**Interação gestual para acessibilidade de vídeos na Web por
pessoas com deficiência visual**

Márcio Maestrello Funes

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências
de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Márcio Maestrello Funes

Interação gestual para acessibilidade de vídeos na Web por pessoas com deficiência visual

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Rudinei Goularte

Coorientador: Profa. Dra. Renata Pontin de Mattos Fortes

USP – São Carlos
Novembro de 2018

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

F319i Funes, Márcio Maestrello
Interação gestual para acessibilidade de vídeos na
Web por pessoas com deficiência visual / Márcio
Maestrello Funes; orientador Rudinei Goularte;
coorientadora Renata Pontin de Mattos Fortes. --
São Carlos, 2018.
130 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2018.

1. acessibilidade. 2. deficiência visual. 3.
interfaces naturais. 4. gestos. I. Goularte,
Rudinei, orient. II. Fortes, Renata Pontin de
Mattos, coorient. III. Título.

Márcio Maestrello Funes

**Gestural interaction for accessibility of Web videos by
visually impaired users**

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP,
in partial fulfillment of the requirements for the
degree of the Master Program in Computer Science
and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. Rudinei Goularte

Co-advisor: Profa. Dra. Renata Pontin de
Mattos Fortes

USP – São Carlos
November 2018

Este trabalho é dedicado, especialmente, aos meus pais Márcio e Solange, que proporcionaram todo suporte para minha formação, para minha filha Suri que me motivaram a continuar na conclusão deste desafio. Dedico também aos meus queridos familiares e amigos que me ajudaram ao longo desta caminhada e que, de alguma forma, colaboraram para esta pesquisa.

AGRADECIMENTOS

Gostaria de agradecer aos meus pais, minha filha Suri e meus familiares por todo o amor, carinho e apoio durante a realização deste mestrado. Mesmo distantes, sei que posso contar com o apoio incondicional de vocês.

Agradeço à Universidade de São Paulo, em especial, ao Instituto de Ciências Matemáticas e de Computação pela estrutura fornecida assim como os diversos professores e funcionários, que sempre se dispuseram a ajudar no que fosse necessário. Agradeço ainda a Capes pelo apoio financeiro despendido para a realização deste trabalho. Agradeço aos meus colegas e amigos do laboratório Intermídia pelos momentos de apoio e descontração que pudemos compartilhar.

Por último, mas não menos importante, agradeço ao meu orientador, Prof. Rudinei Goularte e a minha coorientadora Prof. Renata Fortes, por toda a ajuda e compreensão, além das diversas sugestões e orientações indispensáveis no desenvolvimento deste trabalho.

A todos meu MUITO OBRIGADO!

*“As invenções são, sobretudo,
o resultado de um trabalho de teimoso.”
(Santos Dumont)*

RESUMO

FUNES, M. M. **Interação gestual para acessibilidade de vídeos na Web por pessoas com deficiência visual**. 2018. 131 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

Nos últimos anos o interesse por plataformas de compartilhamento de conteúdo multimídia na Web, como vídeos, tem crescido consideravelmente e atraindo usuários com diferentes perfis, incluindo pessoas com deficiência visual. Para tais usuários, diretrizes de acessibilidade e usabilidade devem ser aplicadas para que o acesso a estes conteúdos multimídias seja garantido. Apesar de iniciativas como a WCAG 2.0 (*Web Content Accessibility Guidelines*) definirem diretrizes sobre a acessibilidade para reprodutores de vídeos na Web, usuários com deficiência visual ainda encontram barreiras e dificuldades para acessar este tipo de conteúdo. Em muitos casos reprodutores de vídeo na Web não oferecem total suporte a acessibilidade, dificultando assim o acesso ao conteúdo por tal público, seja por meio de mouse ou de navegação por teclado via leitores de tela.

Nesse sentido, este trabalho teve como objetivo investigar se interações por meios de gestos fornecem melhoria nos níveis de acessibilidade durante o acesso a vídeos na Web realizado por usuários com deficiência visual, a fim de que as barreiras de interação fossem reduzidas ou eliminadas. Inicialmente estudou-se os diferentes tipos de sensores gestuais, bem como se estes sensores poderiam ser utilizados para atingir o objetivo proposto. A partir desses estudos, foi desenvolvido um *framework* para captura, representação e interpretação de gestos. Em seguida, foi desenvolvida uma aplicação Web, utilizando o *framework* desenvolvido para viabilizar possíveis testes com usuários. Finalmente, um estudo de caso com 38 usuários com deficiência visual foi conduzido, a fim de levantar indícios que ajudem a entender se a utilização de gestos como modo de interação entre pessoas com deficiência visual e reprodutores de vídeo oferece melhoria de acessibilidade.

Palavras-chave: acessibilidade, deficiência visual, interfaces naturais, gestos.

ABSTRACT

FUNES, M. M. **Gestural interaction for accessibility of Web videos by visually impaired users**. 2018. 131 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

Recently the interest in multimedia Web content sharing platforms, such as videos, has increased considerably attracting users with different profiles, including people with visual impairment. For such users, accessibility and usability guidelines should be applied for access guaranteed of these multimedia content. Although initiatives such as WCAG 2.0 (Web Content Accessibility Guidelines) set guidelines for video players, users with visual impairment still find barriers and difficulties to access this type of content. In many cases Web video players do not provide full accessibility support, making it difficult to access content by such a public, either through mouse or keyboard navigation by screen readers.

This paper aims at investigating if gesture interactions provide improve improvement in accessibility levels during access to Web videos by visually impaired users, in order to reduce or eliminate the interactions barriers. We initially analyzed the different types of gesture sensors, as well as if these sensors could be used to achieve the proposed objective. Based on this analysis, we developed a framework for capture, representation and interpretation of gestures. Then, we developed a Web application, using the framework guidelines developed to enable users tests. Finally, we then run a study case with 38 users with visual impairment, in order to get evidence to help understand if the use of gestures as a mode of interaction between people with visual impairments and video players offers improved accessibility.

Keywords: accessibility, visually impaired, natural interface, gesture.

LISTA DE ILUSTRAÇÕES

Figura 1 – Na direita a lousa eletrônica e a esquerda o documento multimídia.	38
Figura 2 – Usuário durante escovação e sistema de pontuação.	39
Figura 3 – Na direita Steve utilizando o <i>EyeTap</i> e na esquerda o <i>Hydraulophone</i>	41
Figura 4 – Exemplo da estrutura de uma Interface Baseada em Gestos	45
Figura 5 – Usuário realizando um gesto tridimensional com o <i>Wii Remote</i>	46
Figura 6 – <i>Myo Armband</i>	46
Figura 7 – Usuário realizando um gesto tridimensional com o <i>Leap Motion</i>	47
Figura 8 – Usuário realizando um gesto tridimensional com o <i>Kinect</i>	48
Figura 9 – Usuário realizando um gesto tridimensional com o <i>Asus Xtion</i>	48
Figura 10 – Usuário controlando uma aplicação por meio de sua <i>Webcam</i>	49
Figura 11 – Exemplo de representação gráfica do gesto a ser realizado.	50
Figura 12 – Interação do usuário com um monitor e <i>smartphone</i>	51
Figura 13 – Usuário realizando um gesto tridimensional e visualizando o <i>feedback</i> da ação.	52
Figura 14 – Comparativo da quantidade de trabalhos encontrados e selecionados por base de pesquisa	58
Figura 15 – Elementos do estudo de caso e suas interações	64
Figura 16 – Visão geral do contexto de aplicação do <i>Gesture4All</i>	65
Figura 17 – Visão completa da Arquitetura do <i>Gesture4All</i>	67
Figura 18 – Aplicação Web desenvolvida para viabilizar o teste	74
Figura 19 – Agrupamento das variáveis coletadas no estudo de caso	76
Figura 20 – Comparativo entre Programas de Leitura e Dificuldades	80
Figura 21 – Sensores preferidos dos usuários do grupo 2	89
Figura 22 – Comparativo entre interações gestuais de forma geral e com o reprodutor de vídeo	91
Figura 23 – Comparativo de tarefas completadas com sucesso (satisfatórias) entre intera- ção usual e gestual	98
Figura 24 – Comparativo de número de erros cometidos entre interação usual e gestual	99
Figura 25 – Comparativo do tempo de cada gesto por tarefa entre interação usual e gestual	99

LISTA DE QUADROS

Quadro 1 – Quantidade de trabalhos encontrados por base de pesquisa	57
---	----

LISTA DE ALGORITMOS

LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Exemplo de uso código implementado no módulo de controle de interface	70
Código-fonte 2 – Exemplo de uso código implementado nos módulos da camada de <i>feedback</i>	71

LISTA DE TABELAS

Tabela 1 – Programa de Leitura (PI) x Dificuldades(D)	80
Tabela 2 – Usuários e suas dificuldades	83
Tabela 3 – Ações no <i>Smartphone</i> de maior facilidade e dificuldade	84
Tabela 4 – Ações na <i>Webcam</i> de maior facilidade e dificuldade	85
Tabela 5 – Ações no <i>Myo</i> de maior facilidade e dificuldade	86
Tabela 6 – Ações no Leap Motion de maior facilidade e dificuldade	88
Tabela 7 – Comparativo de uso entre os sensores por meio da escala Likert	88
Tabela 8 – Comparativo de facilidade entre sensores e ações executadas no reprodutor de vídeo	90
Tabela 9 – Comparativo de dificuldade entre sensores e ações executadas no reprodutor de vídeo	90
Tabela 10 – Interações no Vídeo (iV) x Dificuldades(D)	92
Tabela 11 – Dados sobre personalização, cotidiano e melhorias	92
Tabela 12 – Dados quantitativos gestuais coletados por meio do sensor <i>Smartphone</i>	95
Tabela 13 – Dados quantitativos gestuais coletados por meio do sensor <i>Webcam</i>	95
Tabela 14 – Dados quantitativos gestuais coletados por meio do sensor <i>Myo</i>	96
Tabela 15 – Dados quantitativos gestuais coletados por meio do sensor Leap Motion	97
Tabela 16 – Dados quantitativos coletados por meio da interação usual realizada pelo usuário (leitores de tela e teclado)	97
Tabela 17 – Dados sobre o melhor <i>feedback</i> na interação gestual	100
Tabela 18 – Dados sobre o <i>feedback</i> ideal na interação gestual	101

LISTA DE ABREVIATURAS E SIGLAS

C&A	Captura e Acesso
EEG	Electroencephalogram
eMAG	Modelo de Acessibilidade Brasileiro
HMM	<i>Hidden Markov Models</i>
NUI	<i>Natural User Interface</i>
UAAG	<i>User Agent Accessibility Guidelines</i>
W3C	World Wide Web Consortium
WAI	Web Accessibility Initiative
WCAG	Web Content Accessibility Guidelines
WHO	<i>World Health Organization</i>

LISTA DE SÍMBOLOS

SUMÁRIO

1	INTRODUÇÃO	31
1.1	Motivação	33
1.2	Objetivo	36
1.3	Organização do texto	36
2	REFERENCIAL TEÓRICO	37
2.1	Computação Ubíqua e Interfaces Naturais	37
2.2	Interfaces Baseadas em Gestos	43
2.2.1	<i>Segmentação</i>	44
2.2.2	<i>Reconhecimento</i>	49
2.2.3	<i>Interpretação</i>	52
2.3	Acessibilidade de vídeos na Web	53
2.4	Mapeamento Sistemático	56
2.5	Considerações finais	60
3	ESTUDO DE CASO COM USUÁRIOS CEGOS E PESSOAS COM DEFICIÊNCIA VISUAL	63
3.1	Descrição do <i>framework Gesture4All</i>	64
3.1.1	<i>Contextos de Uso</i>	64
3.1.2	<i>Arquitetura do Gesture4All</i>	66
3.2	Protocolo de condução de testes	71
3.2.1	<i>Planejamento</i>	72
3.2.2	<i>Condução dos testes</i>	72
3.2.3	<i>Coleta de dados</i>	74
3.3	Resultados	76
3.3.1	<i>Sobre o perfil dos usuários</i>	76
3.3.2	<i>Sobre os sensores gestuais utilizados</i>	82
3.3.3	<i>Sobre o uso de gestos</i>	90
3.3.4	<i>Sobre o feedback utilizado</i>	99
3.4	Considerações finais	101
4	CONCLUSÕES	103
4.1	Principais contribuições	105
4.2	Limitações	106

4.3	Trabalhos futuros	107
	REFERÊNCIAS	109
	GLOSSÁRIO	121
ANEXO A	ANEXO A - QUESTIONÁRIO GRUPO 1	123
ANEXO B	ANEXO B - QUESTIONÁRIO GRUPO 2	127

INTRODUÇÃO

Consumir e compartilhar vídeos diariamente tem se tornado uma prática comum entre os usuários da Web. Fatores como redes sociais, plataformas gratuitas de compartilhamento de vídeo, mais acesso a infraestruturas de banda larga e entre outros fatores contribuem para esta prática. Para uma maior compreensão deste comportamento, a empresa Cisco (2016), um dos líderes mundiais sobre tráfego de dados, divulgou suas previsões sobre o tráfego de vídeos na Web, a cada segundo quase um milhão de minutos de conteúdo de vídeo poderá circular pela Web e 82% de todo tráfego na Web, até 2020, poderá ser vídeos online, indicando assim um aumento de quatro vezes entre os anos de 2015 e 2020. Para essas previsões, foram considerados os diversos tipos de vídeos online existentes como: vídeos de curta duração (por exemplo, YouTube¹), vídeos de longa duração (por exemplo, Hulu²), *streaming* de vídeos (por exemplo, Netflix³) e streaming ao vivo de vídeos (por exemplo, Twitch⁴).

Segundo Highq (2017), 74% de todo tráfego mundial em 2017 será de vídeos online e em apenas um mês são enviados mais conteúdos de vídeo para Web do que a TV criou em três décadas (BRAFTON, 2015; CLICKBANK, 2017). Com relação a comportamento na Web, 78% dos usuários consomem vídeos todas as semanas e 55% consomem vídeos todos os dias (GROUPON, 2014), sobre tempo de consumo do vídeo, 65% dos usuários consomem mais que 3/4 do vídeo que escolhem (REELNREEL, 2015).

As redes sociais também possuem uma grande parcela de contribuição para o aumento do consumo de vídeo na Web. Segundo Doubleshot (2016), 500 milhões de pessoas consomem vídeos no Facebook⁵ todos os dias, no Snapchat⁶ são consumidos 10 bilhões de vídeos todos os dias e 82% dos usuários do Twitter consomem vídeos enquanto estão na rede social (DOU-

¹ <https://www.youtube.com>

² <https://www.hulu.com/>

³ <https://www.netflix.com>

⁴ <https://www.twitch.tv/>

⁵ <https://www.facebook.com>

⁶ <https://www.snapchat.com>

BLESHOT, 2016). Com relação aos dispositivos utilizados, 28% dos usuários de *smartphones* assistem ao menos um vídeo todos os dias e 33% dos usuários de *tablets* consomem a média de uma hora de vídeo todos os dias (BRAFTON, 2015).

Dentre os exemplos citados anteriormente, as plataformas de vídeo Netflix e YouTube merecem destaque. O Netflix conta com mais de 93 milhões de membros em mais de 190 países, esses usuários consomem mais de 125 milhões de programas de TV e filmes por dia (CENTER, 2017). O Youtube conta com 1,32 bilhões de usuários onde 4,95 bilhões de vídeos são consumidos todos os dias, isso representa 900 milhões de visitas e 3,25 bilhões de horas de vídeos consumidos todos os meses (YOUTUBE, 2015). Esses dados sobre consumo de vídeo e comportamento de usuários demonstram a importância que pesquisas na área de multimídia possuem para a sociedade, assegurar que todos esses usuários que desejam consumir vídeos online possam acessar conteúdos sem barreiras e dificuldades nunca foi tão necessário.

Com o crescimento dos usuários da Web, questões de acessibilidade tem se tornado cada vez mais frequentes e importantes tópicos para a sociedade, principalmente se tratando de inclusão social. Diversos serviços essenciais como transações bancárias, acesso a portais governamentais e informações sobre o cotidiano são frequentes na Web, garantir que qualquer pessoa com alguma deficiência ou com restrições temporárias de mobilidade torna-se fator de caráter essencial.

Segundo a *World Health Organization* (WHO), a estimativa de pessoas com deficiência visual no mundo é de 285 milhões, sendo 39 milhões cegos e 246 milhões com baixa visão. (WHO, 2014). No Brasil, segundo o censo demográfico do ano de 2010, 45,6 milhões de pessoas declararam ter algum tipo de deficiência, representando 23,9% do total da população. A deficiência visual é a mais frequente, atinge 3,5% dos brasileiros totalizando seis milhões de pessoas que possuem dificuldade severa em enxergar e mais de 506 mil cegas (IBGE, 2012). Segundo Thatcher e Waddell (2003), acessibilidade na Web corresponde a possibilitar que qualquer usuário, utilizando qualquer agente seja ele *hardware* ou *software* alcance a interação desejada dentro de uma página da Web. Apesar de existir diversas diretrizes de acessibilidade, como a WCAG (2008) do W3C e leis federais em diversos países, relacionados à acessibilidade, ainda existem muitas barreiras que impossibilitam o uso por parte de pessoas com deficiência visual.

Um exemplo de barreira de acessibilidade em páginas da Web pode ser encontrado em reprodutores de vídeo. Em muitos casos, programa de leitura de tela não conseguem identificar corretamente os botões que compõem a interface do reprodutor, inviabilizando assim a interação por parte de usuário que fazem uso desse recurso e caracterizando assim uma barreira de acessibilidade. Para melhor compreensão de como a acessibilidade tem sido utilizada em páginas da Web que oferecem reprodutores de vídeos em seus conteúdos, Rosas-Villena *et al.* (2015) realizaram uma pesquisa incluindo os 50 sites mais acessados de acordo com o site Alexa ⁷.

⁷ <http://www.alexa.com/>

Dentre os 50 sites, foram destacados dez sites que permitem assistir e realizar *uploads* de vídeo, todos apresentam controle de *play* e *pause*, são eles: *Facebook*, *YouTube*, *Yahoo*, *Vk*, *MSN*, *Instagram*, *Tumblr*, *Apple*, *XVideos* e *Microsoft*. Três deles (*Instagram*, *Tumblr* e *Microsoft*) não possuem controle de volume e quatro sites (*Youtube*, *Vk*, *MSN*, *XVideos*) possuem configuração de qualidade. Segundo a pesquisa, mais importante que os controles básicos de reprodução de vídeo, para fins de acessibilidade, são características especiais como: *Closed Captions*, *realce*, configuração de texto, acesso via teclado, busca, definições de preferência e configuração da barra de ferramentas.

A pesquisa também faz considerações com relação a tecnologia em que o reprodutor de vídeo foi desenvolvido, como exemplo o *Flash*, muitas vezes não oferece suporte para acessibilidade. A falta destas características especiais e a tecnologia empregada, em muitos casos, impossibilita o acesso ao conteúdo de vídeo por usuários com deficiência visual, caracterizando assim uma barreira de acessibilidade. Apesar de existirem iniciativas que aplicam esforços para melhorar a acessibilidade para vídeos na Web, há ainda a necessidade de desenvolver soluções de acessibilidade para conteúdos de vídeo (ROSAS-VILLENA *et al.*, 2015).

1.1 Motivação

Plataformas de vídeos na Web tem crescido consideravelmente demonstrando o interesse por diversos tipos de usuário incluindo aqueles que possuem algum tipo de deficiência visual. Entretanto, segundo Tanaka e Rocha (2011), muitos desenvolvedores não possuem formação acadêmica que proporcione profunda compreensão sobre questões de acessibilidade para acompanhar esse crescimento.

Segundo Freire, Lara e Fortes (2013), para que pessoas com deficiências visuais possam utilizar sites de forma independente e satisfatória é importante que profissionais e pesquisadores, que trabalham com o desenvolvimento de sistemas Web tenham conhecimentos em acessibilidade, usabilidade e seus recursos tecnológicos. Apesar de todos os reprodutores e plataformas de vídeo demandarem as mesmas funções (*play*, *pause*, *stop*, avançar, retroceder, dentre outras), não existe padronização da localização dessas funções e quais comandos a elas estão associados, o que pode inviabilizar a usabilidade de usuários com deficiência visual.

Para contornar essas barreiras de acessibilidade encontradas na Web (FREIRE; LARA; FORTES, 2013), existem ferramentas que permitem navegar e acessar vídeos de forma autônoma. A ferramenta mais utilizada para esse objetivo são os leitores de tela que efetuam a leitura audível de textos para o usuário.

Programas de leitura de tela são instalados em computadores ou *smartphone*, e realizam a leitura do texto contido na página por meio de navegação em foco, geralmente começando no canto superior esquerdo e avançando no texto por comandos do teclado de forma horizontal até o canto inferior direito quando utilizados em computadores (BORODIN *et al.*, 2010). A navegação

em foco permite utilizar apenas o teclado ou o toque para navegação, assim é possível selecionar um objetivo gráfico da tela e ouvir a leitura do texto ou descrição da imagem ao qual a seleção se encontra.

