues* foram referentes a funcionalidade de *Feedback* (#110, #111) e as demais tiveram uma ocorrência cada para: troca de avatar (#105), uma para a tela de *Quem Somos* (#107) e, por fim, uma para a *Ajuda* (#106).

### 5.2 Issues dos Testes de Usabilidade

Com a análise dos dados observados durante a interação das pessoas participantes com o *VLibrasWidget*, assim das informações declaradas por elas, um conjunto de 9 *issues* foram identificadas, das quais, 4 haviam sido previamente registradas como resultado da avaliação heurística (#89, #97, #99, #108).

Em relação ao nível de severidade das *issues* identificadas exclusivamente a partir da perspectiva dos usuários têm-se: S2 - Média = 2; S3 - Baixa = 3, descritas a seguir.

**Severidade Média [S2]:** #111 [Upgrade] Barra de execução da tradução poderia ser manipulada; #112 [Upgrade] Símbolo do *VLibrasWidget* pode ser mais intuitivo.

**Severidade Baixa [S3]:** #117 [Upgrade] Recurso de legenda pode ser ativado como padrão; #115 [Bug] Palavras com sinalização incorreta; #116 [Upgrade] Tradução na velocidade 3.0 pode ser alterada.

Das das 9 *issues* identificadas, 7 foram percebidas enquanto as pessoas do P1 (surdos) interagiam com o *VLibrasWidget*, sendo 4 delas referentes a sinalização e a tradução (#111, #115, #116, #117) - essenciais no processo de comunicação em Libras. Este fato é consequência da familiaridade das pessoas participantes com a Libras. Adicionalmente, destas *issues*, 3 foram identificadas exclusivamente com o referido perfil (#115, #116, #117) - destacando a relevância de considerar os usuários reais do sistema em contexto de avaliação com foco na melhoria da experiência de uso.

### 5.3 Issues Mapeadas para o VLibrasMóvel

Das *issues* identificadas com a avaliação do *VLibrasWidget*, 4 se aplicam também para a ferramenta *VLibrasMóvel*, são elas: S1 - Alta #94 [Upgrade] Nome técnico das palavras no dicionário pode torná-las pouco compreensíveis; S3 - Baixa: #87 [Bug] Itens no dicionário sem sinal associado, e #116 [Upgrade] Tradução na velocidade 3.0 pode ser alterada; e S4 - Neutra: #90 [Upgrade] Alteração no ícone de ajuda para melhor entendimento.

Este mapeamento foi um alerta para a necessidade de realizar uma avaliação de inspeção heurística específica para a ferramenta *VLibrasMóvel*, tanto para a versão Android, quanto para iOS.

## 6 Lições Aprendidas

A inspeção heurística realizada previamente foi importante como base para o planejamento dos testes com os usuários - na definição das tarefas de teste - diante da possibilidade de evidenciar e entender como a violação das heurísticas geram (ou não) impactos reais na interação entre usuário e sistema. A análise estática da interface é articulada à observação dinâmica do uso em contexto específico, sob uma perspectiva complementar e comparativa.

A participação de diferentes perfis de usuários nos testes de usabilidade proporcionou aprendizados relacionados tanto ao planejamento, quanto à condução da avaliação. Identificou-se a necessidade de adaptar os artefatos utilizados (roteiro de tarefas, formulários, termo de consentimento) para a Libras, de forma a garantir uma comunicação acessível e efetiva. Ainda, a participação da intérprete de Libras antes e durante as sessões de teste se mostrou essencial não apenas para mediar a comunicação, mas também para assegurar que os participantes surdos se sentissem confortáveis e compreendidos durante o experimento. Adicionalmente, ter perfis diversificados possibilitou a identificação de nuances distintas de problemas de interação, o que contribuiu para uma compreensão mais abrangente das necessidades e para uma evolução mais inclusiva e eficaz da aplicação.

