

Avaliação de Usabilidade do SofiaFala: Melhorando a Tecnologia Assistiva para Terapias de Fala

Victor Hugo da Silva Lembor

victorlembor@usp.br

Ribeirão Preto Medical School, University of São Paulo

ABSTRACT

This paper presents a multifaceted usability evaluation protocol for SofiaFala-Fono, an assistive technology for speech therapists. Unlike traditional unimodal approaches, the protocol combines heuristic evaluation by HCI experts, usability testing with therapists, and affective response measurement (SAM questionnaire). This methodological triangulation offers a comprehensive assessment of effectiveness, efficiency, and user satisfaction, providing a replicable framework to support the development of more user-centered assistive technologies in speech therapy.

KEYWORDS

heuristic evaluation, usability testing, self-assessment manikin, human-computer interaction, assistive technology, speech therapy

1 INTRODUÇÃO

Transtornos de fala representam um desafio significativo para o desenvolvimento de crianças e exercício da cidadania. Nesse cenário, as tecnologias assistivas (TAs) emergem como ferramentas cruciais, com o potencial de estender a prática terapêutica para além do ambiente clínico. Ao apoiar o tratamento contínuo, as TAs podem não apenas aprimorar a comunicação, mas também promover o desenvolvimento socioemocional e prevenir o isolamento social. Sistemas dedicados à prática da pronúncia, um subconjunto dessas tecnologias permitem que usuários realizem exercícios de forma autônoma para melhorar sua inteligibilidade [8]. O projeto SofiaFala visa desenvolver TAs para aprimorar a qualidade de vida de pessoas com dificuldades de fala [8, 9, 15, 17]. Nesse ecossistema, foram criados dois aplicativos complementares: SofiaFala-Paciente e SofiaFala-Fono. O primeiro, destinado ao paciente, captura dados de áudio e vídeo durante a execução de exercícios terapêuticos, fornecendo um retorno lúdico-educacional ao usuário e gerando relatórios de desempenho para o fonoaudiólogo. O SofiaFala-Fono, por sua vez, é uma plataforma para o profissional, permitindo que este prepare os exercícios e analise o progresso clínico de seus pacientes com base nos dados recebidos do aplicativo paciente. Para que ferramentas digitais como estas alcancem seu pleno potencial clínico, é imperativo garantir sua usabilidade. A avaliação sistemática de uma *interface* é crucial para assegurar que ela atenda às necessidades de seus usuários e promova uma experiência de uso eficaz e satisfatória [10]. Este estudo foca na Avaliação de Usabilidade (AU) do SofiaFala-Fono, buscando identificar quais são os problemas e as

Alessandra Alaniz Macedo (Advisor)

ale.alaniz@usp.br

Department of Computing and Mathematics, FFCLRP, USP

violações de usabilidade¹.

2 TRABALHOS RELACIONADOS

A digitalização da prática clínica consolidou os aplicativos móveis como ferramentas de apoio na Fonoaudiologia. Embora essa tecnologia ofereça oportunidades para otimizar a terapia e o engajamento do paciente, seu potencial só é plenamente realizado quando a usabilidade é garantida [1]. A usabilidade – medida em que um produto pode ser utilizado por usuários específicos para alcançar objetivos com eficácia, eficiência e satisfação em um contexto de uso – é um pilar para a aceitação e o sucesso clínico de ferramentas digitais na área da saúde [7, 11]. Uma revisão sistemática da literatura sobre sistemas de terapia da fala [1] confirma a crescente atenção à AU. No entanto, o panorama revela o predomínio de abordagens metodológicas unimodais. A prática mais comum consiste em conduzir Testes de Usabilidade (TUs) focados exclusivamente nos usuários finais (pacientes ou fonoaudiólogos) ou em aplicar isoladamente questionários de satisfação, como o *System Usability Scale* (SUS) [3]. Esse métodos, embora valiosos, oferecem uma perspectiva limitada da complexa interação humano-computador.

A principal lacuna identificada na literatura reside, portanto, nessa abordagem metodológica fragmentada. Os estudos existentes, apesar de seu valor, frequentemente negligenciam uma visão holística da Experiência do Usuário (UX). A inspeção por especialistas, como a Avaliação Heurística (AH) [14], é subutilizada, embora seja capaz de identificar problemas de usabilidade de forma rápida e econômica. De modo ainda mais acentuado, a dimensão afetiva da interação – um componente crítico para a satisfação e o engajamento a longo prazo – é pouco explorada com instrumentos psicométricos validados, como o *Self-Assessment Manikin* (SAM) [2]. Essa tendência gera escassez de protocolos de avaliação padronizados e multifacetados, dificultando a análise completa e comparável de aplicativos de fala.