Leitores de tela dependem da forma em que a interface gráfica, como um todo, está desenvolvida e disponibilizada ao usuário. No quesito sites e reprodutores de vídeos, botões de controle sem identificação textual impossibilitam que leitores de tela consigam identificar e informar ao usuário o foco de seleção. Isso impossibilita o usuário de controlar as funcionalidades do vídeo, muitas vezes essa falta de controle causa frustração (AIZPURUA; ARRUE; VIGO, 2013) e o sentimento de exclusão digital (MATTOS; CHAGAS, 2008) e (CARVALHO, 2003) no usuário.

Barreiras em acessibilidade também podem existir dependendo do dispositivo no qual o vídeo está sendo executado. Dispositivos como *smartphones*, *tablets* e televisores geralmente dispõem seus reprodutores de vídeo de modos diferentes da plataforma *desktop*, em alguns casos, esses dispositivos não permitem navegação via teclado ou opções de leitura de tela. Muitas vezes apenas disponibilizam interação via toque ou controle remoto. Essa mudança na interface do dispositivo pode ser uma barreira aos usuários com deficiência visual, quando dificulta a operação das funções necessárias para a reprodução de um vídeo.

Para que exista padronização e acesso à informação, a *Web Accessibility Initiative* (WAI) (W3C, 2012) se dedica a realizar estudos e definição de diretrizes para acessibilidade e usabilidade na Web. Para vídeos, a WAI recomenda os seguintes procedimentos durante o desenvolvimento de um reprodutor de vídeos Web:

- Oferecer as mesmas funcionalidades presentes por meio do uso de interface como o *mouse*.
- Fornecer ao usuário opções de ajuda e documentação sobre as funcionalidades presentes no reprodutor de vídeo como atalhos e modos de uso.
- Indicar de forma visual, qual elemento presente na interface do vídeo está selecionado por meio de seleção de foco.
- Possibilitar a inclusão de audiodescrição e anotações no vídeo.
- Modificar o tamanho e a coloração do texto.

Com relação à deficiência visual, a WAI indica que sites acessíveis possuam: movimentação do navegador via teclado; inclusão do cursor de foco e a movimentação do mesmo via teclado; o uso do botão Enter tendo a mesma funcionalidade do botão esquerdo do *mouse*.

As pesquisas atuais têm apresentado maior enfoque no desenvolvimento de reprodutores (GONZALEZ *et al.*, 2011; ROSAS-VILLENA *et al.*, 2015) e em diretrizes e recomendações (W3C, 2012). Poucos trabalhos têm sido feitos para melhorar o nível de acessibilidade especificamente entre os usuários com deficiência visual e reprodutores de vídeo na Web (a discussão destes trabalhos será evidenciada na [Seção 2.4](#)

Na busca por modos alternativos para melhoria dos níveis de acessibilidade, uma possível investigação no contexto de vídeos na Web, pode ser encontrada em pesquisas na área de Computação Ubíqua, mais precisamente no tema de Interfaces Naturais. O uso do conceito de naturalidade pode permitir que usuários com deficiência visual controlem as funções de reprodutores de vídeo realizando gestos com as mãos. Desta forma, o usuário não depende apenas da interface gráfica, opções de leitura de tela ou da tecnologia presente no reprodutor de vídeo na Web.

A área de Interfaces Baseadas em Gestos tem sido aplicada para os mais diversos usos (LEE, 2010) e está presente no cotidiano dos usuários. A área de dispositivos móveis, por exemplo, tem se beneficiado bastante da utilização de gestos em telas *touchscreen* presentes em *smartphones* para controle de jogos e aplicações, tornando-as mais atrativas ao usuário (ASHBROOK; STARNER, 2010) e (KANE; WOBROCK; LADNER, 2011). Assim, como a captura de gestos realizados sobre uma tela pode beneficiar usuários de dispositivos móveis, a captura de gestos realizados em três dimensões pode auxiliar usuários com deficiência visual na realização de comandos.

Segundo (LIU *et al.*, 2015), pesquisas com Interfaces Baseadas em Gestos começaram a se tornar mais presentes com a chegada de sensores de captura gestual mais baratos e acessíveis. Um destes sensores é o *Microsoft Kinect*⁸, que realiza a captura e o processamento de gestos realizados em três dimensões. Apesar de inicialmente ser criado apenas para o uso em jogos

⁸ <http://www.xbox.com/pt-BR/Xbox-One/accessories/kinect-for-xbox-one>

eletrônicos, pode ser aplicado para estudos e pesquisas na área acadêmica. Contudo, a literatura da área não apresenta evidências de que interfaces baseadas em gestos possam melhorar os níveis de acessibilidade durante o acesso a vídeos por usuários com deficiência visual.

1.2 Objetivo

Diante do exposto, o objetivo deste trabalho é investigar se interações gestuais podem oferecer melhorias de acessibilidade para usuários com deficiência visual quando estes acessam vídeos na Web. Para atingir tal objetivo, durante este trabalho se desenvolveu um *framework* para captura, representação e interpretação de gestos. Este *framework* foi utilizado como ferramenta para viabilizar um estudo de caso envolvendo diferentes tipos de sensores gestuais e usuários com deficiência visual.

1.3 Organização do texto

Neste capítulo foram apresentadas as considerações iniciais desta dissertação, o contexto no qual se insere, suas motivações e objetivos. No [Capítulo 2](#) são apresentados os principais conceitos estudados na literatura científica relacionados ao tema deste trabalho.

No [Capítulo 3](#) é apresentada a abordagem de desenvolvimento de um *framework* de interação gestual para acessibilidade de vídeos na Web multissensorial chamado *Gesture4All*. Neste capítulo ainda é descrito um estudo de caso conduzido para viabilizar a investigação de melhoria de acessibilidade proposta como objetivo dessa dissertação. Por fim, no [Capítulo 4](#) são discutidas as conclusões, incluindo as principais contribuições, algumas limitações, bem como as possibilidades de trabalhos futuros.

REFERENCIAL TEÓRICO

O objetivo do referencial teórico é apresentar e discutir os principais conceitos que envolvem a pesquisa, de modo que possam servir de referência para eventuais leitores que não sejam exatamente especialistas no assunto ([WAZLAWICK, 2017](#)). Além disso, este procedimento é importante, uma vez que discute os pontos de vista de diversos autores.

Nesse sentido, este capítulo apresenta, de maneira sucinta, os conceitos centrais que conduziram ao alcance dos objetivos propostos e desenvolvidos neste trabalho, proporcionando melhor compreensão dos temas abordados. Para tanto o, inicialmente são discutidos os conceitos relacionados à Computação Ubíqua, bem como o a subárea de Interfaces Baseadas em Gestos, tópica de importância para este trabalho. Em seguida, são discutidos os principais conceitos de acessibilidade e acessibilidade de vídeos na Web. E, por fim, são apresentados os trabalhos relacionados, os quais são a base para construção da solução proposta neste trabalho.

2.1 Computação Ubíqua e Interfaces Naturais

Segundo [Weiser \(1993\)](#), o conceito de ubiquidade em computação é aplicado quando a interação do usuário com o computador ultrapassa o uso de interfaces como monitor, teclado e *mouse*, por interfaces invisíveis aos olhos e incorporadas ao ambiente, ou objetos de seu cotidiano. Weiser descreve um mundo onde a computação poderia estar presente em carros, roupas, salas de reuniões, casas e onde mais a imaginação poderia querer. Ao aplicar recursos computacionais em objetos ou lugares, não é alterado seu propósito original mas sim expandidas as possibilidades de uso do mesmo ([WEISER, 1993](#))

Com a evolução de *hardwares* e *softwares*, principalmente dentro da área de Sistemas Embarcados e Computação Móvel, muitos avanços foram possíveis desde 1993 até os dias atuais. Maior acesso a sensores inteligentes, com tamanho reduzido e móveis, assim como *softwares* que integram dados recebidos deste sensores e realizam seu processamento, possibilitam cada

vez mais desenvolver e disponibilizar tecnologias ubíquas ao usuário final.

Como exemplo dessas tecnologias, o sistema *DUMMBO* (*Dynamic, Ubiquitous, Mobile Meeting Board*) (BROTHERTON; ABOWD; TRUONG, 1998) da *Georgia Institute of Technology* ilustrado na Figura 1, foi desenvolvido para uso em salas de reuniões e possibilita anotações e desenhos em uma lousa eletrônica que registra as interações juntamente com o áudio das conversas realizadas. Quando solicitado o sistema disponibilizados na forma de um documento multimídia essas interações.

Figura 1 – Na direita a lousa eletrônica e a esquerda o documento multimídia.

Fonte: Adaptada de Brotherton, Abowd e Truong (1998).

Neste exemplo, nenhum participante da reunião interagiu com uma interface convencional como monitor ou teclado para realizar os desenhos ou realizar a gravação do áudio. A captura das informações foram realizada por meio de sensores presentes na sala e o processamento computacional ocorreu sem a interferência direta do usuário, o computador estava invisível ao usuário mas totalmente presente nos bastidores processando toda a interação.

Para o surgimento de novas aplicações a computação ubíqua possui subáreas que abrangem outros aspectos e ampliam as possibilidades de desenvolvimento, de acordo com Abowd e Mynatt (2000) são: **Captura e Acesso (C&A)**, **Computação Ciente de Contexto** e **Interfaces Naturais**.

Para que exista **Captura e Acesso (C&A)** Truong, Abowd e Brotherton (2001) definem que aplicações ubíquas devem preservar os dados capturados de alguma interação em tempo real e fornecer estes dados a qualquer momento em que a aplicação necessitar. Segundo Abowd e Mynatt (2000) o acesso a informações derivadas de interações devem ser universais e flexíveis. Isso implica que a captura seja o máximo possível automatizada e devidamente organizada.

Aplicações de C&A possuem interesses em capturar e processar dados gerados durante a vida humana provendo assim possibilidade de melhorias na qualidade de vida. Como exemplo o projeto *SymDetector* do *Department of Computer Science and Engineering of Pennsylvania State University* que por meio de um microfone e de um *smartphone* do próprio usuário, monitora o som de sua respiração para detectar possíveis sintomas de doenças pulmonares (SUN *et al.*, 2015).

Outro estudo que aplica C&A foi realizado pela *Graduate School of Information Science and Technology of Osaka University* e a *Osaka University Dental Hospital* resultando no estudo publicado *Evaluating Tooth Brushing Performance with Smartphone Sound Data* (KORPELA *et al.*, 2015) que também utiliza o monitoramento de sons por meio de um *smartphone*. Seu objetivo é medir o desempenho da escovação dentária atribuindo pontuações como ilustrado na Figura 2.

Figura 2 – Usuário durante escovação e sistema de pontuação.

Fonte: Adaptada de Korpela *et al.* (2015).

A computação ubíqua também pode ser **Ciente de Contexto**, permitindo que sistemas modifiquem automaticamente seu comportamento com base em informações obtidas a partir de ambientes físicos e computacionais (ABOWD; MYNATT, 2000). A computação ciente de contexto faz uso dos dados capturados e processados, por meio de um ou mais contextos, como forma de prover informações importantes para a realização de uma tarefa (DEY, 2001). Um dos projetos pioneiros na utilização do conceito de ciência de contexto foi o trabalho de Want *et al.* (1992), chamado *Active Badge*. O trabalho consistia em localizar funcionários dentro de um escritório, para tanto, cada funcionário recebeu um crachá que emitia um sinal infravermelho único que era capturado por sensores espalhados pelo escritório, permitindo assim a rápida localização de um funcionário.

Com o aumento de tipos de sensores diferentes disponíveis no mercado, o baixo custo e o aumento do poder computacional, projetos voltados à ciência de contexto foram direcionados aos mais variados tipos de contextos e resoluções de problemas. Como exemplo, o trabalho de Huang, Yang e Newman (2015), chamado *The Potential and Challenges of Inferring Thermal Comfort at Home Using Commodity Sensors* tem como objetivo encontrar o melhor modo de gerenciar a temperatura de um ambiente para garantir o bem estar do morador, para tanto o sistema recebe diversas informações espalhados pela casa durante diversos dias e ter a ciência do contexto do ambiente, como restrições ou preferência e temperatura externa em diferentes horários do dia. A pesquisa foi realizada em sete casas residenciais convencionais, gerenciando a temperatura visando um conforto a todos os residentes durante o tempo que permaneciam no ambiente.

Além de compreender a **Captura e Acesso** e a **Computação Ciente de Contexto**, a

computação ubíqua também compreende a subárea de **Interfaces Naturais**. Por este ser o tópico de interesse do projeto, ele será abordado com maior profundidade.

A área de Interfaces Naturais, também conhecida como *Natural User Interface* (NUI), permite tornar uma interface mais agradável e fácil de ser compreendida pelo usuário. Segundo [Abowd e Mynatt \(2000\)](#), ao utilizar interfaces naturais como forma de controle é possível aumentar a usabilidade e melhorar a capacidade de aprendizado de um usuário ao interagir com um sistema.

A palavra natural possui diferentes significados dependendo da área ou contexto que é utilizada. Em computação pode-se entender o conceito de naturalidade como sendo uma qualidade da forma com que um usuário utiliza uma interface. Uma interface natural é aquela que possui um alto grau de intuição em sua execução e é eficaz em ser quase imperceptível quando o usuário realiza alguma tarefa ([LEE, 2010](#)).

As interações entre homem e computador sofreram muitas mudanças ao longo do tempo. Desde linhas de comando com entrada de dados apenas por teclados, passando pela inserção do *mouse* como interface para interações gráficas, o homem busca meios de melhorar a interação tornando mais eficiente e natural.

Nos anos 90, Steve Mann foi um dos primeiros pesquisadores a propor novos modos de interações naturais por meio do projeto *EyeTap* ([MANN; FUNG; MONCRIEFF, 1999](#)) ilustrado na [Figura 3](#). Utilizando um óculos como interface, o sistema adicionava informações no *display* do óculos ao usuário. Outros exemplos pioneiros incluem o trabalho de [Schilit, Adams e Want \(1992\)](#), que utiliza a escrita manual em lousas como forma de interação, por meio de câmera o sistema captura a escrita realizada de forma natural pelo usuário e transformando o texto em um documento eletrônico. E trabalho de [Ishii e Ullmer \(1997\)](#), que utiliza a manipulação física de objetos reais para controle de artefatos eletrônicos, criando o conceito de interfaces tangíveis.

Em 2007, Steve voltou a ser referência na área de interfaces naturais com o projeto chamado *Hydraulophone - Natural interfaces for musical expression* ilustrado na [Figura 3](#). O projeto consiste em um teclado musical onde o som é produzido com interação natural por meio da água ([MANN, 2007](#)), um projeto que incentivou outros pesquisadores a buscar novos meios de criar interfaces.

Uma interface deve ser intuitiva o quanto possível, para que o usuário interaja naturalmente com uma tecnologia com o mínimo de instruções prévias consiga atingir seu objetivo tornando-se um usuário experiente. Essa ideia foi bem retratada por [Norman \(1990\)](#) em seu livro *Why interfaces don't work*, segundo ele:

"O problema real com as interfaces é que elas são uma interface. Interfaces ficam no caminho. Eu não quero concentrar minhas energias em uma interface. Quero me concentrar no trabalho ... Eu não quero pensar por mim mesmo em como usar um computador, eu quero pensar por mim mesmo em como fazer o meu trabalho."

Figura 3 – Na direita Steve utilizando o *EyeTap* e na esquerda o *Hydraulophone*.

Fonte: Adaptada de Mann (2007), Mann, Fung e Moncrieff (1999).

Em alguns casos, a interface se torna tão complicada que ao invés de ser um facilitador para se alcançar um objetivo ela se torna o maior problema da aplicação em si. O usuário além de não conseguir realizar suas tarefas fica desconfortável e tem uma opinião de que a aplicação não atende os objetivos que deseja.

Existe também o fato de que muitos usuários não possuem muitas experiências anteriores em computação, quanto mais natural for uma interface, mais fácil será para um usuário inexperiente entender seu funcionamento; criar interfaces mais naturais é um desafio muito importante da computação (LEE, 2010).

Existem diversas tarefas executadas de modo natural que podem ser utilizadas para interação com sistemas computacionais. Tarefas como: fala, gestos, escrita manual e expressão corporal são alguns exemplos de modos naturais que uma interface pode utilizar de meios de complementar ou substituir elementos do paradigma de interfaces gráficas. Ao analisar a literatura é possível identificar os seguintes tipos de aplicações em interfaces naturais:

1. **Reconhecimento de toque (*Touch Recognition*):** Interfaces que reconhecem o toque permitem que usuários selecionem diretamente, geralmente utilizando os dedos da mão, uma representação gráfica na tela ao qual desejam interagir. Diferente do *mouse*, que utiliza o cursor para interagir com representações gráficas, o toque permite diferentes meios de interação, como utilizar dois dedos para ampliar um objeto ou múltiplos toques para iniciar diferentes ações, ampliando os modos de interação de uma interface a tornando mais interativa.

A origem do conceito data de um artigo publicado por Johnson (1965) com título "*Touch display—a novel input/output device for computers*", onde foram descritos os primeiros experimentos com toque em tela. Exemplos relevantes de pesquisa nessa área incluem: o trabalho realizado por Albinsson e Zhai (2003) que explora técnicas de melhoria de

precisão e captura do toque; o trabalho realizado por [Yairi et al. \(2011\)](#) que apresenta um método de interação por toque para usuários com deficiência visual utilizando o som como *feedback*; e o trabalho realizado por [Harrison, Tan e Morris \(2010\)](#) que explora a possibilidade de utilizar a pele humana como interface sensível ao toque.

2. **Reconhecimento de fala (*Speech recognition*):** Interfaces com reconhecimento de fala permitem a usuários interagir com aplicações por meio da análise das palavras e frases ditas em voz alta. Segundo o trabalho de [Vertanen e Kristensson \(2008\)](#), que discute a dificuldades de detecção de erros na captura da fala, um dos métodos usados para reconhecimento utiliza redes neurais que dão pesos a cada palavra de uma frase e interpreta semanticamente o objetivo do usuário com a análise dos valores dos pesos.

O primeiro estudo de reconhecimento de fala documentado surgiu do trabalho de [Davis, Biddulph e Balashek \(1959\)](#), que conseguia apenas reconhecer dígitos numéricos ditos por um único usuário. Já o primeiro sistema comercial utilizando a voz como interação foi chamado de "*Dragon*" ([BAKER, 1975](#)), que utilizava o Modelo Oculto de Markov para o reconhecimento das palavras e frases. Atualmente esse tipo de interface é encontrado facilmente em *smartphones* e *tablets*, dá-se destaque aos aplicativos como: a *Siri* ([LOVATO; PIPER, 2015](#)), presente no sistema operacional IOS da *Apple* e o *Google Now* ([DIAO et al., 2014](#)) presente em dispositivos com o sistema operacional *Android* da *Google*.

3. **Rastreamento dos olhos (*Gaze-tracking*):** Interfaces que capturam os movimentos dos olhos do usuário, podem melhorar significativamente a usabilidade de sistemas computacionais. Após determinar para onde o usuário está olhando, e analisar o trajeto do olho, o sistema pode interferir na tarefa atual e em seu contexto levando a prever um conjunto de possíveis ações para a melhoria da interação ([MORIMOTO; AMIR; FLICKNER, 2002](#)). Segundo [Yang et al. \(2002\)](#), os primeiros trabalhos nesse campo começaram a ser publicados na década de 1990. Desde então, o rastreamento dos olhos se tornou mais preciso, com menor custo computacional e já pode ser encontrado em aplicações voltadas ao usuário final. Para captura dos movimentos geralmente utiliza-se câmeras ou sensores próximos aos olhos, que enviam as imagens em tempo de execução para um sistema que efetua técnicas de processamento de imagens.

4. **Leitura de sinais cerebrais (*Brain-machine*):** outro modo de interação natural possível com uma interface é a utilização de sinais do cérebro para a execução de comandos. Como exemplo, a movimentação de um cursor em um ambiente gráfico pode ser realizada por meio da captura e processamento dos sinais do cérebro de um usuário. ([WANG, 2016](#)).

O uso dos sinais cerebrais pode permitir que uma pessoa com paralisia possa navegar pela internet ou escrever um livro. Um dos maiores desafios desse campo é o desenvolvimento de dispositivos que possam capturar com precisão as variações dos sinais e algoritmos que consigam processar e realizar a interpretação do pensamento do usuário. Como exemplo, o

trabalho de [Evain et al. \(2016\)](#) realizou um teste com usuários que deveriam olhar fixamente para objetos uma tela, por alguns segundos, vestindo um Electroencephalogram (EEG) *headset*¹ em suas cabeças. Os resultados demonstraram a taxa de sucesso e fracasso que o sistema obteve ao saber exatamente e qual momento e para qual objeto o usuário esteve olhando. Outro aspecto interessante deste trabalho é a investigação sobre a importância do *feedback* em sistemas de interfaces naturais, segundo o trabalho parte do sucesso que uma interface natural obtém é devido ao meio estrutural que o *feedback* possui.