A inclusão de questões abertas no questionário pós-teste, ainda que opcionais, foi outro ponto que revelou-se uma estratégia útil

para captar percepções espontâneas dos participantes. Esses comentários livres permitiram que os usuários expressassem aspectos da experiência que não seriam obtidos por escalas padronizadas ou mesmo pela observação da interação. As respostas abertas fornecidas possibilitaram o cruzamento dos dados observados com as questões fechadas do questionário.

A experiência de adotar uma avaliação multidimensional permitiu uma visão complementar sobre os achados, demonstrando-se eficazes ao possibilitar maior abrangência dos resultados. A análise das interações entre usuários e sistema permitiu identificar barreiras específicas no uso do *widget*, cujas causas puderam ser associadas às heurísticas de usabilidade violadas. Algumas dessas questões revelaram-se recorrentes e puderam ser espelhadas para outras plataformas da suíte *VLibras*. Teve-se, portanto, a oportunidade de indicar padrões de dificuldade que podem ser comuns entre as ferramentas web e móvel, favorecendo uma melhoria em escala mais ampla. Esse fator reforça a importância de abordagens centradas no usuário e da consideração das especificidades da comunidade surda no projeto e avaliação de tecnologias acessíveis.

Essas lições sublinham a importância de se adotar uma abordagem centrada no usuário e continuamente iterativa para garantir a melhoria evolutiva da acessibilidade em soluções de tradução de Português para Libras. Em especial, quando se trata de públicos com demandas específicas, torna-se evidente que a acessibilidade não se limita à disponibilização de recursos tecnológicos, mas envolve também a adaptação de processos, artefatos e interações. Portanto, esse aprendizado reforça que garantir acessibilidade é um processo contínuo, contextual e colaborativo, que se fortalece a partir de ciclos sucessivos de análise e redesign, com participação ativa dos próprios usuários.

## 7 Conclusões e Trabalhos Futuros

A realização da avaliação heurística e dos testes com usuários se mostrou valiosa para identificar problemas de usabilidade que interferem na interação com o *VLibrasWidget*, podendo afetar a experiência de uso.

A violação das heurísticas, juntamente com as observações de uso e declarações das pessoas participantes nos testes de usabilidade, foram compiladas em *issues* associados às diferentes características da interface e de interação (representação visual de ícones, navegação, arquitetura da informação, tempo de resposta, entre outros), telas (*Configurações*, *Dicionário*, *Feedback*, *Ajuda*) e funcionalidades (*Feedback*, opacidade, manipulação do *widget* na tela, entre outros) do *VLibrasWidget*.

As 28 *issues* foram registradas com informações que permitem sua reprodutibilidade e solução proposta. A indicação da severidade de cada *issue* é uma informação útil para a priorização das mesmas no *backlog*, sendo seu gerenciamento apoiado pelo uso do GitLab na rastreabilidade de suas correções no ciclo de vida do projeto *VLibras*. Em se tratando da Suíte *VLibras*, algumas *issues* puderam ser mapeadas para outra ferramenta, o *VLibrasMóvel*, sendo este um ganho colateral ao planejamento inicial.

Neste sentido, a equipe do projeto passa a ter as informações necessárias para correção de inconsistências e problemas em elementos da interface e de interação; a partir de recursos concretos para priorizar a evolução da aplicação; reduzindo riscos de manutenibilidade e retrabalhos.

A aplicação de uma avaliação multidimensional permitiu alcançar resultados complementares, e portanto, mais abrangentes. Outro aspecto que favoreceu a identificação de *issues* a partir de óticas distintas foi considerar, além da visão de especialistas, as características de três perfis de usuários (reais e potenciais) do *VLibrasWidget*: pessoa surda, ouvinte iniciante com a aplicação e ouvinte com experiência em tecnologia.

Diante do exposto, tem-se como expectativa que, diante da correção das inconformidades percebidas e implantação das sugestões de melhoria, a evolução do *VLibrasWidget* e *VLibrasMóvel* aconteça de forma alinhada com as necessidades de seus usuários. Como consequência, favorecendo a experiência de uso por prover: maior facilidade de navegação; melhor compreensibilidade sobre funcionalidades do sistema; linguagem e elementos do design condizente com a realidade dos usuários; etc.