Este estudo posiciona-se como uma resposta a essa lacuna, apresentando uma contribuição primariamente metodológica. O trabalho diferencia-se ao propor um protocolo inovador e multifacetado para a avaliação do aplicativo SofiaFala-Fono. Este protocolo integra sinergicamente três abordagens: (1) a AH por especialistas; (2) os TUs com fonoaudiólogos; e (3) a medição da resposta afetiva com o questionário SAM. A premissa é que esta triangulação metodológica oferece uma compreensão mais rica e robusta da UX, capturando as dimensões de eficácia, eficiência, satisfação subjetiva e o componente emocional. Dessa forma, o estudo oferece um modelo replicável que pode orientar o desenvolvimento de ferramentas digitais mais eficazes e bem aceitas na prática fonoaudiológica.

In: V Concurso de Trabalhos de Iniciação Científica (CTIC 2025). Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web (CTIC'2025). Rio de Janeiro/RJ, Brasil. Porto Alegre: Brazilian Computer Society, 2025.
© 2025 SBC – Sociedade Brasileira de Computação.
ISSN 2596-1683

¹Apoiado pela Pró-Reitoria de Cultura e Extensão Universitária da USP (PRCEU-USP) e aprovado pelo CEP do HCFMRP-USP, sob o CAEE n. 95.853.018.0.0000.5440 e o Parecer Consustanciado n. 2.885.905.

3 PROTOCOLO DE AU

Para conduzir a AU, desenvolveu-se um plano de teste em conformidade com as recomendações da literatura para a documentação, o planejamento e a execução de tais avaliações. O documento tem as seguintes seções: finalidade, metas e objetivos do teste; questões de investigação; características dos participantes; método (concepção do teste); cenário de tarefas; ambiente, equipamento e logística; papel do moderador; dados a serem coletados e medidas de avaliação; e conteúdo e apresentação do relatório [16]. Para o Teste de Usabilidade (TU), 3 fonoaudiólogos-clínicos foram recrutados por meio de rede social, e, para a AH, 3 especialistas em AU e *design de interfaces*, como indicado por Jakob Nielsen [11], a partir da indicação de docentes especialistas da área. Embora a literatura sugira 3-5 participantes para detectar a maioria dos problemas de usabilidade, a seleção de 3 avaliadores por método se justifica. Estudos demonstram que, com o aumento do número de testadores, os problemas identificados começam a se repetir, o que reduz a taxa de descoberta de novas falhas e limita a obtenção de novos *insights* significativos. Além disso, o número de tarefas e a complexidade da *interface* neste estudo eram baixos. Essa abordagem também otimiza a alocação de recursos, como esforço, tempo e custos, na sua realização [12, 13]. As Questões de Pesquisa (QPs) para guiar a AU eram:

QP1. O fluxo do *software* atende à forma como o usuário pensa e organiza seu fluxo de trabalho, promovendo uma execução eficiente das tarefas? [16]

QP2. Os usuários conseguem compreender adequadamente os símbolos e ícones? Quais deles apresentam dificuldades de entendimento e como isso impacta a tomada de decisão no uso? [16]

QP3. O aplicativo é intuitivo o suficiente para que novos usuários consigam utilizá-lo sem necessidade de treinamento prolongado?

QP4. Quais são as principais falhas de usabilidade que impedem os utilizadores de concluir as tarefas mais comuns e como isso afeta a experiência geral do usuário? [16]

QP5. O *feedback* fornecido pelo aplicativo é claro e útil para orientar os usuários durante as tarefas?

Este estudo de usabilidade empregou uma abordagem de métodos mistos, combinando dados quantitativos e qualitativos. A condução do estudo baseou-se no método *Testing Multiple User Groups*, que prevê a avaliação de um produto por dois grupos distintos de participantes [16]: um grupo de usuários finais, *e.g.* fonoaudiólogos; e um grupo de avaliadores, composto por especialistas em AU e *design*. Ambos os grupos executaram um conjunto de tarefas no aplicativo SofiaFala-Fono. Durante a interação, coletaram-se métricas de desempenho e as percepções subjetivas dos participantes sobre a experiência de uso. O uso dessa abordagem permitiu obter perspectivas complementares: o grupo de fonoaudiólogos forneceu dados sobre a aplicação do *software* no contexto clínico, refletindo seu raciocínio profissional e suas práticas de trabalho; por sua vez, o grupo de especialistas ofereceu uma análise técnica detalhada, fundamentada em princípios e heurísticas de usabilidade. A comparação entre as análises dos grupos resultou em uma avaliação mais completa e aprofundada da *interface* [16].