5. **Reconhecimento de gestos (Gesture recognition):** com novas pesquisas sendo desenvolvidas buscando meios de tornar uma interação mais natural e a possibilidade de utilizar outras interfaces diferentes das convencionais, os gestos produzidos pelo corpo humano começaram a ser considerado como uma possível forma de interação natural. Como exemplo de uso do conceito de Interfaces Naturais, o aplicativo *RunBuddy*, desenvolvido por estudantes da *Michigan State University* e *College of William and Mary*, monitora os gestos do usuário durante uma corrida. Os dados capturados são utilizados para indicar a frequência de passos e respiração, assim é possível medir a qualidade da corrida executada e sugerir melhorias ([HAO; XING; ZHOU, 2015](#)).

Seria adequado para qualquer sistema que deseje ser natural incorporar em sua concepção os gestos como forma de interação. Essa proposta já é conhecida e faz parte da subárea de Interfaces Naturais. Interfaces baseadas em gestos podem ser uma ótima alternativa para diversas aplicações que desejam ser mais agradáveis ao usuário. Por este ser um tópico de interesse do projeto ele será abordado com maior profundidade na [Seção 2.2](#).

2.2 Interfaces Baseadas em Gestos

Interface Baseada em Gestos é a interface que está pré-configurada ou poderá ser personalizada para receber interações gestuais do usuário. Para a interpretação destas interações gestuais, realiza o processamento das informações capturadas por meio de sensores e dispositivos.

Gestos são utilizados dentro da linguagem humana como uma forma de comunicação não verbal durante toda a vida. Mesmo antes de aprender a fala, o ser humano utiliza a comunicação por gestos de forma natural e espontânea, incluindo expressões faciais, movimentos realizados com as mãos e realizados utilizando todo o corpo ([GHIROTTI; MORIMOTO, 2010](#)).

Segundo [Moro et al. \(2013\)](#), gestos associados a comunicação humana recebem o nome de gestos semióticos e auxiliam na construção do modelo mental que o locutor deseja passar ao ouvinte. Na área da computação, interfaces puramente baseadas em gestos utilizam, principalmente, os gestos semióticos para o desenvolvimento de vocabulários de interação, estabelecendo uma relação sobre o que sistema espera receber como entrada de gestos para

¹ sensor de ondas cerebrais por eletroencefalograma

realizar determinada ação. Além do uso de gestos semióticos, outra subcategoria usada pela computação são os gestos naturais, que descartam a necessidade de aprendizado e treinamento.

Um número considerável de sistemas atualmente tem explorado o uso de gestos como interação e tem demonstrado bons resultados. Muitos desses sistemas possuem gestos pré-configurados que devem ser repetidos pelo usuário por meio de movimentos corporais e posicionamento das mãos (LEE, 2010).

O sistema verifica se o movimento capturado é semelhante a sua pré-configuração, após confirmação o sistema aciona a ação associada aquele gesto. Desde a captura do movimento até a execução da ação, diversas etapas devem ser consideradas para que a interface seja eficiente e totalmente funcional.

Devido ao potencial que sistemas baseados em gestos podem oferecer por meio de interações naturais com o usuário, muitas pesquisas podem ser realizadas para a melhoria de sistemas já existentes e o desenvolvimento de novos sistemas. Devido à variedade de dispositivos e técnicas de captura de gestos é possível abranger diversos requisitos que uma aplicação necessite (ABOWD; MYNATT, 2000).

Segundo Ghirotti e Morimoto (2010), a estrutura de uma Interface Baseada em Gestos pode ser classificada em três etapas como ilustra a Figura 4:

1. **Segmentação:** realiza a captura dos dados brutos coletados por sensores e dispositivos, em um dado momento ou com o passar do tempo.
2. **Reconhecimento:** processa as informações obtidas na etapa de segmentação e identifica qual foi o gesto capturado.
3. **Interpretação:** realiza a comparação do gesto reconhecido na etapa anterior com uma base de dados gestual. A comparação permite identificar qual ação deverá ser executada pelo sistema, uma vez que esta ação está associada a um gesto na base de dados.

Como a estrutura de uma Interface Baseada em Gestos é de grande interesse deste trabalho, cada parte será abordada com maior profundidade da seguinte forma: Subseção 2.2.1 Segmentação, Subseção 2.2.2 Reconhecimento e Subseção 2.2.3 Interpretação.

2.2.1 Segmentação

Dois modos de captura de gestos são geralmente empregados na maioria de estudos em computação ubíqua.

A) Por meio de sensores vestíveis ou próximos do usuário.

B) Por meio de processamento de imagens capturadas durante a interação.

Figura 4 – Exemplo da estrutura de uma Interface Baseada em Gestos

Fonte: Adaptada de [BussinesInsider \(2015\)](#).

A seguir, são descritos os dois modos de captura de gestos, detalhando-se as técnicas e metodologias que usualmente têm sido adotadas.

A) Por meio de sensores:

Dois sensores se destacam quando existe a captura precisa de gestos realizados por usuários, são estes sensores o acelerômetro e o giroscópio. Pode-se entender suas diferenças da seguinte forma:

- **Acelerômetro:** sensor de movimento que captura a variação na aceleração que um corpo exerce no espaço de forma tridimensional, ou seja, nos eixos X, Y e Z. ([BONOMI et al., 2009](#)). Por meio desses valores, é possível determinar o deslocamento de um corpo em movimento.
- **Giroscópio:** sensor de movimento que consegue capturar a mudança de rotação que um corpo exerce no espaço de forma tridimensional, ou seja, nos eixos X, Y e Z ([O'DONOVAN; AYER, 2011](#)). Por meio desses valores, é possível determinar a rotação e a mudança de direção de um corpo em movimento.

A exemplo do uso destes sensores, a dissertação de ([CARVALHO, 2013](#)) propõe o uso do *Wii Remote* para captura dos gestos realizados com as mãos com o objetivo da interação com objetos virtuais. Desenvolvido pela *Nintendo*[®] como interface de controle para o *video-game Nintendo*[®] *WII*, .

Modo de captura do *Wii Remote*: O *Wii Remote* utiliza três acelerômetros, um giroscópio para capturar o posicionamento das mãos do usuário como ilustrado na [Figura 5](#). A captura e

Figura 5 – Usuário realizando um gesto tridimensional com o *Wii Remote*.

Fonte: Adaptada de [Network \(2015\)](#).

o envio dos dados brutos coletados é enviada em tempo real, o controle envia as informações gestuais por uma rede *bluetooth*, a aplicação recebe os dados brutos coletados durante a interação e realiza o processo de tratamentos dos dados para interpretar qual movimento foi realizado.

Outro dispositivo que utiliza sensores de acelerômetro e giroscópio que tem como foco a máxima precisão de captura é o *Myo Armband*, desenvolvida pela *Thalmict*® é uma braçadeira com sensores que ao ser vestida pelo usuário captura o movimento muscular do braço.

O *Myo* consegue identificar o posicionamento gestual das mãos do usuário detectando quais músculos foram estimulados como ilustra a [Figura 6](#), o sistema é capaz de diferenciar se a mão do usuário está aberta ou fechada como ilustrado na [Figura 6](#) apenas analisando a forma que o músculo está disposta ([MULLING; SATHIYANARAYANAN, 2015](#)).

Figura 6 – *Myo Armband*.

Fonte: Adaptada de [Thalmic \(2015\)](#).

A exemplo do uso deste dispositivo vestível, o trabalho ([MCCULLOUGH et al., 2015](#)) explora o uso da identificação dos padrões musculares que usuário possuem a executar cada gestos. Segundo o estudo, o padrão muscular do braço do usuário é alterado a cada novo gesto executado, logo pode ser usado para iniciar uma ação quando se compara o padrão muscular

atual com uma base de dados gestual.

Modo de captura do Myo: Além da análise muscular, o *Myo*, possui acelerômetro e giroscópio para aumentar a abrangência de captura gestual. Os dados brutos são gerados com informações de posicionamento X, Y e Z. Os dados então são enviados por uma rede *bluetooth* ao computador. Após o envio é possível aplicar algoritmos que interpretam os dados a fim de encontrar padrões gestuais.

B) Por meio de processamento de imagens

Leap Motion: Desenvolvido pela *Leap Motion Inc.* é um sensor que captura os gestos das mãos do usuário estando elas a uma certa distância do dispositivo como ilustrado na [Figura 7](#). Sua principal característica é a precisão da captura do movimento, podendo diferenciar em milímetros o movimento gestual. Possui aplicações diversas como em jogos eletrônicos e reconhecimento de linguagem de sinais que segundo [Potter, Araullo e Carter \(2013\)](#) possui grande potencial porém ainda não compreende a totalidade do vocabulário da linguagem de sinais.

Figura 7 – Usuário realizando um gesto tridimensional com o *Leap Motion*.

Fonte: Adaptada de [LeapMotion \(2015\)](#).

Modo de captura do *Leap Motion*: Segundo ([MOSER; TSCHELIGI, 2015](#)), o dispositivo possui três LEDs infravermelhos e duas câmeras internas. Ao usuário colocar sua mão próxima ao dispositivo, a luz infravermelha indica o contorno da mão, as câmeras são desenvolvidas de modo a apenas capturar a luz infravermelha, que é facilmente reconhecida por algoritmos de detecção de bordas.

Kinect: Desenvolvido pela *Microsoft®* é um dos sensores para captura de gestos e movimentos mais conhecido e de fácil uso como ilustrado na [Figura 8](#). Desenvolvido para uso em conjunto com o *video-game Microsoft Xboxtextsuperscript® 360* demonstrou potencial de uso em diversas áreas devido a possibilidade de acesso aos dados capturados por meio de um computador, por sua precisão e resposta rápida durante a captura de dados ([CARTER et al., 2015](#)).

Modo de captura do *Kinect*: Permite de forma eficiente a captura de todo o corpo de vários usuário simultaneamente e ainda acompanhar o posicionamento do olhos do usuário como

Figura 8 – Usuário realizando um gesto tridimensional com o *Kinect*.

Fonte: Adaptada de [Kinect \(2015\)](#).

integração da captura de movimentos. Isso é possível devido à combinação de três elementos: uma **câmera**, um **sensor de profundidade** e um **microfone multi-array**.

A câmera detecta as cores vermelha, verde e azul da cena, possui resolução de 640x480 e 30 frames por segundo. O sensor de profundidade utiliza luz infravermelha para captar as bordas e gerar um modelo 3D, o sensor também mede a distância de cada ponto de objetos da cena. Para captura do som ele utiliza quatro microfones que conseguem isolar a voz de ruídos externos.

Asus Xtion: Desenvolvida pela *Asus* para ser utilizado na captura de movimentos, como ilustrado na [Figura 9](#), para a plataforma PC. Por meio de sensores infravermelho e sistemas RGB (*Red, Green, Blue*) para detecção de cores. Possui ferramentas de desenvolvimento que permitem a implementação de diversas aplicações.

Figura 9 – Usuário realizando um gesto tridimensional com o *Asus Xtion*.

Fonte: Adaptada de [Asus \(2015\)](#).

Câmeras diversas: Devido a algoritmos de processamento de imagens é possível transformar qualquer câmera em um sensor que capture e reconheça gestos como ilustrado na [Figura 10](#).

Figura 10 – Usuário controlando uma aplicação por meio de sua *Webcam*

Fonte: Adaptada de Gizmag (2015).

O que diferencia uma aplicação de outra será basicamente a resolução da câmera e se o custo computacional do processamento solicitado pode ser atendido. O conjunto desses dois elementos permite interações como controle de robôs por gestos (NGUYEN *et al.*, 2011) ou reconhecimento de gestos para vídeo *streaming* (SRIJEYANTHAN; JAYALATH, 2011).

2.2.2 Reconhecimento

Em 1986 Nespoulous e Lecours (1986) propuseram uma topologia de identificação gestual que resultou na classificação de três categorias básicas de gestos. Apesar de outros pesquisadores citarem essas classes com outras nomenclaturas, seus conceitos permanecem os mesmos, são elas:

1. Gestos Miméticos ou Icônicos: São gestos que imitam um comportamento ou formato de um objeto. Por exemplo, quando uma pessoa deseja indicar o tamanho de um objeto ou se referir a um animal imitando seu movimento mais característico.
2. Gestos Dêiticos: São gestos utilizados para indicar um objeto ou ação, para ser interpretado depende do contexto ao qual ele é utilizado. Por exemplo, quando uma pessoa aponta para um objeto que deseja pegar ou realiza o movimento de abrir os braços quando deseja abraçar.
3. Gestos Arbitrários: São gestos que não tem a intenção de imitar algum comportamento ou indicar algum objeto, são aprendidos e utilizados para uma situação específica. Por exemplo, a saudação utilizada no exército ou um aperto de mão.

Aplicando ao contexto da computação ubíqua, os gestos mais utilizados em aplicações computacionais são os gestos arbitrários devido serem facilmente distinguido de outros gestos naturais do ser humano (LEE, 2010). Como o sistema deverá reconhecer gestos e transforma-los em comandos computacionais é necessária a distinção de gestos para não existir conflito de interpretações gestuais (LEE, 2010). Se uma aplicação está configurada para reconhecer um

gesto dêitico quando o usuário apontar para um local, por exemplo, poderá haver momentos em que o usuário deseja apontar para um lugar, mas não necessariamente deseja que o sistema realize a ação naquele momento.

Para que não haja conflito de interpretações gestuais é comum que o usuário precise aprender primeiramente qual gesto está associado a um comando ou configurar a aplicação indicando qual gesto deverá estar associado a uma função. A aplicação pode criar um tutorial de gestos para o primeiro uso ou criar representações gráficas como na [Figura 11](#) simbolizando qual movimento o usuário deverá realizar.

Figura 11 – Exemplo de representação gráfica do gesto a ser realizado.

Fonte: Adaptada de [Designbeep \(2015\)](#).

Casos nos quais gestos diferentes e específicos são necessários a interação pode ser desgastante e tornar a aplicação complicada, perdendo o objetivo principal de ser natural e simples ao usuário ([LEE, 2010](#)). Uma solução para o problema de submeter o usuário a aprendizado de gestos antes do uso da aplicação sem o desgaste, pode estar na utilização de diferentes tipos gestuais em conjunto para facilitar a interação. Utilizar parte de gestos que são simples e naturais ao usuário, porém adaptados para que sejam facilmente diferenciados pela aplicação, podem trazer muitos benefícios.

Além dos conceitos sobre o tipo de gesto que a aplicação deve reconhecer, outro fator é indispensável na construção de sistemas baseados em gestos, o desenvolvimento da interface deve considerar qual deverá ser abrangência dimensional do gesto ([MITRA; ACHARYA, 2007](#)).

Dimensões podem ser representadas por informações de comprimento, largura e altura de um objeto no espaço, no contexto de reconhecimento gesticular existem dois tipos: **gestos bidimensionais** e **gestos tridimensional** detalhados a seguir:

A) Gestos bidimensionais: são interações executadas sobre uma superfície física, geralmente plana, onde os dedos do usuário executam movimentos de forma bidimensional, ou seja, utilizando informações de altura e largura ([HOGGAN et al., 2013](#)).

Como exemplo, essa forma de interação é muito presente na computação móvel a partir do design dos *smartphones* e *tablets* para interações com telas *touch screen*. Como ilustrado na

Figura 12, uma tela *touch screen* pode reconhecer movimentos bidimensionais como a quantidade de dedos sobre a tela e o movimento que o(s) dedo(s) executam.

Figura 12 – Interação do usuário com um monitor e *smartphone*.

Fonte: Adaptada de [Tecnundo \(2015\)](#).

A grande maioria dos *smartphones* e *tablets* atuais abandonaram seus teclados físicos passando a incorporar o teclado como uma interface lógica interna juntamente com comandos gestuais de toque para entrada de dados. Cada vez mais são empregados estudos para melhorar o design de gestos em dispositivos móveis para uma interação natural e entender melhor como proporcionar uma experiência melhor para todos os tipos de usuário ([ASHBROOK; STARNER, 2010](#)) e ([KANE; WOBBROCK; LADNER, 2011](#)).

Não somente na computação móvel são empregados conceitos de gestos bidimensionais, qualquer tecnologia que possui telas sensíveis ao toque ou que consiga reconhecer o toque por meio de uma superfície também utilizam gestos bidimensionais. Monitores e televisores são exemplos de interfaces que podem utilizar esse conceito assim com painéis de carro, brinquedos ([REVELLE et al., 2005](#)), móveis e qualquer superfície sensível pode ser utilizadas como interface baseada em gestos bidimensionais ([GRIJNCU; NACENTA; KRISTENSSON, 2014](#)).

B) Gestos tridimensionais: abrangem movimentos executados no espaço contendo informações de altura, largura e profundidade ([MANDERS et al., 2008](#)) e ([STURMAN; ZELTZER, 1993](#)). Essa característica denota um nível maior de liberdade e complexidade devido a não existir uma superfície de contato, para reconhecer os gestos muitos dispositivos podem ser empregados para captura desses movimentos tridimensionais.

A escolha da forma de captura define o modo de interação e pode ser crucial para obtenção de dados precisos. Uma vez que o *feedback* geralmente ocorre em tempo real, o usuário realiza o gesto e espera a resposta da aplicação durante a execução como ilustrado na [Figura 13](#).

Dependendo do objetivo que a interface deseja alcançar, o desenvolvimento do sistema pode restringir o uso dos gestos para captura em duas dimensões ou em três dimensões. Como exemplo, os *smartphone*, em sua grande maioria, restringe o uso para a captura dos gestos realizados sobre a superfície da tela *touch screen* de forma bidimensional, ou seja, apenas gestos realizados enquanto existe contato na tela.

Figura 13 – Usuário realizando um gesto tridimensional e visualizando o *feedback* da ação.

Fonte: Adaptada de [Deskeng \(2015\)](#).

Em alguns casos, uma mesma interface pode ser dinâmica e possibilitar interações bidimensionais e tridimensionais, essas características devem levar em consideração o modo de captura dos gestos ([LEE, 2010](#)). Utilizando o exemplo de *smartphone*, citado anteriormente, aplicações podem utilizar a câmera presente no dispositivo móvel para também reconhecer gestos tridimensionais, dessa forma uma mesma interface pode receber informações da interação bidimensional realizada com toques na tela assim como por meio de gestos tridimensionais realizados na área de abrangência da câmera.

Características como a quantidade de usuários simultâneos e a abrangência de captura de movimento do usuário (todo o corpo ou apenas as mãos) devem definir a escolha do modo de captura. Usuários geralmente possuem facilidade de aprendizagem de vocabulários gestuais utilizando apenas a observação, isso torna a adaptação da interfaces mais natural e agradável ao usuário, mesmo sem ele antes possuir experiências anteriores com esse tipo de interface ([GHIROTTI; MORIMOTO, 2010](#)).

2.2.3 Interpretação

Segundo o trabalho de [Schlomer et al. \(2008\)](#), uma das técnicas possíveis para a interpretação de gestos, é a utilização de três componentes: um **quantizador** para realizar a segmentação, um **modelo** para atuar como reconhecimento e um **classificador** para atuar como o interpretador.

Quantizador: por causa dos dispositivos produzirem dados de forma vetorial e em grande escala, uma técnica de **quantização** pode ser aplicada para realizar o agrupamento dos dados. O agrupamento permite obter a média dos dados brutos gestuais e assim conseguir maior precisão e rapidez no reconhecimento gestual. Sensores como acelerômetro e giroscópios podem enviar dados de até milésimos de segundos com informações espaciais como X, Y e Z, como esses valores são muito próximos à média é necessária para obter um valor aceitável para que o reconhecimento. Nessa técnica podem ser aplicados algoritmos como o **K-means**. O algoritmo *K-means* foi utilizado pela primeira vez no trabalho de [Macqueen \(1967\)](#), e consiste em agrupar grandes quantidades de dados com base em singularidades, o K, de *K-means*, é a quantidade de

centroídes (pontos centrais dos grupos) que serão criados e ajudará a encontrar a similaridade dos dados gestuais.

Modelo: um dos possíveis modelos que podem ser utilizados após os agrupamentos dos dados é o *Hidden Markov Models* (HMM) ou Modelo Oculto de Markov. Foi descrito pela primeira no final dos anos 60 no trabalho de Baum e Petrie (1966), e tem sido utilizado para o reconhecimento de outras interfaces naturais como: voz (LEE *et al.*, 1990), (RABINER; WILPON; SOONG, 1989) e escrita manual (KUNDU; HE; BAHL, 1988), (CHEN; KUNDU; ZHOU, 1994). O HMM utiliza processos escondidos que consistem em conjuntos de estados conectados por transições com probabilidades (autômato finito), enquanto que os processos observáveis (não escondidos) consistem de um conjunto de saídas ou observações, cada qual podem ser emitido por cada estado de acordo com alguma saída da função de densidade de probabilidade (OLIVEIRA; MORITA, 2010).