A fim de garantir o compromisso em prover acessibilidade e inclusão digital a partir da Suíte *VLibras*, se faz necessário que as melhorias que foram propostas, sigam um ciclo contínuo de revisão e ajuste baseado no conhecimento de especialistas, no *feedback* de usuários, mantendo um registro organizado das *issues*, monitorando da evolução das aplicações.

Há planejamento de novos ciclos de avaliação para validar e ampliar os achados apresentados. Pois, a condução de testes de usabilidade está prevista de forma sistemática no cronograma de execução do projeto. Quando as correções estiverem finalizadas, assim como quando novas funcionalidades forem disponibilizadas, outros testes de usabilidade devem ser conduzidos. Esta estratégia permite o refinamento contínuo do *VLibrasWidget*, considerando a perspectiva de especialistas e usuários.

Como trabalho futuro, no contexto de evolução da Suíte *VLibras*, a naturalização da sinalização é entendida como prioridade, sobretudo com a inclusão da representação das emoções a partir de expressões faciais e corporais. Diante dos resultados obtidos com a atual avaliação multidimensional do *VLibrasWidget*, são previstos outros ciclos semelhantes para esta e outras ferramentas.

## ACKNOWLEDGEMENTS

Agradecemos à Secretaria Nacional dos Direitos da Pessoa com Deficiência (SNDPD) do Ministério dos Direitos Humanos e da Cidadania (MDHC) e à Secretaria de Governo Digital (SGD) do Ministério da Gestão e da Inovação em Serviços Públicos (MGI) pelo apoio para realização dessa pesquisa. Agradecemos também às pessoas envolvidas no projeto *VLibras* no âmbito do Laboratório de Aplicações de Vídeo Digital (LAViD) da Universidade Federal da Paraíba (UFPB), e a todos que participaram desse estudo - usuários e intérprete.

## REFERENCES

- [1] Acesso à Informação Pública: uma introdução à Lei nº 12.527, de 18 de novembro de 2011. Brasil, Controladoria-Geral da União.
- [2] A. C. Aguiar, L. E. Melo Alves, Y. P. Costa Aguiar P. Lima, T. Maritan, and D. Omai. 2024. Challenges and Lessons Learned to Improve the Deaf User Experience: Compliance with ABNT NBR 17060 Standard in Mobile Applications. *Proceedings*