As heurísticas e suas respectivas questões (Ver Tabela 1 do Material Suplementar)² nortearam esta avaliação e baseiam-se em

[6], que consolida os trabalhos de Jakob Nielsen [14] e Bruce Tognazzini [18]. Para este estudo, as questões foram traduzidas para o português e adaptadas ao escopo do teste. A análise quantitativa dos resultados foi conduzida com base no método de [5], que apresenta uma abordagem para o cálculo do Percentual de Usabilidade (UP). Essa metodologia emprega uma escala de respostas com pesos específicos, conforme a Tabela 1.

O *Self-Assessment Manikin (SAM)* é um instrumento de avaliação pictórico e não verbal, projetado para mensurar as dimensões emocionais de satisfação, motivação e domínio. A escala usa um conjunto de figuras humanoides que representam diferentes níveis de intensidade para cada uma das 3 dimensões, permitindo ao participante avaliar sua própria resposta afetiva a um estímulo [2]. No presente TU, o SAM foi aplicado para capturar a reação emocional dos usuários durante a interação com o aplicativo, complementando os demais dados coletados sobre a experiência de uso.

O questionário aplicado aos fonoaudiólogos (Ver Tabela 2 do Material Suplementar)² baseou-se nos critérios de usabilidade da norma ISO 9241-11:2018 [7], que define os seguintes componentes:

- *Eficácia:* a exatidão e completeza com que os usuários alcançam metas especificadas;
- *Eficiência:* recursos empregados para alcançar resultados;
- *Satisfação do Usuário:* o grau em que as respostas físicas, cognitivas e emocionais do usuário, resultantes do uso de um sistema, produto ou serviço, atendem às suas necessidades e expectativas.

O questionário foi adaptado de uma avaliação prévia do SofiaFala [8]. O processo de adaptação envolveu a exclusão de questões do estudo original e a adição de novos itens, formulados para abranger as funcionalidades específicas da nova versão do SofiaFala-Fono. Além disso, registraram-se os comentários espontâneos dos participantes durante a sessão de teste para compor a análise qualitativa.

Neste estudo, a ferramenta *Remote User eXperience Artificial Intelligence LABoratory (RUXAILAB)*³ foi empregada para aplicar: um questionário direcionado aos especialistas, baseado em heurísticas; e outro com questões personalizadas para os fonoaudiólogos [4, 5]. A plataforma também viabilizou a análise dos dados, por meio da geração automática de relatórios, que incluem diversas métricas de usabilidade e podem ser exportados para uso posterior.

As sessões de avaliação foram conduzidas em junho de 2025, com duração de até 60 minutos para cada participante. Esse tempo foi distribuído entre uma fase inicial de apresentação e instruções (pré-teste), a execução de doze tarefas (Ver Tabela 3 do Material Suplementar)² com duração de 30 minutos e uma entrevista final de esclarecimento (pós-teste).

Os testes com os fonoaudiólogos ocorreram em seus consultórios reais, enquanto para os especialistas em AU o ambiente foi simulado para replicar o contexto de uso. A interação com a *interface* foi capturada por observação direta da tela do dispositivo *Android* nas sessões presenciais e por transmissão de tela nas sessões remotas, realizadas por videoconferência quando a avaliação presencial não era viável. Independentemente da modalidade, o moderador registrou anotações sobre o comportamento dos usuários, esclareceu dúvidas e realizou perguntas de acompanhamento para aprofundar a compreensão das ações e decisões dos participantes.

³Uma plataforma *web* de código aberto projetada para a condução remota de TUs e avaliações heurísticas. Disponível em: <https://ruxailab-prod.web.app/>

²Material Suplementar disponível em: <https://github.com/lembor/ctic2025-apendice>

Tabela 1: Respostas para questões das heurísticas

Opção	Valor	Descrição
SIM, em TODOS os casos	1.0	O que está sendo avaliado foi cumprido em todos os casos verificados.
Sim, mas faltam casos	0.66	O que está sendo avaliado nem sempre foi cumprido, mas na maioria dos casos.
Nem sempre	0.33	O que está sendo avaliado praticamente não foi cumprido em nenhum caso.
NÃO, em NENHUM caso	0.0	O que está sendo avaliado não foi cumprido em nenhum dos casos.
Não aplicável	-	O que está sendo avaliado não é encontrado no “site”, mas não é relevante.
Não é um problema	-	O que está sendo avaliado não é um problema.
Impossível verificar	-	O que está sendo avaliado não pode ser cumprido.