Classificador: segundo Schlomer *et al.* (2008), ao usuário executar diversos gestos em sequência, o quantizador agrupa os dados que possuem similaridades (como gestos que indicam movimentação para cima) e aplica o HMM para comparar o dado agrupado com outros grupos previamente cadastrados. Por fim a última etapa é a classificação para calcular a probabilidade que um gesto executado (processador pelo quantizador) é o mesmo que o gesto cadastrado no sistema (HMM). Um dos possíveis algoritmos para essa etapa é a classificação Bayesiana, que calcula a probabilidade que uma amostra desconhecida (novo gesto) pertença a cada uma das classes possíveis (Classes criadas pelo HMM), ou seja, predizer a classe mais provável. O sistema somente considera um gesto como reconhecido se a probabilidade é bem próxima de 100%.

2.3 Acessibilidade de vídeos na Web

Acessibilidade na Web refere-se a permitir a percepção, o entendimento, a navegação e a interação com a Web, por todas as pessoas, que livres de barreiras (ou com barreiras reduzidas), ainda que contando com recursos de Tecnologia Assistida², também possam utilizar ou contribuir com a geração de conteúdo (Web 2.0) e desenvolvimento de sistemas Web (GRILLO; FORTES, 2014). Portanto, em termos gerais, a definição de acessibilidade na Web visa garantir o uso da Web, principalmente, por pessoas com deficiência.

Acessibilidade é um direito previsto por lei e aplicado a todo território nacional. Segundo a Lei nº 10.098 de 19 de dezembro de 2000, artigo 2º parágrafo I (FEDERAL, 2000) acessibilidade é a "possibilidade e condição de alcance para utilização, com segurança e autonomia, de espaços, mobiliários, equipamentos urbanos, edificações, transportes, informação e comunicação, inclusive seus sistemas e tecnologias, bem como de outros serviços e instalações abertos ao

² Termo que se refere a recursos ou serviços que buscam ampliar as habilidades de pessoas com deficiência

público, de uso público ou privados de uso coletivo, tanto na zona urbana como na rural, por pessoa com deficiência ou com mobilidade reduzida".

Seguindo as definições presentes na Lei nº 10.098 foram aprovados em 2004 no Brasil o Decreto nº 5.296, que determinou prazos para tornar acessível todos os portais públicos, de interesse público ou que possuía alguma ligação com órgãos públicos. Para cumprimento deste decreto foi criado o Modelo de Acessibilidade Brasileiro (eMAG), elaborado pelo Departamento de Governo Eletrônico, analisando as diretrizes de acessibilidade internacionais (EMAG, 2005).

As diretrizes internacionais de acessibilidade são definidas pela Web Content Accessibility Guidelines (WCAG), elaborada como resultado da Web Accessibility Initiative (WAI). A WAI é uma organização mantida pelo World Wide Web Consortium (W3C)³ que tem como objetivo definir princípios e regras de interfaces e desenvolvimento de sistemas que sejam acessíveis a pessoas com necessidade especiais (W3C, 2012)

Apesar de existirem leis e diretrizes internacionais para acessibilidade ainda muitos desafios para pessoas com deficiência visual ao acessarem a Web permanecem. Esses desafios estão ligados principalmente com relação à usabilidade que um site pode oferecer, usuários devem conseguir perceber, entender, navegar interagir e contribuir com a Web (FREIRE; LARA; FORTES, 2013).

Existem *softwares* que permitem que usuários com deficiência visual naveguem entre conteúdos de vídeos na Web de forma acessível, os mais utilizados para esse objetivo são os leitores de tela que como seu nome sugere faz a leitura audível de textos que estão na tela do usuário. Leitores de tela realizam a leitura selecionando textos presentes na interface de forma sequencial, geralmente começando no canto superior esquerdo e avançando por meio de comandos do teclado de forma horizontal até o canto inferior direito (BORODIN *et al.*, 2010). Apesar de muito úteis, leitores de tela dependem da forma em que o reprodutor de vídeo está desenvolvido, a utilização de imagens como menus sem descrição, por exemplo, faz com que o leitor de tela não consiga identificar e informar ao usuário de forma correta a opção em seleção. Assim, o usuário não consegue encontrar o conteúdo que deseja ou demora muito tempo até encontrar o conteúdo desejado, se sentindo frustrado (AIZPURUA; ARRUE; VIGO, 2013) e excluído digitalmente (MATTOS; CHAGAS, 2008) e (CARVALHO, 2003).

Vídeos na Web

Quando se trata de vídeos na Web, a barreira de acessibilidade ocorre dependendo da estrutura em que o reprodutor foi desenvolvido em termos de linguagem e tecnologias. Nestes casos, os leitores de tela não são capazes de identificar e informar ao usuário qual botão está em seleção de forma correta, esse problema torna impossível ao usuário selecionar qual botão deseja pressionar.

Interfaces como botões de *play*, *pause*, *stop*, volume, avançar, retroceder, botões de

³ <http://www.w3c.org>

navegação e seleção do vídeo podem mudar de plataforma para plataforma e também podem alterar seu *layout* de acordo com o dispositivo físico que o vídeo está sendo executado, estas características se não acessíveis dificultam ou até impossibilitam o uso de vídeos por usuários com deficiência visual.

Como exemplo, vídeos executados em dispositivos móveis geralmente possuem sua interface totalmente baseada em interação por *touch screen*. Um usuário com deficiência visual pode ter muita dificuldade em operar a interface, mesmo utilizando leitores de tela para dispositivos móveis (RICO; BREWSTER, 2009).

Os trabalhos de Gonzalez *et al.* (2011) e Rosas-Villena *et al.* (2015), destacam um avanço na área de acessibilidade para vídeos na Web por meio do desenvolvimento de um reprodutor chamado *Acessible HTML5 Media Player*, que oferece uma interface mais acessível, comparada a outros reprodutores, por ser utilizado facilmente em qualquer site devido a estruturação de sua linguagem.

Ainda no sentido de contribuições em acessibilidade a WCAG elaborou um documento onde definem diretrizes específicas para diversos aspectos da Web incluindo vídeos. No documento WCAG 2.0 na Seção de Acessibilidade Multimídia (WACG, 2012), a audiodescrição é apresentada como uma importante diretriz de acessibilidade para uso de vídeos.

Por meio da opção de audiodescrição é possível incluir áudios geralmente em momentos que não prejudiquem o áudio original do vídeo. Este áudio possui alguma descrição sobre um local, uma cena ou algo relevante para o entendimento do conteúdo do vídeo que só seria possível compreender visualmente, uma paisagem ou uma expressão que não possui sons é um exemplo de inserção de audiodescrição (WACG, 2012).

Como exemplo, Champin *et al.* (2010) realizou um estudo sobre acessibilidade para usuários com deficiência visual por meio da inserção de audiodescrição em cenas estrategicamente escolhidas de modo a não prejudicar o entendimento do áudio do vídeo. O enriquecimento do vídeo por meio de audiodescrições permite que o usuário tenha entendimento do conteúdo com a descrição da cena, ação ou elementos fundamentais.

Alguns estudos abordam a possibilidade do uso da audiodescrição de forma automática, como Hong *et al.* (2010) e Hong *et al.* (2011) que exploram do reconhecimento do movimento da boca em cenas que existam pessoas falando, dessa forma, pode-se reduzir as chances de inserir uma audiodescrição em cenas que existam fala de personagens.

Já Rosas-Villena *et al.* (2013) desenvolveram uma metodologia para ajudar no modo de desenvolvimento de vídeos acessíveis por meio de princípios da *User Agent Accessibility Guidelines* (UAAG) da W3C. Para avaliação da metodologia foi desenvolvido um reprodutor de vídeo chamado *Facilitas* que possibilitou a avaliação de heurísticas de acessibilidade em vídeos seguindo a metodologia proposta.

Perrenoud e Phan (2012) afirmam que questões de acessibilidade devem ser levantadas e

investigadas com o intuito de verificar se usuários com deficiências visuais conseguem utilizar plataformas de vídeos de forma satisfatória e também se páginas da Web têm contemplado as diretrizes de acessibilidade em vídeos. Apesar de muitos desses usuários citados por Perrenoud e Phan (2012) não conseguirem enxergar o conteúdo visual do vídeo, devido a um alto grau de deficiência, existe interesse no conteúdo de áudio do vídeo.

A união das áreas de acessibilidade e computação ubíqua pode criar alternativas para transpor diversas barreiras que são encontradas por usuários com interesse em vídeos na Web. Na Seção 2.4 serão discutido o estado da arte de trabalhos que exploram conceitos de computação ubíqua e acessibilidade.

2.4 Mapeamento Sistemático

Esta seção apresenta um mapeamento sistemático elaborado com o objetivo de identificar, avaliar e interpretar a pesquisa disponível e relevante sobre a seguinte questão de pesquisa: **"Como estão direcionadas as pesquisas em acessibilidade que utilizam conceitos de interações gestuais?"**. Ademais, busca-se também identificar quais são as técnicas, algoritmos, *frameworks* e demais recursos para capturar e processar dados de movimento gestual.

Bases selecionadas: foram selecionadas as seguintes bases científicas: *Scopus*⁴, *ACM Digital Library*⁵ e *IEEE Xplore*⁶.

CrITÉrios de Inclusão: trabalhos relacionados com as áreas de acessibilidade, gestos e interfaces Naturais.

CrITÉrios de Exclusão: trabalhos que não estejam em português ou inglês e trabalhos sem PDF disponível.

Termos adotados: para padronização e melhoria no resulta de buscas, foram adotados os seguintes termos:

1. Acessibilidade: "acessibilidade", "*accessibility*" e "*assistive*".
2. Gestos: "gestos", "**gesture**", "*gestural*" e "*gesture-based*".
3. Interfaces Naturais: "interfaces naturais", "*natural interface*" e "*natural user interface*".

Operadores de busca: após testes, adotou a seguinte lógica de operadores:

(gesto OR gesture OR gesture-based OR gestural)

AND

("interface natural" OR "natural interface" OR "natural user interface")

⁴ <https://www.scopus.com/>

⁵ <http://dl.acm.org/>

⁶ <http://ieeexplore.ieee.org/>

AND

(acessibilidade OR accessibility OR assistive)

Strings de busca: as *strings* de busca foram criadas para buscar os termos adotados presentes no título, *abstract* e palavras chaves de cada artigo. Abaixo são apresentadas as *strings* utilizadas em cada base:

ACM: (gesture OR gesture-based OR gestural OR gesto) AND ("interface natural" OR "natural interface" OR "natural user interface") AND (acessibilidade OR accessibility OR assistive)

IEEE: (((((((((gesto) OR gesture) OR gestural) OR gesture-based) AND acessibilidade) OR accessibility) OR assistive) AND interface natural) OR natural interface) OR natural user interface)

Scopus: (TITLE-ABS-KEY(gesto) OR TITLE-ABS-KEY(gesture) OR TITLE-ABS-KEY(gestural) OR TITLE-ABS-KEY(gesture-based)) AND (TITLE-ABS-KEY(acessibilidade) OR TITLE-ABS-KEY(accessibility) OR TITLE-ABS-KEY(assistive)) AND (TITLE-ABS-KEY("interface natural") OR TITLE-ABS-KEY("natural interface") OR TITLE-ABS-KEY("natural user interface"))

Trabalhos encontrados e selecionados: no [Quadro 1](#) são indicadas as quantidades de resultados obtidos em cada base de pesquisa utilizando a lógica de busca antes apresentada, a base de pesquisa *ACM* apresentou mais trabalhos relevantes sendo maior que as outras duas bases somadas. A quantidade de estudos selecionados é apresentada na [Figura 14](#), nota-se que apesar da base *ACM* prover a maior quantidade de resultados menos da metade apenas foi selecionado para leitura completa, nas bases *Scopus* e *IEEE* quase todos os artigos encontrados foram selecionados de interesse do objetivo da mapeamento sistemática.

Quadro 1 – Quantidade de trabalhos encontrados por base de pesquisa

Bases	Resultados
ACM	49
IEEE	23
Scopus	17
Total	89

Fonte: Dados da pesquisa.

Após a seleção e leitura dos resumos dos 89 artigos, foram selecionados 49 artigos para leitura completa. Durante a análise dos artigos, foram encontrados trabalhos relacionados com o uso de gestos para jogos, *frameworks* de captura gestual tridimensional, recomendações sobre a modelagem de sistemas para captura gestual de pessoas com necessidades especiais dentre outros. Percebe-se que não há nenhuma pesquisa relacionada com métodos de melhoria do nível de acessibilidade a reprodutores de vídeo utilizando interfaces naturais.

Figura 14 – Comparativo da quantidade de trabalhos encontrados e selecionados por base de pesquisa

Fonte: Dados da pesquisa.

A seguir, são descritos os trabalhos selecionados no mapeamento sistemático, que melhor responderam à pergunta da busca realizada, agrupados por dois temas de relevância da pesquisa, **Gestos aplicados em melhorias de acessibilidade** e **Captura e reconhecimento de gestos em contextos diversos**.

- **Gestos aplicados em melhorias de acessibilidade:** como exemplo de acessibilidade o trabalho de [Tavares et al. \(2013\)](#) intitulado *Hefestos Smart Wheelchair*, permite que cadeirantes e idosos obtenham informações sobre acessibilidade por meio de um monitor acoplado em suas cadeiras de rodas. Uma rede sem fio permite a conexão desses monitores e um sistema provê informações à cadeira de rodas sobre a localização em um mapa de recursos como: rampa de acesso, posto de atendimento, elevador, banheiros com acessibilidade e dentre outros. Isso permite que pessoas com deficiência motora ou idosos encontrem facilmente pontos de interesse utilizando o menor percussor.

Outros trabalhos tem como foco a melhoria do nível do acesso a locais físicos de interesse do usuário, essa é a proposta do trabalho de [Subbu, Gozick e Dantu \(2013\)](#) intitulado *LocateME* que tem por objetivo permitir que *smartphones* identifiquem pontos magnéticos espalhados em locais que podem ser de interesse durante o deslocamento do usuário, ao se aproximar de um ponto magnético o *smartphone* notifica o usuário que ali pode existir algo de interesse do mesmo. Apesar de existirem tecnologias como *Wi-fi* e redes móveis o projeto aposta na captura de assinaturas magnéticas para melhor precisão da localização dos pontos de interesse, isso permite que usuários com deficiência visual consigam obter informações sobre o ambiente durante seu deslocamento se a necessidade de estar conectados a uma rede sem fio.

Projetos também utilizam sensores espalhados pelo ambiente para auxiliar usuários com deficiência visual ao caminharem em ambientes fechados. Esse é o objetivo do projeto

Low-cost Stereo Vision System for Supporting the Visually Impaired's Walk desenvolvido por pesquisadores da *Prince of Songkla University* que utiliza o processamento de imagens em para detectar obstáculos no ambiente e fornecer um *feedback* em tempo real ao usuário (LIMNA; TANDAYYA; SUVANVORN, 2009).

Outro projeto chamado *Capti Narrator* desenvolvido por pesquisadores da *Charmtech Labs* permite criar *playlists*, por meio de uma interface natural com o conteúdo de navegadores, repositórios de arquivos como, *Dropbox* e *Google Drive* entre outros. A aplicação faz a leitura do conteúdo textual de um site ou arquivo e cria um arquivo de áudio que pode ser exibido em forma de lista, a leitura é feita de forma narrativa tendo o controle de velocidade de leitura feito pelo usuário. Por meio desse projeto um usuário com deficiência visual pode acessar facilmente informações textuais mesmo de sites que não possuam acessibilidade (BORODIN *et al.*, 2014).

Alguns projetos têm por objetivo a dinamização de uma interface que possa se adaptar dependendo do perfil e do ambiente ao qual está o usuário no momento da utilização. Essa é a proposta dos pesquisadores da *University of the Basque Country* com o estudo *Automatically Generating Tailored Accessible User Interfaces for Ubiquitous Services*, em seu conceito aplica o método EGOKI (ABASCAL *et al.*, 2011) que se encarrega de gerar uma interface acessível com base no contexto de uso, independentemente do tipo de aplicação o método modifica o *design* para facilitar o acesso, por exemplo, de usuários com alguns tipos de deficiências ou idosos.

No quesito acesso a informações governamentais deve existir uma grande preocupação com acessibilidade, com a popularização de *smartphones* muitos serviços de interesse público são acessado por meio de dispositivos móveis. Melhorar a acessibilidade dessas interfaces é fundamental para deficientes e usuários não experientes é a proposta de pesquisadores da *Universite Abderrahmane Mira* ao propor um *framework* com design focado no usuário para acesso de serviços governamentais (DJEDDI; DJILALI, 2015).

- **Captura e reconhecimento de gestos em contextos diversos:** a captura de gestos pode ser realizada de diversos modos e com diferentes dispositivos. Devido essa diversificação muitos estudos apresentam métodos de captura gestual diferentes porém focados em um determinado público alvo ou determinado escopo, como é o exemplo do trabalho Morelli e Folmer (2011) que utilizou o dispositivo *Kinect* para captura dos gestos realizados pelo corpo de um usuário deficiente visual e o controle *WiiMote* para envio de *feedbacks* táteis (vibração) sobre a interação do usuário. Isso permitiu melhorar o nível de acessibilidade que usuários com deficiência visual possuem em jogos que inicialmente não ofereciam suporte em acessibilidade.

Ainda sobre o uso dos conceitos de gamificação e captura gestual, o trabalho realizado por Qamar *et al.* (2014) que propõe a reabilitação de pacientes com algum tipo de paralisia.

Utilizando o dispositivo *Kinect*, os movimentos do usuário são capturados de forma não invasiva e permite o acesso à internet interagindo com *browsers*, assim usuário sentem maior inclusão digital na Web.

Existem também projetos que se preocupam com a qualidade do ensino que usuários com deficiência visual recebem em escolas públicas. Esse é o objetivo do estudo chamado *Haptic user interface design for students with visual impairments* desenvolvido pela *Virginia Tech* que tem foco no ensino de conceitos relacionados a calor e temperatura de forma tátil, o objetivo do estudo está na estruturação de interfaces hápticas que auxiliam no processo de ensino e aprendizagem desses conceitos em sala de aula (KIM, 2009).

Outra área de grande interesse na captura gestual é a Realidade Aumentada, devido a proporcionar experiências imersivas com o usuário. Como exemplo desse interesse o projeto *GestureChess*, desenvolvido por Dias *et al.* (2012), utiliza a captura gestual por meio do dispositivo *Kinect* e do *framework* OpenNI, onde seu objetivo é desenvolver atividades motoras e cognitivas utilizando gestos e a realidade aumentada com dois usuários interagindo simultaneamente.

2.5 Considerações finais

Neste capítulo foram apresentados alguns dos principais conceitos que estão envolvidos na fundamentação teórica dos estudos desenvolvidos neste trabalho.

Desde Weiser (1993), a computação ubíqua tem atraído novas pesquisas no meio acadêmico e agregando novas definições e aplicabilidades. Um dos objetivos presentes dentro da computação ubíqua são o estudo e desenvolvimento de Interfaces Naturais. Segundo Norman (1990), muitas vezes a interface é tão complicada que chama mais a atenção do usuário do que a própria funcionalidade que a aplicação possui. Usuários não querem ter o trabalho de pensar sobre a interface, querem desempenhar a função da aplicação ao qual necessitam sem maiores problemas, nesse contexto interfaces naturais facilitam a interação e usabilidade tornando-se transparentes durante seu uso.

Interações com interfaces baseadas em gestos podem existir por meio de gestos dimensionais realizados por usuários (MITRA; ACHARYA, 2007), tanto em forma bidimensional como, em superfícies de *smartphones* (HOGGAN *et al.*, 2013) ou executadas de forma tridimensional no espaço, (MANDERS *et al.*, 2008) e (STURMAN; ZELTZER, 1993). Utilizando sensores para a captura de gestos por meio de conceitos como, aproximação física (MOSER; TSCHELIGI, 2015) e (POTTER; ARAULLO; CARTER, 2013), vestíveis (MULLING; SATHIYANARAYANAN, 2015) e móveis (WILLIAMS, 2010) ou com a captura de gestos por processamento de imagens (CARTER *et al.*, 2015) e (NGUYEN *et al.*, 2011) é possível criar inúmeras possibilidades de interação com o usuário e tornar interfaces antes complicadas em agradáveis e usáveis.

Entendendo os detalhes e potencial apresentado pela computação ubíqua no nível de interfaces naturais e interfaces baseadas em gestos, a área de acessibilidade pode se beneficiar por meio de interfaces focadas em necessidades especiais (PERRENOUD; PHAN, 2012). Alguns dos problemas apresentados neste capítulo, foram às barreiras de acessibilidade que existem em reprodutores de vídeo na Web. Usuários com deficiência visual possuem interesse pelo conteúdo presente em plataformas de vídeo na Web, independentemente do nível de deficiência que cada usuário possui a acessibilidade deve ser garantida. Apesar de muitas pesquisas no estado da arte utilizarem gesto para melhorias ao usuário final, não foram encontradas pesquisas que discutissem melhorias no nível de acessibilidade para pessoas com deficiência visual no contexto de vídeos na Web, e esta lacuna de pesquisa pretende-se desenvolver no presente trabalho.