- of the XXIII Brazilian Symposium on Human Factors in Computing Systems (IHC 2024) (2024). doi:10.1145/3702038.3702041
- [3] Associação Brasileira de Normas Técnicas 2022. ABNT NBR 17060:2022 - Acessibilidade em comunicação na prestação de serviços.
  - [4] Simone Barbosa, Bruno Santana da Silva, Milene Silveira, Isabela Gasparini, Ticianne Darin, and Gabriel Barbosa. 2024. *Interação Humano-Computador e Experiência do Usuário*. Leanpub.
  - [5] E. Benveniste. 1976. *Problemas de Linguística Geral*. Ed. da Universidade de São Paulo, Chapter vol. 8.
  - [6] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry* (1995).
  - [7] J. P. Colling and C. Boscarioli. [n. d.]. Avaliação de Tecnologias de tradução Português-Libras visando o uso no ensino de crianças surdas.
  - [8] C. H. d. A. Monteiro, L. F. I. L. Pecoraro, and A. R. Corbo et al. 2012. Utilização de novas tecnologias na popularização da Língua Brasileira de Sinais.
  - [9] J. E. R. de Queiroz. 2001. *Abordagem híbrida para a avaliação da usabilidade de interfaces com o usuário*. Ph. D. Dissertation. Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, Campina Grande.
  - [10] J. E. R. de Queiroz and D. de Sousa Ferreira. 2009. A Multidimensional Approach for the Evaluation of Mobile Application User Interfaces. *Human-Computer Interaction. New Trends (HCI 2009)* (2009).
  - [11] F. M. F. de Sá, L. E. M. Alves, A. C. Aguiar, and Y. P. Costa Aguiar. 2024. Theoretical test using Libras to obtain the National Driver's License: Exploratory usability assessment with interpreters and deaf people. *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems (IHC 2024)* (2024). doi:10.1145/3702038.3702051
  - [12] Instituto Brasileiro de Geografia e Estatística 2024. Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua.
  - [13] Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência) 2015. Lei nº 13.146, de 6 de julho de 2015.
  - [14] Manuella A.C.B. Lima, Tiago M.U. de Araújo, and Erickson S. de Oliveira. 2015. Incorporation of Syntactic-Semantic Aspects in a LIBRAS Machine Translation Service to Multimedia Platforms. *Proceedings of the 21st Brazilian Symposium on Multimedia and the Web (WebMedia '15)* (2015). doi:10.1145/2820426.2820434
  - [15] Jakob Nielsen. 1994. 10 Usability Heuristics for User Interface Design. *Nielsen Norman Group* (1994).
  - [16] Jakob Nielsen. 1994. How to Rate the Severity of Usability Problems. *Nielsen Norman Group* (1994).
  - [17] Jakob Nielsen. 1994. Usability inspection methods. *Conference Companion on Human Factors in Computing Systems (CHI '94 Companion)* (1994). doi:10.1145/259963.260531
  - [18] Jakob Nielsen. 2000. Why You Only Need to Test with 5 Users. *Nielsen Norman Group* (2000).
  - [19] Erik Da Costa Nunes, Guilherme Ribeiro, Ingrid Monteiro, and Enyo Gonçalves. 2023. Digital Accessibility at the Brazilian Symposium on Human Factors in Computing Systems (IHC): An Updated Systematic Literature Review. *Proceedings of the 22nd Brazilian Symposium on Human Factors in Computing Systems (IHC '23)* (2023). doi:10.1145/3638067.3638075
  - [20] L. S. Reis, T. M. U. Araujo, Y. P. Costa Aguiar, and M. A. C. B. Lima. 2020. A Methodology for Assessing Translation Quality Generated by Sign Language Machine Translators: A Case Study with VLibras. In *Lecture Notes in Computer Science*, Vol. 12189, Springer International Publishing, 132–146 (2020). doi:10.1007/978-3-030-49108-6\_10
  - [21] L. S. Reis, T. M. U. Araujo, Y. P. C. Aguiar, and M. A. C. B. Lima. 2023. A Multilevel Methodology to Assess the Quality of Machine Translators to Brazilian Sign Languages. *XI Latin American Conference on Human Computer Interaction (CLIHC 2023)* (2023). doi:10.1145/3630970.3631050
  - [22] L. S. Reis, T. M. U. Araujo, Y. P. C. Aguiar, M. A. C. B. Lima, and A. S. S. Sales. 2018. Assessment of the Treatment of Grammatical Aspects of Machine Translators to Libras. *Anais do XXIV Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres, Salvador, BA, Brazil. Sociedade Brasileira de Computação (SBC)* (2018).
  - [23] L. S. Reis, A. S. S. Sales, M. D. F. C. Lima, T. M. U. Araujo, and Y. P. C. Aguiar. 2017. Avaliação de Usabilidade do Aplicativo VLibras-Móvel com Usuários Surdos. *Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres, Gramado, Brazil. Sociedade Brasileira de Computação (SBC)* (2017).
  - [24] Yvonne Rogers, Helen Sharp, and Jennifer Preece. 2013. *Design de interação*. Bookman Editora.
  - [25] Maristela C Vieira, Ygor Corrêa, Lucila MC Santarosa, and MCV BIAZUS. 2014. Análise de expressões não-manuais em avatares tradutores de Língua Portuguesa para Libras. In *XIX Conferência Internacional sobre Informática na Educação-TISE.J.* (2014).
  - [26] World Health Organization 2021. World Report on Hearing. World Health Organization, Geneva.