Fonte: Metodologia adaptada de [5].

4 RESULTADOS E DISCUSSÃO

Nesta seção são descritos e discutidos os resultados da AH e do TU, que incorporou o SAM.

4.1 Análise Heurística

Para quantificar os achados da AH do SofiaFala-Fono, foi empregada a metodologia do cálculo do UP, de [5]. O aplicativo alcançou um UP médio de 54,53%, que indica uma usabilidade problemática e a necessidade de melhorias significativas na interface. Os resultados individuais de cada avaliador são detalhados na Tabela 2. Metodologicamente, o cálculo do índice (Equação 1) desconsidera as heurísticas classificadas como “Não aplicável” ou “Não é um problema”. A incerteza metodológica associada às heurísticas marcadas como “Impossível verificar” é tratada com a definição de um intervalo de confiança, delimitado por um cenário otimista (UP_β) e um pessimista (UP_γ). Estes cenários indicam, respectivamente, a usabilidade caso todas as incertezas se resolvessem como problemas inexistentes (pontuação 1.0) ou como violações críticas (pontuação 0.0).

$$UP_i = \frac{\sum_{i=1}^M (Score_i) + 0.6 \cdot \sum_{j=1}^N (Score_j) + 0.3 \cdot \sum_{k=1}^O (Score_k) + Z \cdot \sum_{l=1}^P (Score_l)}{(M + N + O + X) + P} \quad (1)$$

Tabela 2: Resultados do UP por Avaliador (Av.), com os cenários Base (UP_α), Otimista (UP_β), Pessimista (UP_γ) e média geral

Av.	UP_α	UP_β	UP_γ	Resultado Sugerido
1	60.89%	0.639	0.562	60.89% (+4.94% / -7.69%)
2	57.57%	0.576	0.576	57.57% (+0.00% / 0.00%)
3	45.14%	0.480	0.428	45.14% (+6.40% / -5.26%)
Média	54.53%	0.565	0.522	54.53% (+3.78% / -4.32%)

Fonte: Metodologia adaptada de [5].

Para além da pontuação geral, a análise da severidade média dos problemas da Figura 1 permitiu identificar as áreas que necessitam de atenção prioritária. As heurísticas consideradas mais críticas foram a H11, revelando falhas na proteção do trabalho do usuário por falta de salvamento automático; a H13, que indica pouca autonomia e feedback inadequado sobre o status do sistema; e a H8,

relacionada à falta de mecanismos para prevenção de erros. Com um grau de criticidade ligeiramente inferior, mas ainda relevante, destacam-se a H6, que aponta para uma baixa flexibilidade e eficiência de uso, com ausência de atalhos para acelerar tarefas; a H15, que sugere problemas de latência com atrasos perceptíveis na resposta do sistema; a H3, que indica dificuldades no controle e liberdade do usuário para navegar e desfazer ações; e a H5, que aponta para uma alta carga cognitiva, exigindo que o usuário memorize fluxos em vez de reconhecer padrões de forma intuitiva. Em conjunto, estes sete pontos devem ser o foco das próximas atualizações do software.

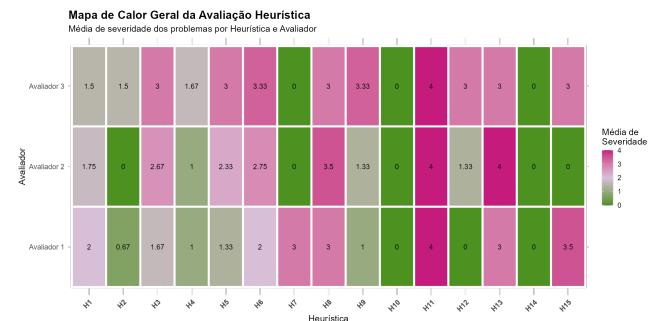


Figura 1: Mapa de Calor Geral da Avaliação Heurística.