A investigação da possibilidade do uso de gestos que podem ser aprendidos sem necessidade de instruções visuais, em interfaces computacionais, para interação com vídeos é justificável devido ao potencial que pode exercer na melhoria do nível de acessibilidade se comparado a interfaces convencionais, atualmente disponíveis para usuários com deficiência visual. Neste trabalho pretende-se a melhoria de acessibilidade de pessoas com deficiência visual ao interagirem com vídeo na Web, para obtenção deste resultado o uso de interações gestuais poderá ser empregado para verificar possíveis melhorias na acessibilidade comparadas aos modos de interações atuais, como teclado, mouse e leitor de tela.

No próximo capítulo será apresentado um estudo de caso que possibilita obter indícios de pesquisa sobre acessibilidade e interações gestuais. Desse modo, este trabalho viabiliza a investigação sobre melhorias no nível de acessibilidade proposta nesta pesquisa.

ESTUDO DE CASO COM USUÁRIOS CEGOS E PESSOAS COM DEFICIÊNCIA VISUAL

Este capítulo descreve um estudo de caso conduzido com o objetivo de levantar indícios que ajudem a entender se a utilização de gestos como modo de interação entre pessoas com deficiência visual e reprodutores de vídeo oferece melhoria de acessibilidade.

O estudo de caso foi aprovado por avaliação de comitê de ética e recebeu o CAAE¹ nº 59269316.7.0000.5390. O desenvolvimento da metodologia e a análise de dados utilizada seguiu as orientações de (MACKENZIE, 2012) e o modo de condução de testes de acessibilidade foi baseado no trabalho de (FREIRE; LARA; FORTES, 2013).

Para melhor entendimento deste capítulo, a Figura 15 representa os elementos utilizados para a realização do estudo de caso e seu fluxo de interação, o conhecimento desses elementos é fundamental para o entendimento do estudo. Inicialmente o condutor dos testes escolhido interage com os participantes utilizando como guia o protocolo de estudo de caso, esse procedimento visa oferecer uma experiência agradável e semelhante a todos.

Os participantes realizam interações gestuais por meio dos sensores (*Smartphone*, *Webcam*, *Myo* e *Leap Motion*), interações estas que são recebidas, interpretadas e executadas por uma aplicação Web desenvolvida como protótipo para viabilizar os testes, a aplicação foi desenvolvida utilizando o *framework Gesture4All* desenvolvido especificamente para este propósito.

Cada um dos elementos descritos e ilustrados na Figura 15 estão descritas em detalhes nas próximas seções. Assim, este capítulo está organizado do seguinte modo, a Seção 3.1 discute o desenvolvimento de um *framework* chamado *Geture4All*, ao qual possibilita interação por meio de gestos com reprodutores de vídeo. A Seção 3.2 descreve o protocolo de estudo de caso utilizado para condução de testes que auxiliem a alcançar o objetivo supracitado. A Seção 3.3 apresenta os resultados e discussões pertinentes ao estudo de caso e, por fim, na Seção 3.4 as

¹ Certificado de Apresentação para Apreciação Ética

Figura 15 – Elementos do estudo de caso e suas interações

considerações finais sobre este capítulo serão abordadas.

3.1 Descrição do *framework Gesture4All*

Esta seção tem por finalidade apresentar a abordagem utilizada na concepção do *framework Gesture4All*. O *framework* foi desenvolvido com o objetivo principal de viabilizar a condução do estudo de caso apresentado neste capítulo. O protótipo desenvolvido, semelhante a uma aplicação Web, permite a interação por meio de gestos e dá suporte a quatro sensores gestuais.

Para melhor entendimento do contexto ao qual o *Gesture4All* se propõe a atuar, a [Subseção 3.1.1](#) descreve uma situação do cotidiano de um usuário com baixa visão que deseja consumir um vídeo na Web e dois cenários possíveis. O primeiro cenário descreve as ações deste usuário em seu cotidiano sem o uso do *Gesture4All*, o segundo cenário descreve como é possível melhorar a acessibilidade deste usuário utilizando o *framework* para interação com o reprodutor de vídeo.

3.1.1 Contextos de Uso

Um usuário com baixa visão está navegando pela Web por meio de seu *smartphone*, para auxílio na realização desta tarefa, faz uso de um *software* de leitura de tela. Ao se deparar com um reprodutor de vídeo presente no site em que se encontra decide consumir o vídeo, e portanto, utiliza o leitor de tela para interagir com a interface do reprodutor de vídeo (*play*, *pause*, controles de volume e etc). Após algumas tentativas, percebe que o leitor de tela não foi capaz de selecionar os botões da interface do reprodutor de vídeo devido sua tecnologia focar em interações com o *mouse*, desse modo o usuário não conseguiu atingir seu objetivo de consumir o vídeo.

Para melhorar o nível de acessibilidade do cenário apresentado, o *Gesture4All* é adicionado entre o usuário e a interface controlando as funções presentes no reprodutor de vídeo baseado nas interações gestuais realizadas pelo usuário, o fluxo de interação pode ser observado na [Figura 16](#).

Figura 16 – Visão geral do contexto de aplicação do *Gesture4All*

A seguir, são descritos cada um dos intens presentes no contexto da aplicação, os quais são esquematizados na [Figura 16](#).

1. Pessoa com baixa visão: inicia a interação do modelo por meio do uso de sensores gestuais (presentes no *smartphone* ao realizar um gesto em três dimensões).

2. *Gesture4All*: é responsável por intermediar a interação do usuário com a interface do reprodutor de vídeo na Web. Esta intermediação é iniciada ao receber os dados dos sensores gestuais realizados pelo usuário, que serão processados e realizarão a ação na interface do reprodutor baseada na semântica que o *Gesture4All* está configurado. Além disso, também fornece *feedback* adicionais resultantes da interação gestual para maior usabilidade e percepção da ação realizada. O *Gesture4All* não substitui o uso de outros *inputs* como teclado e *mouse* ou *softwares* de leitura de tela, mas sim, permite uma forma alternativa de interação dos usuários que encontram barreiras de acessibilidade ao interagir com reprodutores de vídeos na Web.

3. Interface do reprodutor de vídeo: para validar o modelo de interação e o uso do *framework*, foi utilizado nos experimentos um reprodutor de vídeo Web desenvolvido utilizando a linguagem HTML 5. Apesar do foco do *framework* ser a interface do reprodutor de vídeo na Web, o modelo de interação poder ser adaptado para um uso mais generalista interagindo com outras interfaces, como por exemplo: interface do navegador (scroll, voltar, avançar e etc), games, menus e etc.

O cenário de acesso a vídeos na Web por tecnologias *mobile* é um bom exemplo de como o *Gesture4All* pode ser utilizado, uma vez que o site que contém o reprodutor de vídeo Web faz uso da implementação e orientação que o *Gesture4All* recomenda. O usuário com baixa visão pode interagir com a interface do reprodutor de vídeo por meio de um *smartphone* realizando gestos em três dimensões enquanto segura o aparelho. O *Gesture4All* por sua vez, captura os

movimentos realizados utilizando o sensor gestual e interage com a interface do reprodutor de vídeo. Será apenas necessário que o usuário saiba qual gesto, configurado previamente pelo desenvolvedor da página Web, está semanticamente associado a uma função específica do reprodutor de vídeo, exemplo: executar um gesto movendo o aparelho para cima irá iniciar o vídeo e um gesto para baixo irá pausar o vídeo.

Para que o usuário saiba se sua interação gestual atingiu a interação esperada, dois tipos de *feedbacks* são por ele recebidos como exemplificado na [Figura 16](#). O primeiro *feedback* recebido pelo usuário é aquele que o reprodutor de vídeo possui usualmente e foi implementado pelo desenvolvedor do reprodutor, o *framework* não impede que os *feedbacks* usuais dos reprodutores sejam executados. O segundo tipo são os *feedbacks* que podem ser utilizados e estão previstos na arquitetura do *Gesture4All*, estes *feedbacks* adicionais tem a função de melhorar a compreensão do usuário se o gesto por ele executado foi corretamente interpretado pelo reprodutor de vídeo, são eles: o tátil, sendo sentido pelo usuário por meio da vibração dos sensores *smartphone* e *Myo*. O sonoro, sendo reproduzido um áudio com o nome da função que o reprodutor de vídeo executou (exemplo de palavras contidas no áudio: *play*, *pause*, aumentar volume e etc) tocado toda vez que o gesto relacionado a ele é reconhecido, e por fim, o visual que faz com que uma cor (que poderá ser escolhida) pisque no *background* da página Web, este último *feedback* foi adicionado para aqueles usuários que conseguem ter percepção a cores. O uso desses *feedbacks* são opcionais e podem ser ativados ou desativados de acordo com a preferência do usuário.

3.1.2 Arquitetura do *Gesture4All*

Após entendimento sobre o contexto e as melhorias na acessibilidade que o *Gesture4All* pode oferecer, esta seção irá aprofundar quais são as camadas e módulos que a arquitetura do *framework* utiliza. A discussão realizada aqui explana sobre o projeto de um *framework* Web, sua função é permitir o desenvolvimento de uma protótipo de aplicação Web que deem suporte aos quatro sensores gestuais escolhidos para coleta de dados.

O *Gesture4All* foi construído utilizando a linguagem de programação *JavaScript* para ser compatível com as estruturas de desenvolvimento Web atuais e ser facilmente integrado a outras estruturas já existentes. Sua arquitetura está dividida em três camadas sendo elas **Aquisição**, **Representação** e **Feedback** e um módulo chamado **Controle de Interface** como exemplificado na [Figura 17](#).

Camada de Aquisição: é responsável por garantir que a captura dos dados vindos dos sensores ocorra de modo correto para serem utilizados nas próximas camadas. Como entrada, recebe dados brutos dos sensores (baixo nível) realizando o processamento de reconhecimento gestual e como saída envia para a próxima camada qual gesto foi realizado pelo usuário (alto nível). O objetivo desta camada não é competir com técnicas já existentes propondo novas técnicas de reconhecimento gestual, mas sim, incluir as técnicas e procedimentos já existentes e permitir futuras implementações de novas técnicas que poderão vir a existir. O resultado final da

Figura 17 – Visão completa da Arquitetura do Gesture4All

Camada de Aquisição será o gesto já reconhecido e pronto para receber uma semântica baseada em qual interface se dará a interação do usuário, os detalhes do modo ocorre o reconhecimento serão explicadas adiante. Esta camada possui os módulos **Captura** e **Reconhecimento** que serão discutidos a seguir.

Módulo de Captura: possui quatro sensores que foram inicialmente adicionados no *Gesture4All* para fornecer opções diferentes de interações gestuais, são eles: *Smartphone*, *Webcam*, *Myo Armband* e *Leap Motion*.

O *smartphone* foi escolhido devido a sua popularidade atualmente, bem como a possível familiaridade com o aparelho, pelos participantes do estudo de caso. Segundo (KAMAL; LI; LANK, 2014), existem dois motivos principais para o uso deste sensor para reconhecimento gestual. O primeiro está ligado a seu potencial sensorial, *smartphones* modernos possuem um conjunto de sensores de movimentação, incluindo acelerômetros, giroscópios e câmeras que podem trabalhar de forma integrada isto possibilita a captura de movimentos sem a necessidade de percepção visual e sem o toque na tela, sendo assim ideal para o contexto deste trabalho. Outro motivo para sua escolha está no benefício da variedade de *feedbacks* que o sensor permite utilizar, sendo este: tátil devido a vibração que o aparelho emite podendo ter sua intensidade alterada, sonoro devido aos altos-falantes podendo ter seu volume alterado e visual devido sua tela. Os *feedbacks* tátil e sonoro são ideias para o uso de tecnologias assistivas e o visual pode ser empregado para usuários que possuem percepção de luz.

A *Webcam* foi selecionada como um sensor gestual para os testes com usuários devido a existência de técnicas de reconhecimento gestual baseadas em processamento de imagens presentes no estado da arte e, diversos *notebooks* e computadores já possuem este dispositivo. A exemplo de uso deste sensor, o trabalho de (SCHWALLER; LALANNE; KHALED, 2010) utiliza a *Webcam* para o reconhecimento de gestos focados nos dedos do usuário com taxas de acerto de 85,95%.

O *Myo Armband* foi selecionado devido sua precisão de captura baseada na estimulação muscular, em formato de uma braçadeira deve ser vestida pelo usuário até que fique bem presa ao braço porém de forma confortável, possui bateria interna e envia e recebe dados por meio de sinal *bluetooth*. Este sensor, permite reconhecer os gestos realizados detectando quais músculos foram estimulados, para isto ele captura os dados brutos gerados durante a performance do gesto utilizando informações espaciais como: altura, largura, profundidade ao qual o sensor se encontra em relação ao espaço. Outra característica é o reconhecimento de estimulação muscular por meio do sensor EMG², por meio dos dados gerados durante a estimulação muscular, é possível aplicar algoritmos que interpretam os dados a fim de encontrar padrões gestuais. O sensor é capaz de diferenciar se a mão do usuário está aberta ou fechada apenas analisando a forma que o músculo está disposto (MULLING; SATHIYANARAYANAN, 2015).

O *Leap Motion* utiliza infravermelho e câmeras para captar movimentos precisos e simultâneos dos dedos dentro de centésimos de milímetros, devido esta característica este sensor também foi selecionado. Seu uso tem se destacado em pesquisas relevantes na área de interação natural e tecnologias assistivas, como exemplo, os trabalhos recentes de (MAPARI; KHARAT, 2016) que obteve 90% de precisão no reconhecimento da Linguagem Americana de Sinais e (AFYOUNI *et al.*, 2017) que apresenta um *framework* para o reconhecimento de gestos focados na reabilitação de pacientes que sofreram paralisia cerebral e perderam movimentação dos membros superiores.

Como cada sensor possui características próprias de captura de gestos, o Módulo de Sensores contém as instruções de como cada sensor realiza a captura dos dados e quais *drives*, *plugins* ou recursos de linguagens de programação são necessários para o funcionamento de cada sensor. Caso exista a intenção de adicionar um novo sensor ao *framework*, será neste Módulo que ocorrerá as configurações para a nova adição.

Módulo de Reconhecimento: este módulo é responsável por receber os dados gerados no módulo anterior e realizar o reconhecimento gestual. Aqui podem ser implementadas diferentes técnicas de reconhecimento gestual ou soluções existentes no estado da arte ou no mercado, este módulo é flexível e admite novos sensores/técnicas de reconhecimento. Em todos os sensores escolhidos para uso, para que fossem integrados ao *framework* a seguinte sequência de eventos foi estabelecida: como entrada o *framework* recebe os dados brutos coletados dos sensores, para o *Smartphone* e *Myo* os dados coletados são advindos do acelerômetro e giroscópio presentes em seus respectivos *hardwares* e para a *Webcam* e *Leap Motion* a captura de imagem.

Em todos os casos são utilizados algoritmos que aplicam técnicas tanto de processamento de dados brutos (advindos de acelerômetro e giroscópio) quanto de processamento de imagens. Por meio da lógica presente nesses algoritmos, é verificado em tempo de execução qual o sentido espacial que o sensor (*hardware*) ou a mão do usuário (captura de imagem) está se dirigindo, como exemplo, se os dados brutos capturados são positivos e estão tendendo para

² Estudo ou método para o registro gráfico ou sonoro das correntes elétricas geradas em músculo ativo

valores números maiores, entende-se que o sensor ou a mão do usuário está sendo erguida no ar. Para integração com o *Gesture4All*, foi desenvolvida uma aplicação Web para cada sensor que ao ser carregada pelo navegador faz a chamada da função correspondente a execução do reconhecimento gestual, como resultado a função chamada devolve uma variável contendo o gesto reconhecido, variável esta passada como parâmetro para a próxima camada do *Gesture4All*.

Foram utilizadas as seguintes soluções de reconhecimento gestual:

Smartphone: foi utilizado o *framework Tilt-and-Tap* (GERONIMO; ARAS; NORRIE, 2015) devido sua sua precisão na captura, facilidade na implementação e fácil integração com tecnologias Web.

Webcam: foi utilizada a biblioteca de reconhecimento gestual para *Webcam Gest.js*³, que permite o reconhecimento das bordas da mão humana e permite ajuste de velocidade na realização do gesto além de reconhecer o sentido do movimento gestual.

Myo Armband: um algoritmo foi implementado seguindo as orientações de desenvolvimento do fabricante do sensor⁴ utilizando a linguagem JavaScript, que permite total integração com tecnologias Web e facilidade de implementação.

Leap Motion: um algoritmo foi implementado seguindo as orientações de desenvolvimento do próprio fabricante do sensor⁵ na linguagem JavaScript, que permitiu total integração com tecnologias Web e facilidade de implementação.

Camada de Representação: esta camada tem a função de tornar o gesto genérico por meio da desassociação tanto de qual sensor foi utilizado na captura quanto como qual técnica foi utilizada para o reconhecimento gestual. Ela é necessária para que as próximas camadas e módulos não fiquem dependentes dos sensores ou técnica de reconhecimento gestual. Esta camada possui os módulos **Abstração** e **Semântica** que serão discutidos a seguir.

Módulo de Abstração: para abstração do gesto reconhecido, este módulo é responsável pelo desenvolvimento de classes gestuais genéricas de forma conceitual, que permitem que os próximos módulos possam trabalhar em alto nível. Para melhor entendimento a abstração do gesto Inclinar, no contexto do *smartphone*, pode ser entendida do seguinte modo. Os usuários podem "Inclinar", "Sacudir" e "Rodar" o *smartphone* para executar gestos, o *framework* por sua vez deve reconhecer quando um destes movimentos fora executado pelo processamento em tempo de execução do dados gerados pelo sensor gestual, reconhecimento este feito no módulo anterior. Uma vez feito o reconhecimento do gesto executado do mundo real, o módulo de abstração permite a associação de classe aos movimento reconhecidos, como "inclinarEsquerda" e "inclinarDireita". Este módulo recebe como entrada o gesto reconhecido, desassocia o gesto do sensor e da técnica de reconhecimento e informa ao próximo módulo qual classe abstrata foi

³ <https://github.com/hadimichael/gest.js>

⁴ <https://github.com/thalmiclabs/myo.js>

⁵ <https://github.com/leapmotion/leapjs>

ativada.

Módulo de Semântica: este módulo é responsável por definir quais semânticas cada classe (definida no módulo anterior) deverá possuir de acordo com a interface que se pretende a interação de forma conceitual, aqui o *Gesture4All* poder ser personalizado para interagir com diferentes interfaces. Como exemplo para a classe "inclinaEsquerda" e a semântica associada foi de diminuir o volume do vídeo e para a classe "inclinaDireita" o aumento do volume.

Módulo de Controle da Interface: este módulo tem por função executar o que foi definido no módulo semântica diretamente na Interface pretendida na interação por meio da implementação de código. Para uso é imprescindível o conhecimento da linguagem e tecnologias que a interface possui para saber a melhor forma de realizar interações, devido a arquitetura do *framework* ser modularizada, aqui o desenvolvedor não precisa ter conhecimento sobre todas as etapas anteriores que foram executadas anteriormente, permitindo que desenvolvedores com conhecimentos diferentes possam trabalhar em conjunto. Como exemplo, no [Código-fonte 1](#) é associada uma função chamada "inclinaEsquerda", função está já definida no módulo de semântica, o controle do áudio no reprodutor de vídeo. Cada vez que o usuário faz um gesto inclinando o sensor para a esquerda a função "inclinaEsquerda" irá realizar uma chamada de função interagindo diretamente no reprodutor.

Código-fonte 1 – Exemplo de uso código implementado no módulo de controle de interface

```
1: <script src="js/gesture4all.js"></script>
2:
3: function inclinaEsquerda(){
4:   setVolume(-.1); //chamada da função de diminuir o volume
   presente no arquivo gesture4all.js
5: }
```

Camada de Feedback: devido alguns sensores permitirem meios alternativos de *feedback*, esta camada tem por objetivo fornecer *feedback* extra ao usuário para melhorar sua percepção sobre o *status* da interação com a interface e se sua ação atingiu o objetivo esperado. Apesar de estar representada na [Figura 17](#) como uma interação opcional, ela tem um papel muito importante no contexto de pessoa com baixa visão.

Módulo Sonoro: neste módulo são criadas diversas funções que executam um áudio (previamente gravado) todas as vezes em que são chamadas, as funções deverão ser instanciadas dentro de uma classe abstrata. No contexto de vídeo na Web, por exemplo, pode-se ter uma função de *feedback* sonoro com o nome de "audioPlay()", que emite um som de voz humana dizendo a palavra "play" toda vez que o gesto semanticamente associado a iniciar o vídeo foi executado.

Módulo Tátil: neste módulo é criada a função que permite que os sensores que, possuem

suporte a *feedback* tátil, sejam utilizados. Nos sensores atuais suportados pelo *Gesture4All* a vibração é o meio de *feedback* tátil suportado pelos sensores *Myo* e *smartphone*. Outra configuração realizada neste módulo é a intensidade e o tempo em que a vibração deverá ser executada para o usuário.

Módulo Visual: neste módulo são criadas diversas funções que inserem cores no *background* da aplicação Web para aqueles usuário com baixa visão que tem percepção a cores. Também é fornecido *feedback* textual sobre o resultado da ação do usuário, o texto poderá ser lido pelo *software* de leitura de tela, as características de como o texto deverá aparecer (coloração, tamanho e fonte) e a duração de exibição da cor no *background* da aplicação podem ser totalmente configuradas para ajuste de contraste.

No [Código-fonte 2](#), é exemplificado o modo de implementação dos *feedbacks* dentro da classe correspondente ao gesto realizado. Cada vez que o usuário fizer um gesto inclinando o sensor *smartphone* para a esquerda, a função "inclinaEsquerda", irá realizar as chamadas de funções presentes no *framework* que definem os comportamentos dos *feedbacks*.

Código-fonte 2 – Exemplo de uso código implementado nos módulos da camada de *feedback*

```
1: <script src="js/gesture4all.js"></script>
2:
3: function inclinaEsquerda(){
4:
5: //feedbacks
6:   audioMenosVolume(); //audível
7:   blinkMenosVolume(); //visual
8:   window.navigator.vibrate(50); //tátil
9:
10: }
```

3.2 Protocolo de condução de testes

Para a realização deste estudo de caso, foi definido um protocolo de condução de testes que inclui atividades importantes sobre o planejamento, condução e análise de dados. O protocolo utilizado neste trabalho, foi adaptado a partir do protocolo descrito no trabalho de ([FREIRE; LARA; FORTES, 2013](#)), que por sua vez, foi desenvolvido para testes de acessibilidade em páginas da Web.

O protocolo aqui utilizado foi desenvolvido para garantir a coleta de dados de forma a viabilizar uma análise que atenda o objetivo proposto neste trabalho, sendo assim fundamental para a execução da metodologia. Devido este objetivo, o protocolo possui duas principais adaptações do protocolo ao qual se baseia, a primeira é a definição de tarefas relativas ao uso de

gestos de forma a simular situações do cotidiano de um usuário na Web. A segunda é a adição da Fase 3 - Ambientação que será descrita com maior detalhe a seguir.

Segundo Freire, Lara e Fortes (2013), a realização de testes envolvendo questões de acessibilidade além de seguir os diversos princípios fundamentados na área de Interação Humano-Computador, deverá sofrer adaptações e considerar particularidades que cada usuário venha a possuir. Este protocolo, compreende estas adaptações e particularidades, garantindo assim uma coleta satisfatória de dados. O protocolo foi dividido em três fases, **Planejamento dos testes** (Subseção 3.2.1), **Condução dos testes** (Subseção 3.2.2), **Coleta de dados** (Subseção 3.2.3) que serão discutidos a seguir.

3.2.1 Planejamento

Os testes foram realizados com diferentes usuários e em épocas diferentes objetivando uma coleta de dados homogênea, o único pré-requisito estipulado para participação esteve em possuir alguma deficiência visual e experiência com o acesso de vídeos na Web. Os participantes foram divididos em dois grupos, o Grupo 1 e Grupo 2, para melhor identificação.

Grupo 1: no Grupo 1, que será aqui chamado de G1, os testes foram realizados no segundo semestre do ano de 2016 com 18 participantes, 11 homens e 7 mulheres entre 17 e 58 anos.

Grupo 2: no Grupo 2, que aqui será chamado de G2, os testes foram realizados no primeiro semestre do ano de 2017 com 20 participantes, 11 homens e 9 mulheres entre 14 e 61 anos seguindo o mesmo protocolo utilizado no G1.

Hardware e software: foi utilizado um *notebook* com o sistema operacional *Windows 10*, processador *i7*, 8 *gigabytes* de memória RAM e conexão com a Internet. Os usuários poderiam escolher entre a versão 2016.2 do leitor de tela NVDA⁶ ou a versão 14.0.9002 do leitor de tela JAWS⁷, para acessar os vídeos na Web dois navegadores: *Internet Explorer 8.0* ou *Firefox 35*.

Sensores gestuais: após análise dos sensores de captura gestual utilizados em trabalhos relacionados e disponíveis no mercado, quatro sensores foram selecionados para integrar as práticas no estudo de caso, sendo estes: *Smartphone*, *Webcam*, *Myo* e *Leap Motion*, maiores detalhes sobre os sensores estão descritos na Subseção 3.1.2.

3.2.2 Condução dos testes

Os testes pertencentes a esta fase de condução foram realizados em cinco etapas consecutivas, sendo estas: *Perfil*, *Receptibilidade*, *Ambientação*, *Tarefas* e *Avaliação*. As etapas foram conduzidas pelo condutor dos testes e abaixo estão elencadas as suas descrições e motivações.

⁶ <https://www.nvaccess.org/download/>

⁷ <https://www.freedomscientific.com/Downloads/JAWS>

Fase 1 - Perfil: consistiu em uma entrevista estruturada para coleta de dados sobre o perfil dos usuários seguindo as definições estabelecidas no planejamento dos testes [Subseção 3.2.1](#), o questionário utilizado para a entrevista estruturada no Grupo 1 encontra-se no **Anexo A** e o questionário utilizado no Grupo 2 encontra-se no **Anexo B**. Segundo ([BARBOSA; SILVA, 2010](#)), a entrevista estruturada assegura a validade de uma análise comparativa das respostas por toda a amostra de entrevistados devido as perguntas serem sempre as mesmas. Outra técnica considerada foi o uso de formulários eletrônicos juntamente com tecnologia assistiva ou impressão em braille, porém ambas estratégias dependem do conhecimento prévio do usuário o que não foi estabelecido como pré-requisito.

Fase 2 - Receptibilidade: consistiu em apresentar aos participantes os recursos disponíveis para a realização dos testes, descritos no planejamento, em uma conversa sem formalidades. Esta etapa foi de grande importância pois, permitiu o ajuste das condições mais favoráveis ao usuário (escolha do navegador, configurações de teclado, volume e etc), auxiliou o condutor dos testes ao conhecer às experiências prévias dos participantes e também incentivou-os a ficarem mais confiantes, visto que na maioria dos casos, este foi o primeiro contato com um estudo de caso. No contexto de interfaces baseadas em gestos, esta etapa auxiliou os participantes a entender o que são os sensores de captura gestual e o modo de funcionamento.

Fase 3 - Ambientação: esta etapa foi adicionada especificamente após considerar que uso de sensores gestuais poderia ser um paradigma totalmente desconhecido para os usuário, e sem a devida ambientação, a interação nos teste poderia ser prejudicada ou até mesmo inviabilizada. Os participantes puderam aprender qual gesto executava qual ação no reprodutor de vídeo em cada um dos sensores utilizados e tiveram a oportunidade de interagir com o vídeo e o sensor sem nenhuma tarefa ou objetivo, de modo livre, aprender com tentativas e erros e construir um modelo mental de interações gestuais.

Fase 4 - Tarefas: consistiu em explicar e acompanhar as tarefas que os usuários deveriam realizar. Objetivando a análise comparativa, foi requisitado que os usuários fizessem as mesmas tarefas tanto no modo de interação usual (ao qual estão acostumados em seu dia a dia) com *software* de leitura de tela e teclado, quanto no modo gestual com o uso dos sensores. As tarefas escolhidas fazem referência às ações básicas de interação com reprodutores de vídeos na Web, sendo estas:

Tarefa 1: reproduzir o vídeo

Tarefa 2: ajustar o volume de acordo com sua preferência.

Tarefa 3: após reprodução total do vídeo, navegar até um trecho de sua preferência.

Tarefa 4: após assistir o trecho escolhido pausar o vídeo.

Tarefa 5: voltar até o início do vídeo e pausar.

Fase 5 - Avaliação: após a execução das tarefas, o condutor dos testes realizou uma

entrevista semiestruturada para coletar dados sobre a experiência dos participantes ao realizar as tarefas, o questionário utilizado para a entrevista semiestruturada no Grupo 1 encontra-se no **Anexo A** e o questionário utilizado no Grupo 2 encontra-se no **Anexo B**. A escolha da entrevista semiestruturada permitiu seguir um roteiro preestabelecido, assim comparando as respostas dos participantes, porém sem excluir a possibilidade de coletar dados antes não estruturados.

3.2.3 Coleta de dados

Para viabilizar os testes foi desenvolvida uma aplicação Web seguindo as diretrizes de acessibilidade da [WCAG \(2008\)](#), contendo um reprodutor de vídeo em seus conteúdos. Uma tela dessa aplicação é mostrada na [Figura 18](#) e utiliza as linguagens HTML 5, CSS 3 e JavaScript que foram escolhidas pois são amplamente utilizadas na Web, simulando assim uma interação cotidiana do usuário.

Figura 18 – Aplicação Web desenvolvida para viabilizar o teste

Aos participantes foi disponibilizado um *notebook* e um *smartphone* com acesso a internet, leitores de tela e navegadores como citado anteriormente. Também foi disponibilizado aplicação Web, demonstrada na [Figura 18](#), para que os participantes executassem as tarefas em sua ordem de sequência. As tarefas foram executadas de dois modos, um usual e outro gestual. Primeiramente os usuários executam as tarefas da maneira ao qual estavam habituados, chamado aqui de modo usual, utilizando leitores de tela, teclado e *mouse*. Logo após, os participantes executaram as mesmas tarefas porém utilizando sensores gestuais, chamado aqui de modo gestual, para interação com o reprodutor de vídeo. A sequência do uso dos sensores foi feita de modo aleatório para que a coleta de dados não obedecesse apenas um viés de sequência de sensores garantindo maior consistência na estudo de caso.

Para atingir o objetivo proposto nesta pesquisa, foram selecionadas variáveis provenientes das fases citadas na [Subseção 3.2.1](#) além de observação direta e indireta da realização do estudo de caso, todos os participantes tiveram seus testes filmados, com respectivo consentimento, para auxílio da coleta de dados. Para melhor entendimento, estas variáveis foram divididas em quatro grupos, sendo este: Perfil, Sensor, Gestos e *Feedback*.

- Grupo Perfil: para este grupo 13 **variáveis qualitativas** e uma *variável quantitativa* (idade) e sua análise possibilita entender qual o perfil dos participantes do estudo de caso e qual sua relação com a tecnologia e experiências prévias no contexto desta pesquisa. As variáveis deste grupo são: gênero, idade, escolaridade, acuidade visual, tecnologia, interesse por vídeos, frequência de acesso, dispositivo mais utilizado, uso de programas de leitura de tela, site de vídeos mais utilizado, conteúdo de vídeo de maior interesse, dificuldades e barreiras de acessibilidade, controlar as funções do reprodutor de vídeo e interesse por aprender.
- Grupo Sensor: para este grupo cinco **variáveis qualitativas** descrevem às experiências dos usuários durante o uso dos sensores no estudo de caso, sendo estas variáveis: interação com o sensor gestual de forma geral, interação com o sensor *Smartphone*, interação com o sensor *Webcam*, interação com o sensor *Myo* e interação com o sensor *Leap Motion*.
- Grupo Gestos: para este grupo **variáveis quantitativas e qualitativas** foram coletadas para uma análise mais profunda sobre a relação dos usuários com a experiência gestual única de cada um, sendo essas:
 - Variáveis quantitativas: interação gestual de forma geral, interação gestual com o reprodutor de vídeo, interação com o sensor gestual, personalização de gestos, uso de gestos no cotidiano e melhoria em acessibilidade. Totalizando
 - Variáveis quantitativas: tarefas completadas com sucesso, número de erros cometidos e tempo gasto tentando efetuar uma tarefa.
- Grupo *Feedback*: para este grupo quatro **variáveis quantitativas** foram selecionadas para verificar como se deu a relação entre *feedback* e usuário, sendo estas: aural, visual e tátil.

A análise destas variáveis permite não somente compreender qual o perfil dos usuários com deficiência visual que possuem interesse em vídeos na Web, bem como verificar se existe melhoria nas interações com reprodutores de vídeo na Web por meio do uso de gestos, análise esta que será discutida na [Seção 3.3](#).

3.3 Resultados

Nesta seção, são apresentados os resultados obtidos a partir da condução de testes descrita na Seção 3.2. Estes resultados, por sua vez, tem por objetivo auxiliar na identificação de indícios ou evidências de melhorias na acessibilidade no uso de interfaces gestuais. Para a discussão dos resultados, utilizou-se a comparação das variáveis descritas e agrupadas no protocolo de condução de teste em ambos os grupos.

Segundo (MACKENZIE, 2012), a análise comparativa entre grupos de usuários permite validar os testes por meio de resultados próximos, uma vez que os testes foram executados com o mesmo protocolo. Portanto, à análise comparativa desta variáveis foi utilizada como ferramenta para atingir o objetivo do trabalho, pois auxilia a explicar diferenças entre diferentes perfis de usuários e encontrar possíveis barreiras.

As variáveis coletadas foram agrupadas em quatro áreas principais de interesse. Como demonstra Figura 19, os quatro grupos são: Perfil, Sensor, Gestos e *Feedback*. O resultado da análise dos grupos será discutido nas seções seguintes.

Figura 19 – Agrupamento das variáveis coletadas no estudo de caso

A discussão apresentada a seguir procura constatar, por meio dos resultados obtidos, que o uso de interações gestuais apresenta melhoria no nível de acessibilidade de vídeos na Web, tanto na realização de gestos para ações no reprodutor de vídeo quanto na interação com um sensor gestual. Esta conclusão pode ser constatada por meio das discussões que permeiam os quatro grupos a seguir discutidos.

3.3.1 Sobre o perfil dos usuários

Segundo MacKenzie (2012), a coleta de informações que permitem mapear o perfil dos participantes de um estudo é muito relevante pois possibilita uma análise correlacional ao se comparar essas variáveis e obter conhecimento sobre quem são os participantes do estudo, esse conhecimento além de ser útil em trabalho futuros permite também verificar se a população estatística é adequada. Devido o objetivo deste trabalho verificar melhorias para usuários, entender

quais são às características desses usuários é fundamental para que as soluções obtidas realmente impactem o cotidiano do público-alvo.

Ambos os grupos apresentaram resultados similares em Gênero e Idade. O G1 possui 61% dos usuários do sexo masculino e 39% feminino, o G2 56% do sexo masculino e 44% feminino, o G1 possui idades entre 17 e 58 anos e o G2 entre 14 e 61 anos. Apesar da maioria dos perfis ser masculino o experimento conseguiu coletar dados de forma abrangente não focando em apenas um gênero ou faixa etária. Com relação a Escolaridade, os grupos apresentaram o mesmo índice de 50% dos usuários com o Ensino Médio Completo, porém o G1 possui maior incidência de usuários com menores escolaridade (17% Fundamental Incompleto, 11% Ensino Médio Incompleto, 11% Superior Incompleto e 11% Superior Completo) e o G2 possui usuários com maiores índices de escolaridade (6% Fundamental Incompleto, 11% Ensino Médio Completo, 17% Superior Incompleto, 6% Superior Completo e 11% Pós-graduação Completa). Ambos os grupos uma boa quantidade de usuários com recursos a educação e escolaridade.

Com relação a Acuidade Visual, o G1 apresentou a incidência de mais da metade de usuários possuírem Cegueira Total com 61% dos usuários, Próximo à Cegueira 11% e Baixa Visão Profunda 17%, o G2 apresentou índices mais homogêneos onde 28% dos usuários possuem Cegueira Total, 17% Próximo à Cegueira, 11% Baixa Visão Profunda, 22% Baixa Visão Severa e 22% Baixa Visão Moderada. O público-alvo do estudo de caso não foi delimitado utilizando como critério a acuidade visual dos usuários devido a peculiaridade que cada pessoa com deficiência visual possui ao interagir com vídeo na Web. Apesar do interesse nas respostas e avaliação dos usuários com menores acuidades visuais, cada usuário é único e o fato de possuir maior acuidade visual não está totalmente relacionados com melhores resultados de acessibilidade.

Os usuários também relataram sobre o uso de algum recurso especial de auxílio como lupas e bengala, em ambos os grupos 72% necessitam de algum recurso e 28% não fazem uso, a maioria dos usuários fazem uso da bengala para locomoção, os usuários que relataram não utilizar esses recursos frequentemente fazem parte da porcentagem de usuários com Baixa Visão Moderada e Baixa Visão Severa.

Por meio da análise do interesse do usuário em Tecnologias é possível analisar a experiência citada no item anterior. O G1 possui 72% de usuários com mais de cinco anos de experiência no uso do computador e o G2 com 67% das respostas dos usuários (28% foi a soma-tória das respostas Entre 1 e 5 anos e Menos que 1 anos e o G2 possui 33% na mesma somatória), demonstrando assim que a maioria em ambos os grupos são usuários com certa bagagem de conhecimento e interação com tecnologia. A presença de mais da metade de usuários com certa experiência permite uma coleta mais focada nas barreiras relacionadas a acessibilidade do que nas barreiras enfrentada devido a falta de experiência com tecnologia. Um usuário inexperiente poderá ver algumas ações do uso do teclado ou do sistema operacional como uma barreira, caso que não ocorre com maior frequência em usuários experientes.

Devido o objetivo desta pesquisa estar situado na área de vídeos na Web, os usuários

foram questionados sobre suas experiências anteriores, para o mesmo foram coletadas às seguintes variáveis: **interesse por vídeos, frequência de acesso, dispositivo mais utilizado, uso de programas de leitura de tela, site de vídeos mais utilizado, conteúdo de vídeo de maior interesse, dificuldades e barreiras de acessibilidade, controlar as funções do reprodutor de vídeo, interesse por aprender.**

O primeiro item é de grande importância para esta investigação e verifica o **interesse dos usuários** com relação ao acesso a vídeos na Web. No G1 94% dos usuários acessam vídeos na Web e no G2 100% dos usuários também possuem o mesmo interesse, apenas um usuário respondeu não ter interesse em vídeos na Web representando 6% no G1. Ao ser questionado, o usuário relatou que as diversas barreiras somadas a inexperiência fazem das experiências de acesso a vídeos na Web (dentre outras experiências com tecnologias) serem frustrante e desanimadoras, os dados deste usuário foram mantidos na análise devido o interesse em sua opinião após fazer uso do interações gestuais.

Aos usuários foi pedido que classificassem sua **frequência de acesso a vídeos**, a grande maioria em ambos os grupos consome vídeos diariamente, no G1 78% acessam vídeo Diariamente, 11% Semanalmente e 11% Mensalmente já no G2 56% acessa vídeo Diariamente, 28% Semanalmente e 17% Mensalmente. Além da frequência os usuários responderam sobre seu **dispositivo mais utilizado**, no G1 por pouca diferença os usuários preferiam utilizar o computador com 50% das respostas, 45% prefere smartphones, 6% utilizam outros meios como aparelhos de televisão. Já no G2 os dispositivos móveis foram mais citados com 56% das respostas, 39% utiliza o computador e 6% outros dispositivos.

Além dos dispositivos, os usuários responderam sobre o uso de **programas de leitura de tela**. Os grupos obtiveram porcentagens bem diferentes sendo que no G1 as respostas foram 61% Sim e 39% Não, no G2 28% Sim e 72% Não. Estes resultados estão relacionados com a porcentagem de acuidade visual dos grupos que no G1 é maior (72% Próximo à Cegueira e Cegueira Total) que no G2 (45% Próximo à Cegueira e Cegueira Total). Apesar de usuários com maiores acuidades visuais responderem afirmativamente sobre utilizar programas de leitura de tela em suas interações com os dispositivos a incidência é maior entre pessoa com Cegueira Total ou Próximo à Cegueira.

Sobre **site de vídeos mais utilizado**, o *Youtube*⁸ recebeu a maior incidência de respostas em ambos os grupos com 94% dos usuários no G1 e 83% no G2. No G1 apenas um usuários afirmou preferir outros sites sem especificar o motivo e no G2 uma parcela de 17% apontou as Redes Sociais como plataforma mais utilizada para consumo de vídeos.

O **conteúdo de vídeo de maior interesse** também foi abordado durante a entrevista estruturada, no G1 50% dos usuários consomem mais Informações gerais, 11% Notícias diversas e 39% opta pela Entretenimento, números similares foram obtidos no G2 sendo 28% consomem

⁸ <https://www.youtube.com>

mais Informações de seu interesse, 6% Notícias diversas, 56% Entretenimento e 11% vídeos sobre Educação. Esses números estão relacionados a parcela de 11% de usuários que possuem a Escolaridade de Pós-Graduação pois os mesmos usuários apontaram utilizar vídeos educacionais em seu dia-a-dia devido a necessidade de pesquisas acadêmicas.