4.2 Análise do Teste de Usabilidade

A eficácia do aplicativo, medida pela taxa de conclusão das tarefas, apresentou resultados variados. A taxa de sucesso do Participante 1 foi de 83,33%, enquanto 2 e 3 alcançaram 66,67% cada.

A avaliação da experiência do usuário revelou uma acentuada dicotomia no desempenho do aplicativo, com tarefas de baixa complexidade gerando uma resposta positiva, em forte contraste com fluxos de trabalho essenciais que resultaram em frustração. Ações atômicas, como Entrar no aplicativo (T2) e Excluir um exercício (T7), geraram uma resposta emocional altamente positiva, com pontuações médias próximas da máxima (≥ 8.67) nas dimensões de satisfação, motivação e domínio do questionário *Self-Assessment Manikin* (SAM) – instrumento que permitiu quantificar a dimensão afetiva da interação, um componente pouco explorado na literatura – (Ver Figura 2). Em contrapartida, fluxos de trabalho mais complexos, notadamente a edição de dados de pacientes (T6) e a listagem

de treinos (T11), resultaram em uma experiência marcadamente negativa, registrando as menores pontuações (≤ 1.67) e indicando frustração e perda de controle. As respostas do questionário de usabilidade (Ver Figura 3) corroboraram estes achados, apontando para falhas funcionais específicas, como a incapacidade do sistema de salvar alterações (Q10) e a ausência de *feedback* de confirmação na atribuição de treinos (Q7). Adicionalmente, foi reportada uma falha de *feedback* na tarefa de Filtrar pacientes (T8); embora a interação tenha sido considerada satisfatória, a falta de um resultado correto, conforme indicado na Q8, resultou em uma baixa pontuação de Controle (4.33). Em síntese, os resultados indicam que, embora as funcionalidades básicas sejam robustas, falhas críticas em fluxos de trabalho essenciais comprometem a viabilidade do aplicativo, exigindo uma reformulação dos módulos de edição e gestão de treinos para uma experiência de uso coesa e eficaz.

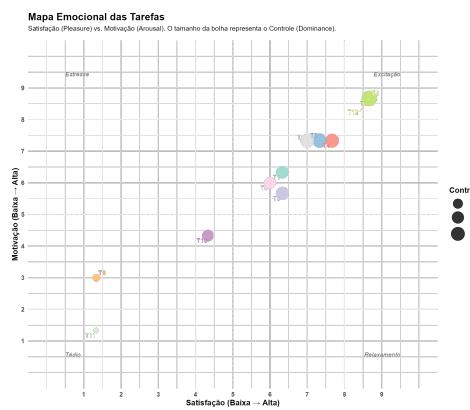


Figura 2: Visão Geral do Mapa Emocional das Tarefas.

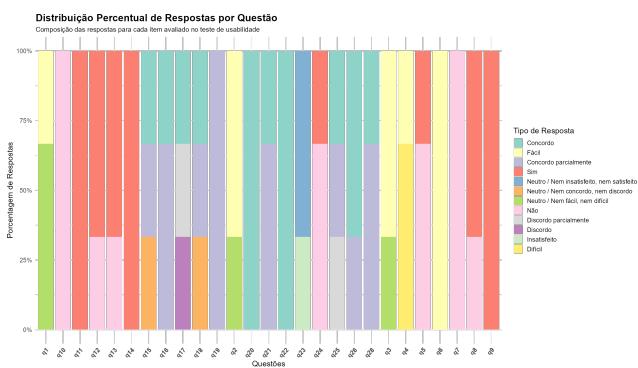


Figura 3: Distribuição Percentual de Respostas por Questão.

5 CONCLUSÃO E TRABALHOS FUTUROS

Este estudo respondeu à carência de *frameworks* de avaliação integrados na Fonoaudiologia ao propor e validar um protocolo de AU multifacetado que se diferencia das abordagens unimodais predominantes na literatura. Ao integrar sinergicamente a AH, TUs e a medição afetiva com o SAM, o protocolo oferece uma análise mais

completa do que métodos isolados. A aplicação deste protocolo ao SofiaFala-Fono demonstrou seu valor diagnóstico, revelando falhas críticas de usabilidade (UP de 54,53%) e frustração do usuário em fluxos essenciais. Portanto, a principal contribuição não reside apenas nos achados específicos do aplicativo, mas no próprio protocolo como um modelo replicável para diagnósticos de UX holísticos e açãoáveis no campo. Os trabalhos futuros seguirão duas frentes: a curto prazo, o redesenho iterativo do SofiaFala-Fono para corrigir as falhas identificadas e garantir sua viabilidade clínica. A médio prazo, a validação do protocolo será expandida, aplicando-o a outras tecnologias assistivas e com grupos de usuários mais amplos.