Outro aspecto de grande interesse da pesquisa foi questionar aos usuários se somando toda sua experiência própria consideravam enfrentar **barreiras de acessibilidade ou dificuldades** no acesso de vídeos na Web, no G1 61% disseram Sim e 39% disseram Não, no G2 100% disseram Sim e nenhum usuário respondeu negativamente. Isso implica em uma diferença de 39% a mais que o G2 afirmou enfrentar dificuldades ou barreiras, essa diferença não era esperada ou se ocorresse provavelmente seria esperada de forma contrária devido às características que cada grupo possui analisadas anteriores.

O G1 apresenta maiores índices de usuários com menor acuidade visual (Cegueira Total e Próximo à Cegueira), menores índices de Escolaridade (o que poderia estar relacionado com acesso ao conhecimento), apresenta o maior índice de usuários com Mais de Cinco Anos de experiência com utilização de tecnologia e mais usuários consumindo Diariamente vídeos na Web, então *porque 100% do G2 afirmam enfrentar dificuldades e/ou barreiras de acessibilidade?*

Para responder a essa pergunta primeiro verifica-se a associação existente entre as variáveis Programa de Leitura (PI) e Dificuldades (D) dos grupos 1 (G1) e 2 (G2) por meio do teste de Chi-Quadrado, seguindo as orientações de (MACKENZIE, 2012), para testar as seguintes hipóteses:

Hipóteses nula (H₀): não há associação entre as variáveis, ou seja, as respostas sobre utilizar Programas de Leitura e sentir Dificuldades são independentes.

Hipóteses alternativa (H_a): há associação entre as variáveis, ou seja, as respostas sobre utilizar Programas de Leitura e sentir Dificuldades não são independentes.

O cálculo efetuado utilizou o nível de significância (alfa) de 5% (0.05) e os resultados encontram-se na [Tabela 1](#).

Devido ao p-valor ser menor que o valor de significância ($0.0001 < 0.05$) rejeita-se a hipótese nula (H₀) de não associação, ou seja, existem evidências para afirmar que uma variável está associada a outra. No [Figura 20](#) é possível analisar como essa associação ocorre entre os perfis de usuários do G1 e G2.

Apesar das características atribuídas anteriormente ao G1, ao comparar as variáveis **programas de leitura de tela** nota-se que o G1 possui uma relação equilibrada entre o uso de leitores de tela (61% Sim e 39% Não) e o encontro de dificuldades (61% Sim e 39% Não), ou seja, usuários com menores acuidades visuais sentem mais necessidade de possuir o conhecimento de uma tecnologia assistiva ao interagir com a Web, essa correlação pode ser verificada ao observar a mesma relação no G2 onde a grande maioria não utiliza leitores de tela (28% Sim e 72% Não) o que está totalmente relacionado aos dados sobre encontrar barreiras e dificuldades (100% Sim

Tabela 1 – Programa de Leitura (Pl) x Dificuldades(D)

Chi-quadrado	G1-PL	G1-D	G2-PL	G2-D	Total
Sim	61%	61%	28%	100%	
Esperado	62.50	62.50	62.50	62.50	250
Observado	0.04	0.04	19.04	22.50	
Não	39%	39%	72%	0%	
Esperado	37.50	37.50	37.50	37.50	150
Observado	0.06	0.06	31.74	37.50	
Total	100	100	100	100	400

Resultado
 chi-quadrado: 110.976 Grau de liberdade: 3
 P-valor: 0.0001 Resultado: Rejeita-se H₀

Figura 20 – Comparativo entre Programas de Leitura e Dificuldades

e 0% Não).

Mesmo o G2 apresentando maiores índices de acuidade (Baixa Visão Moderada, Baixa Visão Severa e Baixa Visão Profunda), apresentando levemente maiores índices de Escolaridade, possuir mais usuários com Menos de Um Ano de experiência com utilização de informática e mais usuários consumindo Diariamente vídeos na Web, o perfil do G2 é de usuários que não se sentem totalmente obrigados a procurar o conhecimento de leitores de tela. Outro fator que se pode atribuir a essa diferença é a definição de "enfrentar barreiras de acessibilidade" que cada usuário atribui a si mesmo. Não espera-se que estes usuários possuam total conhecimento das diretrizes de acessibilidade recomendadas na Web e devido a esse fator foi incluída na entrevista estruturada a palavra "dificuldades" justamente para auxiliar o usuário na elaboração de sua pergunta. Conclui-se então que os usuários, mesmo possuindo maior acuidade visual, poderiam

ter maiores ganhos de acessibilidade utilizando leitores de tela e não apenas utilizando a visão que possui para interação com vídeos na Web.

Outro interesse relacionado ao objetivo desta pesquisa, está no comportamento que o usuário apresenta ao **controlar as funções do reprodutor de vídeo**, este interesse permite compreender quais são as preferências do usuário em interagir com reprodutores de vídeo e assim levantar possíveis barreiras. No G1 61% dos usuários afirmaram realizar diversas interações ao longo do vídeo (ajustar o volume, retroceder ou avançar o vídeo e etc) e 39% preferem não realizar interações, no G2 94% dos usuários afirmam realizar diversas interações ao longo do vídeo e 6% não o fazem.

Essa diferença, entre os grupos, de 33% em realizar diversas interações está totalmente correlacionada com a variável anterior sobre o conhecimento de tecnologias assistivas como leitores de tela. Quando o reprodutor de vídeo oferece total suporte a leitores de tela, facilmente os usuários que possuem tal conhecimento conseguem realizar qualquer tipo de interação que desejarem. Já usuários que não se sentem motivados a procurar esse conhecimento (devido possuir menor acuidade visual) ao se depararem com sites sem acessibilidade não conseguem realizar interações pretendidas e optam por apenas não realizar interações.

Ainda sobre o comportamento de interação com a interface do reprodutor de vídeo, os usuários foram questionados sobre qual tipo de interação mais realizam ou tem preferência por realizar durante suas experiências. No G1 67% responderam controlar o tempo de vídeo (Avançar e Retroceder) e 33% utilizar o recurso de Pausar o vídeo. O G2 apresentou resultados diferentes nesse nível onde 61% responderam ajustar o volume e 39% controlar o tempo de vídeo (Avançar e Retroceder).

No âmbito do consumo de vídeos por pessoas com deficiência visual, a interação principalmente dos recursos de avançar e retroceder um vídeo e o volume possuem maior destaque devido o consumo de vídeo por muitos desses usuários serem do áudio do vídeo, poder ouvir novamente um trecho ou aumentar o volume para melhor entendimento são fundamentais para a compreensão do conteúdo de vídeo e quando não são de fácil acessibilidade criam barreira e frustração a estes usuários.

Ao comparar os grupos sobre o comportamento de interação com a interface do reprodutor de vídeo na Web, uma comparação interessante está nas respostas sobre Pausar e controlar o Volume em ambos os grupos. Nota-se que não houve nenhuma resposta sobre Pausar o vídeo no G2 e não houve nenhuma resposta sobre Volume no G1. Geralmente os usuários recorrem a interações sempre que julgam necessário para melhoria das experiências de consumo de vídeo. A pergunta realizada na entrevista que possibilitou a coleta deste dado, teve como objetivo coletar a interação que o usuário julga ser mais utilizada por ele.

Portanto, para os usuários com menor de acuidade visual, geralmente cegos, (a grande maioria dos usuários do G1) a possibilidade de ouvir um trecho não compreendido é mais

importante e substancial para a compreensão do conteúdo do que as outras interações e quanto mais esta função não é acessível, maior será a barreira de acessibilidade. Para os usuários do G2 que possuem maior de acuidade visual e fazem uso da visão que possuem para interação, o controle de tempo não é o mais substancial dando lugar para o controle do vídeo ocupar o primeiro lugar nas respostas destes usuários.

Como último assunto da entrevista semiestruturada os usuários puderam demonstrar seu **interesse por aprender** uma nova forma de acessar vídeo na Web, diferente ao qual estavam acostumados. Em ambos os grupos 100% dos usuários responderam afirmativamente por desejar conhecer uma nova forma de interação com vídeos na Web incluindo aquelas que utilizassem novas tecnologias. Essa porcentagem demonstra que os usuários desejam e precisam de novos modos de interação, o que motiva a busca por alcançar o objetivo desta pesquisa.

3.3.2 Sobre os sensores gestuais utilizados

No grupo 1, os testes de interação gestual ocorreram apenas com um sensor, o *Myo*. Os testes com o primeiro grupo tiveram caráter exploratório pois, na época em que ocorreram os testes, não foram encontrados estudo publicados sobre o uso deste sensor com pessoas com deficiência visual. Os testes com o grupo 1 provaram ser promissores, porém, o uso de apenas um sensor não seria suficiente para aprofundamento na análise do uso de sensores gestuais no contexto desta pesquisa. Devido este motivo, nos testes com o grupo 2 foram adicionados mais 3 sensores, cada um possuindo uma abordagem de captura gestual diferente.

Nesta seção, será discutido a análise comparativa entre os grupos utilizando os dados coletados com o uso do sensor *Myo*. Após esta discussão, será discutida a análise dos quatro sensores coletados exclusivamente no grupo 2.

Análise comparativa entre os grupos com o sensor *Myo*

No G1 72% dos usuários afirmaram sentir algum tipo de dificuldade ao interagir com o sensor *Myo* e 28% não sentiram qualquer dificuldade, o G2 apresentou um resultado 6% mais positivos onde 67% dos usuários afirmaram ter sentido algum tipo de dificuldade e 33% não sentiram qualquer dificuldade durante o uso do sensor. Essa pequena diferença demonstra que perfis diferentes de usuários tiveram opiniões semelhantes com relação ao sensor.

Também lhes foi questionado sobre qual tarefa realizada com o *Myo* foi mais difícil, inicialmente foram apresentados quatro opções de respostas, sendo estas: **Play / Pause**, **Avançar / Retroceder** e **Volume**. Para esta interação gestual foram adotadas as configurações de reconhecimento gestual padrão do sensor *Myo* que foram implementadas na linguagem JavaScript, dentro do *framework Gesture4All* por meio de funções fornecidas pela própria fabricante como descrito na [Subseção 3.1.2](#). A intenção do experimento esteve em oferecer aos usuários uma configuração de gestos já existente e utilizada no mercado, para coletar os dados gerados de suas opiniões. Para melhor análise das respostas dos usuários algumas considerações sobre a

execução dos gestos em cada tarefa executada se faz necessária.

Destruar o sensor: por padrão o *Myo* apenas executa o reconhecimento gestual quando é destravado, para este procedimento o usuário deve efetuar o gesto de tocar os dedos polegar e anelar rapidamente duas vezes, após esse procedimento um *feedback* tátil por meio de vibração é enviado ao *Myo* e um *feedback* visual é exibido no monitor, o destravamento permanece por alguns segundos antes de voltar ao estado inicial.

Play e Pause: para estas ações no reprodutor de vídeo, o *Myo* reconhece o mesmo gesto de espalmar as mãos separando bem os dedos. Caso o vídeo encontra-se em reprodução, este gesto executa a ação de pausar o vídeo e vice-versa.

Avançar / Retroceder: para facilitar a associação de gestos com a ação no reprodutor de vídeo, para Avançar ou Retroceder, o usuário deve realizar o gesto chamado de "onda" para a direita se desejar Avançar e para a esquerda se desejar Retroceder. Este gesto consiste em permanecer com o braço parado, suspenso no ar, e levemente inclinar apenas a mão para a direita ou para a esquerda.

Volume: para controlar o Volume o usuário deverá realizar o gesto de fechar o punho e rotacionar para a direita se desejar aumentar o volume ou para a esquerda se desejar diminuir o volume.

Após apresentação dos detalhes de como cada ação no reprodutor de vídeo foi executada pelo usuário utilizando o *Myo*, a [Tabela 2](#) sumariza as porcentagens referentes às opiniões dos usuários quanto a dificuldade de interação com o sensor.

Tabela 2 – Usuários e suas dificuldades

Dificuldade	Grupo 1	Grupo 2	Total
Nenhuma	28%	33%	61%
Play / Pause	11%	28%	39%
Volume	22%	17%	39%
Diversas	17%	17%	34%
Avançar / Retroceder	22%	5%	27%

Apesar de não oferecido aos usuários as opções de resposta Nenhuma inicialmente ela foi incorporada após ser citada por usuários que afirmaram não consideram alguma interação gestual com o *Myo* difícil a ponto de utilizar outra classificação. No total 61% não alegaram dificuldades sendo 28% no G1 e no G2 33%. Essa grande porcentagem permite entender que mesmo um dispositivo gestual vestível que não faz parte do cotidiano do usuário pode ser bem aceito pelos mesmos.

A ações de *Play*, *Pause* e *Volume* aparecem em primeiro lugar como dificuldade e foram citadas por 39% no total dos usuários. Outra categoria adicionada posteriormente foi a *Diversas* que reúne principalmente dificuldades motoras ao realizar os gestos. *Avançar* e *Retroceder*

aparece em último lugar com 28% de citações no total. Após observação da interação gestual realizada por estes usuários notou-se que essas dificuldades poderiam ser diminuídas com o treino do sensor e da execução dos gestos. Como citado na [Seção 3.2](#), a Fase 3 - Ambientação, foi adicionada ao protocolo de testes para auxiliar os usuários a se ambientarem com um sensor gestual. Devido a esta fase, os usuários tiveram a oportunidade de interagir por alguns minutos com o sensor para melhores resultados ao chegarem nas tarefas no reprodutor de vídeo. Apesar desta ambientação, para todos os usuários utilizar um sensor gestual vestível foi uma experiência completamente nova, portanto, várias seções de testes e treinamentos seriam ideais para verificar se a dificuldade se dá apenas por inexperiência de uso ou outro fator diretamente ligado a questões motoras ou gestuais.

Análise dos quatro sensores utilizados no grupo 2

Foi observado após os testes com o grupo 1 que apenas a interação com um sensor gestual não seria suficiente para a real compreensão sobre possíveis ganhos em acessibilidade, portanto, para o segundo grupo manteve-se o sensor *Myo* para comparação de resultados e foram acrescentados os sensores: *Webcam*, *Myo* e *Leap Motion*, como discutido na [Subseção 3.2.1](#). Abaixo serão discutidos os resultados obtidos tanto na forma individual de cada sensor como um comparativo entre ambos.

1) Smartphone: o *smartphone* também foi muito bem classificado tendo apenas respostas do tipo "Muito Fácil" e "Fácil" no nível de facilidade. Outro fator que agradou que contribuiu para este resultado está em sua popularidade, onde 55% dos usuários relataram utilizar mais o *smartphone* para acessar vídeo do que outros dispositivos. Ao serem questionados sobre sua preferência por este sensor, os usuários citaram diversos fatores como: fácil acesso de compra, permitir todos os *feedbacks*, possibilidade de conexão estável com a Internet e baixo custo.

1.1) Smartphone Facilidades e Dificuldades: para maior profundidade na compreensão sobre as experiências dos usuários com o sensor *smartphone*, a [Tabela 3](#) permite um comparativo entre as facilidades e dificuldades relatadas sobre as interações gestuais realizadas para ações no reprodutor de vídeo.

Tabela 3 – Ações no *Smartphone* de maior facilidade e dificuldade

Ações no Smartphone	Facilidades	Dificuldades
Play	30%	5%
Pause	5%	-
Aumentar Volume	-	-
Diminuir Volume	-	-
Avançar	-	5%
Retroceder	-	5%
Todas	65%	-
Nenhuma	-	85%
Total	100%	100%

Os usuários foram questionados sobre qual ação no reprodutor de vídeo julgavam ter realizado com maior facilidade e dificuldade utilizando gesto por meio do *smartphone*. A ação "Play" foi citada como a mais fácil com 30% das respostas e a opção "Todas" obteve 65% de afirmativas, já em dificuldades não foi obtido porcentagem significativa em alguma ação sendo 85% de respostas "Nenhuma". Tais resultados indicam que, de forma geral, este sensor oferece grande facilidade de interação durante as execução das tarefas. Inicialmente as opções "Todas" e "Nenhuma" não contavam entre alternativas dadas aos usuários, elas foram incorporadas após o primeiro usuário ter citado de forma espontânea que "Todas" ações foram fáceis para ele, consequentemente, "Nenhuma" também foi adicionada como complemento.

2) Webcam: a Webcam foi o sensor classificado com mais resposta negativas, tendo as respostas "Difícil" e "Muito Difícil" representando 35%. Um dos fatores que foi observado durante a interação, foi o fator velocidade e ruídos na imagem. Para reconhecimento do gesto a biblioteca utilizada baseia-se na velocidade e direção com que a mão passa sobre a Webcam, muitos usuários tiveram dificuldades de ajustar a velocidade do movimento, outro fator que não contribui para uma boa experiência ocorre quando outros objetos estão sobre o alcance da câmera, como o rosto do usuários ou objetos atrás do mesmo. O posicionamento correto da Webcam é fundamental para uma boa captura gestual.

2.1) Webcam Facilidades e Dificuldades: para maior profundidade na compreensão sobre as experiências dos usuários com o sensor Webcam, a Tabela 4 permite um comparativo entre as facilidades e dificuldades relatadas sobre as interações gestuais realizadas para ações no reprodutor de vídeo.

Tabela 4 – Ações na Webcam de maior facilidade e dificuldade

Ações na Webcam	Facilidades	Dificuldades
Play	35%	5%
Pause	5%	5%
Aumentar Volume	-	20%
Diminuir Volume	-	-
Avançar	-	5%
Retroceder	-	5%
Todas	25%	40%
Nenhuma	35%	20%
Total	100%	100%

Os usuários foram questionados sobre qual ação no reprodutor de vídeo julgavam ter realizado com maior facilidade e dificuldade utilizando gesto por meio da Webcam. A ação "Play" foi citada como a mais fácil com 35% das respostas, a opção "Nenhuma" facilidade também obteve 35% das respostas contra 25% da opção "Todas". Isso demonstra que não houve uma tendência nos usuário sobre a facilidade de interação com este sensor. Sobre as dificuldades, quase todas as ações obtiveram alguma resposta, sendo que na ação "Aumentar Volume" obteve 20% de respostas sendo a mais citada. A opção "Todas" as dificuldade obteve 40% de afirmativas.

Esses resultados indicam que a *Webcam* apresentou maiores dificuldades do que facilidades aos usuários.

Um dos fatores observados pelos pesquisadores durante a observação direta e indireta do estudo de caso, que podem explicar este resultados, está no quesito de noção espacial. Os usuário ficavam confusos sobre onde deveria executar os gestos no ar para que a interação fosse capturada de forma satisfatória, a noção espacial da posição e distância da câmera é essencial para seu uso sem dificuldades. Outro fator está no modo de reconhecimento do gesto da *Webcam* utilizado pelo *framework*, em muitos casos o algoritmo de processamento de imagem não conseguiu diferenciar as mãos e braços de outros objeto que estivesse em sua área de captura como seu próprio rosto ou tronco.

3) Myo: o *Myo* também recebeu classificações positivas, as respostas "Muito Fácil" e "Fácil" somadas representam o total de 70% fazendo desse sensor uma ótima opção para captura de gestos. Duas características peculiares observada durante a interação com este sensor estão no quesito de usabilidade. Por ser vestível, o *Myo* possui um movimento padrão que deve ser realizado sempre que o usuário quiser realizar um gestos, no manual de instruções do sensor este movimento é chamado de "destravar" e consiste em tocar rapidamente os dedos anelar e polegar por duas vezes. Isso impede que todos os movimentos que o usuário faz sejam interpretados como tentativas de interação gestual, alguns usuários não conseguiram realizar este gesto padrão o que impossibilitou a realização de qualquer interação. Outro fator de usabilidade observado está na anatomia do usuário, alguns usuários enfrentaram dificuldades de vestir o sensor devido a espessura do braço, em alguns usuários mesmo vestindo o sensor o reconhecimento muscular ficou prejudicado devido a espessura do mesmo.

3.1) Myo Facilidades e Dificuldades: para maior profundidade na compreensão sobre a experiências dos usuários com o sensor *Myo*, a [Tabela 5](#) permite um comparativo entre as facilidades e dificuldades relatadas sobre as interações gestuais realizadas para ações no reprodutor de vídeo.

Tabela 5 – Ações no *Myo* de maior facilidade e dificuldade

Ações no Myo	Facilidades	Dificuldades
Play	25%	20%
Pause	5%	5%
Aumentar Volume	10%	5%
Diminuir Volume	5%	15%
Avançar	10%	-
Retroceder	-	5%
Todas	30%	15%
Nenhuma	15%	35%
Total	100%	100%

Os usuários foram questionados sobre qual ação no reprodutor de vídeo julgavam ter realizado com maior facilidade e dificuldade utilizando gesto por meio do *Myo*. Os resultados

demonstraram não existir uma tendência dos usuários sobre o sensor ser muito fácil ou muito difícil para uso. Quase todas as ações foram citadas tanto no quesito facilidade e dificuldade, demonstrando que os usuário ficaram divididos sobre este sensor.