Agradecimentos. Os autores agradecem o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2024/15912-0) e da PRCEU-USP por meio do Programa Unificado de Bolsas (PUB).

REFERÊNCIAS

- Geertruida Aline Attwell, Kwabena Ebo Bennin, and Bedir Tekinerdogan. 2022. A Systematic Review of Online Speech Therapy Systems for Intervention in Childhood Speech Communication Disorders. *Sensors* 22, 24 (2022). doi:10.3390/s22249713
- Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59. doi:10.1016/0005-7916(94)90063-9
- John Brooke. 1996. SUS – a quick and dirty usability scale. 189–194.
- Marc Capdevila, Karine Rodrigues, Afra Pascual, Toni Granollers, and Èric Monné. 2024. Whitepaper: Remote usability laboratory to perform user and expert evaluations. In *Proceedings of the XXXIII International Conference on Human Computer Interaction* (Lleida, Spain) (*Interacción '23*). Association for Computing Machinery, New York, NY, USA, Article 20, 9 pages. doi:10.1145/3612783.3612804
- Marc Gonzalez Capdevila, Karine Aparecida Pistili Rodrigues, Tales Augusto Sartório Furlan, and Toni Granollers. 2025. Quantifying heuristic evaluation. *Computer Standards Interfaces* 92 (2025), 103891. doi:10.1016/j.csi.2024.103891
- Toni Granollers. 2018. Usability evaluation with heuristics, beyond nielsen's list. In *The Eleventh International Conference on Advances in Computer-Human Interactions (ACHI 2018)*, 1–6.
- International Organization for Standardization. 2018. ISO 9241-11:2018 Ergonomics of human-system interaction – Part 11: Usability: Definitions and Concepts.
- Alessandra Alaniz Macedo, Vinícius de S. Gonçalves, Patrícia P. Mandrá, Vivian Motti, Renato F. Bulcão-Neto, and Kamila Rios da Hora Rodrigues. 2024. A mobile application and system architecture for online speech training in Portuguese: design, development, and evaluation of SofiaFala. *Multimedia Tools and Applications* (07 Aug 2024). doi:10.1007/s11042-024-19980-5
- Fernando Meloni, Bianca B. Sichieri, Patrícia Mandrá, Renato De Freitas Bulcão-Neto, and Alessandra Alaniz Macedo. 2022. Detection and Evaluation of Speech Intelligibility with Speech Tool. In *2022 XLVIII Latin American Computer Conference (CLEI)*, 1–9. doi:10.1109/CLEI56649.2022.9959936
- Christine Murad, Cosmin Munteanu, Benjamin R. Cowan, and Leigh Clark. 2019. Revolution or Evolution? Speech Interaction and HCI Design Guidelines. *IEEE Pervasive Computing* 18, 2 (2019), 33–45. doi:10.1109/MPRV.2019.2906991
- J. Nielsen. 1994. *Usability Engineering*. Elsevier Science.
- Jakob Nielsen. 2000. *Why You Only Need to Test with 5 Users*. www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ Acesso em: 18/08/25.
- Jakob Nielsen and Thomas K. Landauer. 1993. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) (*CHI '93*). Association for Computing Machinery, New York, NY, USA, 206–213. doi:10.1145/169059.169166
- Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256. doi.org/10.1145/97243.97281
- Pedro Henrique D'Almeida Giberti Rissato and Alessandra Alaniz Macedo. 2021. SofiaFala: Software Inteligente de Apoio à Fala. In *Anais Estendidos do XVII Simpósio Brasileiro de Sistemas Multimídia e Web* (Minas Gerais). SBC, 91–94. doi:10.5753/webmedia_estendido.2021.17620
- J. Rubin, D. Chisnell, and J. Spool. 2011. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley.
- Francisco Carlos Souza, Alinne Souza, Yethzabeth Carolina, Patricia Pupin, and Alessandra Macedo. 2019. An Analysis of Visual Speech Features for Recognition of Non-articulatory Sounds using Machine Learning. *International Journal of Computer Applications* 177 (11 2019), 1–9. doi:10.5120/ijca2019919393
- B Tognazzini. 2014. First Principles of Interaction Design (Revised & Expanded). <https://asktug.com/atc/principles-of-interaction-design/> Acesso em: 18/08/25.