Sobre facilidades encontradas as ações "Aumentar Volume" e "Diminuir Volume" obtiveram 25% de respostas e a ação "Play" obteve 25%, a opção "Todas" as facilidade obteve a maior porcentagem com 30% de respostas contra a 15% da opção "Nenhuma" facilidade. Os resultados obtidos com relação às dificuldades demonstram que ação de "Play" obteve 20% das respostas e a opção "Diminuir Volume" 15%, já a opção "Nenhuma" obteve o maior índice de respostas contando com 35% contra 15% de respostas "Todas" as dificuldades.

Aos comparar os resultados com outros sensores na opção "Play" uma particularidade ocorre com o sensor *Myo*, os resultados de facilidade (25%) e dificuldade (20%) estão muito próximos. Um dos fatores observados pelos pesquisadores durante a observação direta e indireta do estudo de caso, que podem explicar este resultados, está na característica física dos usuários. Devido este sensor ser vestível, diferente dos outros, a espessura do braço onde o sensor foi posicionado influência na captura do gesto. Apesar do sensor permitir o ajuste de tamanho, em braços com espessuras muito grandes ou muito finas a captura dos gestos não ocorreu de forma satisfatória, os resultados obtidos por este sensor estão diretamente ligados com as características físicas dos usuários o que explica a heterogeneidade das porcentagens obtidas na [Tabela 5](#).

4) Leap Motion: este sensor obteve os resultados com maior facilidades segundo os usuários, as repostas "Muito Fácil" e "Fácil" somadas representam o total de 100%. Realizando uma comparação dos sensores baseada nas resposta e na observação do condutor dos testes, de todos os sensores o *Leap Motion* possui a menor exigência de realização dos gestos devido seu foco de reconhecimento gestual estar no movimentos dos dedos. Outro fator positivo para o *Leap Motion* está sobre sua forma física, diferente do *smartphone* que exige que o usuário segure o dispositivos com as mãos ou do *Myo* que o usuário vesta o sensor todo o momento da interação, o *Leap Motion* fica localizado sobre a mesa e somente é ativado quando o usuário posiciona uma das mão sobre o sensor e realiza um gesto suave.

4.1) Leap Motion Facilidades e Dificuldades: para maior profundidade na compreensão sobre a experiências dos usuários com o sensor *Leap Motion*, a [Tabela 6](#) permite um comparativo entre as facilidades e dificuldades relatadas sobre as interações gestuais realizadas para ações no reprodutor de vídeo.

Os usuários foram questionados sobre qual ação no reprodutor de vídeo julgavam ter realizado com maior facilidade e dificuldade utilizando gesto por meio da *Leap Motion*. Os resultados demonstraram uma forte tendência de facilidade fazendo deste sensor o mais fácil para uso segundo os usuários. Sobre facilidades a ação "Aumentar Volume" obteve 10% de respostas e a opção "Todas" as facilidades obteve 90% de respostas, os resultados obtidos em dificuldades ajudam a colaborar com este resultados onde, 95% dos usuários responderam "Nenhuma" dificuldade encontrada na execução de ações com este sensor.

Tabela 6 – Ações no Leap Motion de maior facilidade e dificuldade

Ações no Leap Motion	Facilidades	Dificuldades
Play	-	5%
Pause	-	-
Aumentar Volume	10%	-
Diminuir Volume	-	-
Avançar	-	-
Retroceder	-	-
Todas	90%	-
Nenhuma	-	95%
Total	100%	100%

5) Qual sensor obteve os melhores resultados:

5.1) Sensor mais fácil no geral

Aos serem entrevistados sobre a facilidade de uso dos sensores de forma geral, foi dado ao usuário 5 opções de resposta baseadas na escala Likert (Muito Fácil, Fácil, Nem Fácil Nem Difícil e Difícil). O sensor *Leap Motion* obteve o maior índice de classificações "Muito Fácil" com 70% das respostas, em segundo lugar o *Smartphone* com 50%, e em terceiro o *Myo* com 20%, a *Webcam* não recebeu nenhuma classificação "Muito Fácil" como demonstra a [Tabela 7](#).

Tabela 7 – Comparativo de uso entre os sensores por meio da escala Likert

Sensores	Muito Fácil	Fácil	Nem Fácil Nem Difícil	Difícil	Muito Difícil	Total
Smartphone	50%	45%	5%	0%	0%	100%
Webcam	0%	15%	50%	30%	5%	100%
Myo	20%	50%	15%	15%	0%	100%
Leap Motion	70%	30%	0%	0%	0%	100%
Percentual Correspondente	35%	35%	18%	11%	1%	100%

Ao comparar os níveis da escala Likert utilizada, a classificação "Muito Fácil" obteve 35% das respostas e a classificação "Fácil" também obteve 35%, juntas representam 70% do percentual correspondente indicando que no geral os usuários tiveram ótimas experiências com os todos os sensores. As classificações "Difícil" e "Muito Difícil" juntas correspondem a 12% e a classificação intermediária "Nem Fácil Nem Difícil" ocupa a segunda posição com 18%.

5.2) Sensor preferido no geral

Apesar da classificação positiva com relação ao sensor *Leap Motion* no quesito facilidade, outro sensor foi o mais citado quando os usuários foram questionados "Qual dispositivo você mais gostou?" como demonstra a [Figura 21](#). O *smartphone* ficou em primeiro lugar com 65% dos usuários, seguido do *Leap Motion* com 30% e em terceiro o *Myo* com 5%, ao serem questionados sobre sua preferência por este sensor os usuários citaram diversos fatores, como fácil acesso de

compra, permitir todos os *feedbacks*, possibilidade de conexão estável com a Internet e baixo custo.

Figura 21 – Sensores preferidos dos usuários do grupo 2

5.3) Sensores e ações executadas no reprodutor de vídeo

5.3.1) *Facilidade entre os sensores e ações*: para uma compreensão mais profunda entre a relação dos sensores e as ações em níveis de facilidade e dificuldade, a [Tabela 8](#) permite uma análise comparativa dos resultados obtidos. A resposta "Todas" as facilidades foi a que mais se destacou utilizando o percentual comparativo dentro todos os sensores com um total de 53%, pode-se concluir que mais da metade do usuário considerou grande facilidade ao executar ações no reprodutor de vídeo por meio de gestos utilizando os sensores gestuais testados, a ação "Play" ficou em segundo lugar nesta classificação com 23% das respostas, o que deve ser considerado como ótimo resultado levando em conta que a ação de iniciar o vídeo pode ser entendida como mais importante dentre as ações propostas, a opção "Nenhuma" facilidade ocupa o terceiro lugar com 13% das respostas indicando que poucos usuários não perceberam nenhuma facilidade nos sensores de forma geral.

5.3.1) *Dificuldade entre os sensores e ações*: com relação a dificuldades relatadas de forma geral entre todos os sensores, a [Tabela 9](#) permite uma análise comparativa dos resultados obtidos. A resposta "Nenhuma" dificuldade obteve 59% do percentual comparativo dentro todos os sensores indicando que na grande maioria os usuários não consideraram nenhuma dificuldade ao utilizar os sensores, em segundo lugar com 14% dos usuários optaram encontraram muitas dificuldades no uso de todos os sensores. Apesar do ótimo resultado, todas as ações receberam alguma porcentagem de respostas indicando alguma dificuldade, se somadas representam 28% do percentual correspondente, isso indica que de forma geral os sensores apresentam alguma dificuldade de uso, porém, não maior que as facilidades que os sensores proporcionam na

Tabela 8 – Comparativo de facilidade entre sensores e ações executadas no reprodutor de vídeo

Facilidades	Smartphone	Webcam	Myo	Leap Motion	Percentual Correspondente
Todas	65%	25%	30%	90%	53%
Play	30%	35%	25%	-	23%
Nenhuma	-	35%	15%	-	13%
Aumentar Volume	-	-	10%	10%	5%
Pause	5%	5%	5%	-	4%
Avançar	-	-	10%	-	3%
Diminuir Volume	-	-	5%	-	1%
Retroceder	-	-		-	0%
Total	100%	100%	100%	100%	100%

acessibilidade de vídeo na Web por meio de gestos.

Tabela 9 – Comparativo de dificuldade entre sensores e ações executadas no reprodutor de vídeo

Dificuldades	Smartphone	Webcam	Myo	Leap Motion	Percentual Correspondente
Nenhuma	85%	20%	35%	95%	59%
Todas	-	40%	15%	-	14%
Play	5%	5%	20%	5%	9%
Aumentar Volume	-	20%	5%	-	6%
Retroceder	5%	5%	5%	-	4%
Diminuir Volume	-	-	15%	-	4%
Pause	-	5%	5%	-	3%
Avançar	5%	5%	-	-	3%
Total	100%	100%	100%	100%	100%

3.3.3 Sobre o uso de gestos

Após a realização do roteiro de atividades definida no protocolo de condução de testes, os usuários foram entrevistados sobre suas experiências ao executar interações gestuais, por meio dessa entrevista às seguintes variáveis foram coletadas: **interação gestual de forma geral, interação gestual com o reprodutor de vídeo, interação com o sensor gestual, personalização de gestos, uso de gestos no cotidiano e melhoria em acessibilidade**. O levantamento e discussão destas variáveis permite verificar como se deu a experiência dos participantes e se houveram melhorias na acessibilidade dos mesmos.

Com relação a variável **interação gestual de forma geral** no G1 22% dos usuários classificaram como sendo Muito Fácil o uso de gesto como forma de interação, 56% Fácil e 22% Difícil. No G2 os usuários classificaram a interação gestual de forma mais positiva, 33% responderam como sendo Muito Fácil e 67% Fácil não havendo outras classificações, para os 22% dos usuários que utilizaram a classificação Difícil no G1 foi questionado o motivo de suas respostas, todos alegaram ter dificuldades com o aprendizado de novas tecnologias e que apesar dessa classificação não se sentiram desmotivados a continuar um novo aprendizado.

Também lhes foi questionado sobre sua **interação gestual com o reprodutor de vídeo**, mais precisamente sobre sua classificação de acessar as funcionalidades do reprodutor de vídeo na Web por meio de gestos. Ambos os grupos ficaram próximos em suas classificações como demonstrado na [Figura 22](#), no G1 22% classificou como Muito Fácil, 56% Fácil, 11% como Nem Fácil e Nem Difícil e 11% como Difícil, no apresentou uma classificação levemente mais negativa onde 17% classificou como Muito Fácil, 50% Fácil, 17% como Nem Fácil e Nem Difícil e 17% como Difícil.

Figura 22 – Comparativo entre interações gestuais de forma geral e com o reprodutor de vídeo

De modo geral a classificação Fácil representada com a cor vermelha na [Figura 20](#) é predominante na opinião dos usuários e a classificação Muito Difícil em nenhum momento foi citada, o que motiva a investigação de interações gestuais como forma de ganho de acessibilidade. Outra análise está na comparação dos itens "G1 gestoGeral" e "G1 gestoTarefas" onde a quantidade de respostas Muito Fácil (56%) e Fácil (22%) é idêntica e a classificação Difícil (22%), representada pela cor verde, cai pela metade com o aumento da classificação Nem Fácil e Nem Difícil (11%), demonstrando assim um aumento positivo na experiência do usuário na interação gestual com o reprodutor de vídeo. Ao fazer a mesma análise com os dados do G2 nos itens "G2 gestoGeral" e "G2 gestoTarefas" nota-se o comportamento oposto ao G1 onde alguns usuários ao interagirem com as funções do reprodutor a classificaram como Nem Fácil e Nem Difícil e Difícil, diminuindo assim as outras classificações. Este comportamento não era esperado e leva ao questionamento da razão de ter ocorrido.

Uma possível hipótese para este comportamento encontrado pode estar no perfil que cada grupo possui, como citado anteriormente os usuários do G2 fazem 33% mais interações no reprodutor de vídeo que os usuários do G1, portanto estes usuários que atribuem bastante importância em poder interagir com o reprodutor de vídeo, ao encontrarem alguma dificuldade na interação gestual, podem ter classificado de forma mais negativa devido a importância que

este comportamento tem em seu estilo de interação. Para aqueles usuários do G1 que não tem o comportamento de fazer diversas interações no vídeo, encontrar uma dificuldade na interação gestual no reprodutor pode não parecer tão negativo quanto é para os usuários do G2. Para testar essa hipótese formula-se um teste de Chi-Quadrado e verifica-se se existe associação entre as variáveis *interageVídeo* (iV) e *Dificuldades* (D) da seguinte forma:

Hipóteses nula (H₀): não há associação entre as variáveis, ou seja, as respostas sobre realizar Interações no Vídeo e sentir Dificuldades são independentes.

Hipóteses alternativa (H_a): há associação entre as variáveis, ou seja, as respostas sobre realizar Interações no vídeo e sentir Dificuldades não são independentes.

O cálculo efetuado utilizou o nível de significância (alfa) de 5% (0.05) e os resultados encontram-se na [Tabela 10](#).

Tabela 10 – Interações no Vídeo (iV) x Dificuldades(D)

Chi-quadrado	G1-iV	G1-D	G2-iV	G2-D	Total
Sim	61%	61%	94%	100%	
Esperado	79.00	79.00	79.00	79.00	316
Observado	4.10	4.10	2.85	5.58	
Não	39%	39%	6%	0%	
Esperado	21.00	21.00	21.00	21.00	84
Observado	15.43	15.43	10.71	21.00	
Total	100	100	100	100	400
Resultado					
chi-quadrado:	79.204	Grau de liberdade:		3	
P-valor:	0.0001	Resultado:		Rejeita-se H ₀	

Devido o p-valor ser menor que o valor de significância ($0.0001 < 0.05$) rejeita-se a hipótese nula (H₀) de não associação, ou seja, existem evidências para afirmar que existe relação entre usuários que atribuem bastante importância em poder interagir com o reprodutor de vídeo, ao encontrarem alguma dificuldade na interação gestual, podem ter classificado de forma mais negativa devido a importância que este comportamento tem em seu estilo de interação.

Para melhor análise dos resultados as variáveis **personalização de gestos**, **uso de gestos no cotidiano** e **melhoria em acessibilidade** foram sumarizados na [Tabela 11](#) abaixo:

Tabela 11 – Dados sobre personalização, cotidiano e melhorias

Variáveis	G1 Sim	G1 Não		G2 Sim	G2 Não
personalizaGestos	78%	22%		33%	67%
usariaGestos	100%	0%		94%	6%
melhoraAcessibilidade	94%	6%		100%	0%

Com o objetivo de compreender melhor a experiência obtida com o uso da interação gestual os usuários foram questionados sobre personalização dos gestos, no G1 78% dos usuários

demonstraram vontade de poder personalizar os gestos de acordo com sua vontade e no G2 apenas 33% dos usuários demonstrou a mesma vontade. Uma possível correlação pode estar no perfil que cada grupo demonstrou com relação às dificuldades encontradas, os usuários do G1 demonstraram um maior percentual de dificuldades ao analisar a variável interação com o sensor gestual, o que poderia levar a compreender que para os usuários a possibilidade de criar um gesto personalizado poderia diminuir as dificuldades encontradas.

Com relação ao nível **uso de gestos no cotidiano** 94% dos usuários de ambos os grupos afirmaram que utilizariam interações gestuais em seu dia-a-dia, o mesmo ocorreu ao serem questionados sobre a percepção de melhoria na acessibilidade comparando com experiências anteriores e o método usual de acessar vídeos na Web ao qual estão acostumados. Apenas 6% dos usuários do G2 afirmaram negativamente sobre usar gestos em seu dia-a-dia e a mesma porcentagem de 6% afirmou no G1 não sentirem melhora na acessibilidade.

Devido a experiência de usuário para usuário ser única, para esta parcela de 6% de usuários foi questionado o motivo de suas resposta, um usuário afirmou que não teria interesse em usar interações gestuais devido ao alto valor que os sensores podem tem em sua opinião. Sobre não sentir melhorias na acessibilidade um usuário afirmou estar tão acostumado com o uso do teclado e do leitor de tela que seria mais rápido executar uma ação no modo usual comparado com a sua experiência no modo gestual.

Também foi observado que dados quantitativos poderiam ser coletados para uma uma análise mais abrangente com relação da interação entre usuários e gestos. Esta análise permite a possibilidade de obter de novas informações e conhecimento além dos já discutidos por meio da análise qualitativa realizada. As variáveis quantitativas consideradas para esta análise foram: **tarefas completadas com sucesso, número de erros cometidos e tempo gasto tentando efetuar uma tarefa.**

Todos os testes foram gravados com a autorização dos mesmos e a análise dos vídeos de cada interação permitiu a obtenção de valores quantitativos. A variável "tarefas completadas com sucesso" foi obtida por meio da contagem de quantas tarefas foram completadas de modo satisfatório, esta variável foi coletada para compreender se os usuários conseguiram executar as tarefas com maior ou menor grau de satisfação permitindo análises de melhoria na experiência de interação. Foi considerado como modo satisfatório à tarefa completada pelo usuário do início ao fim. Como exemplo, a tarefa 1 requitava que o usuário iniciasse o vídeo por meio da ação de *play*, todos os usuário que conseguiram realizar esta tarefa mesmo após algumas tentativas, foram consideradas como modo satisfatório.

Consequentemente a obtenção de tarefas completadas de modo insatisfatório, foi a obtenção de tarefas não completadas. Este valor permite compreender se os usuário tiveram maior ou menor grau de insatisfação ao realizar a interação. O valor de tarefas insatisfatórias foi obtido quando os usuários, mesmo após repetidas tentativas (por volta de 10), não conseguiam executar uma tarefa ou quando os usuário conseguiam executar a tarefas mas isso lhe demandava

muito esforço físico para executar os gestos.

Também foi obtido a variável "número de erros cometidos" durante a execução de cada tarefa por cada usuário, esta variável foi coletada para compreender se uma determinada tarefa executada por meios gestuais é mais ou menos difícil ao usuário e estabelecer uma relação entre melhorias na interação. Toda vez que um usuário executava um gesto mas não alcançava a resposta esperada, foi considerado como um erro cometido. Todos os usuário passaram pela fase de ambientação onde puderam livremente executar gestos como interação com o reprodutor, durante esta fase muitos erros foram cometidos como parte do processo de aprendizado e tais erros não foram considerados para a obtenção desta variável.

Por fim, outra variável coletada foi o "tempo gasto na tentativa de executar um gesto". Esta variável permite a análise de quanto tempo o usuário emprega durante a interação o que permite estabelecer relações de melhoria ao comparar o tempo gasto em interações sem a interação gestual. Os dados coletado foram agrupados em três classificações, sendo estas: menos que 2 segundos, entre 2 e 5 segundos e mais que 5 segundos. Esta classificação foi necessária devido a rapidez em que os gesto são realizados, em sua grande maioria o gesto leva apenas alguns segundos (menos de 5 segundos) para ser executado pelo usuário.

Após consideração sobre quais variáveis quantitativas foram coletadas e a motivação que cada uma possui, a seguir serão discutidos os resultados iniciando sobre a perspectiva de cada sensor gestual, isso devido ao fato que os gestos executados em cada sensor são diferentes entre si.

Na [Tabela 12](#) os dados quantitativos gestuais coletados por meio do sensor *Smartphone* são apresentados. Nota-se que o total de interações no modo satisfatório é de 95%, demonstrando que os gestos empregados nas tarefas por meio do *smartphone* propiciaram uma experiência bastante satisfatória aos usuários, esta afirmação pode ser confirmada ao observar o valor de erros cometidos e de tempo gasto. Apenas 16 foi o total de erros cometidos e os usuário executaram os gestos em menos que 2 segundos em 90% das tarefas. Uma das explicações sobre o ótimo resultado alcançado, está na familiaridade que os usuários já possuem com o mesmo, outro fator está no modo de captura gestual onde os usuário faziam movimentos de inclinar o sensor para executar os gestos.

Pode-se concluir que no quesito interações gestuais por meio do sensor *smartphone* os usuário obtiveram um ótimo desempenho sendo este sensor bem adequado para interações gestuais no contexto deste trabalho de pesquisa.

Na [Tabela 13](#) os dados quantitativos gestuais coletados por meio do sensor *Webcam* são apresentados. Nota-se que o total de interações no modo satisfatório é de 50%, demonstrando que os gestos empregados nas tarefas por meio da *webcam* não propiciaram uma experiência plenamente satisfatória aos usuários, esta afirmação pode ser confirmada ao observar o valor de erros cometidos e de tempo gasto, 377 foi o total de erros cometidos e os usuário executaram os

Tabela 12 – Dados quantitativos gestuais coletados por meio do sensor *Smartphone*

Smartphone	Interações			Tempo do gesto		
	Satisfatórias	Insatisfatórias	Erros cometidos	menos que 2s	entre 2 e 5s	mais que 5s
Tarefa 1	16	4	7	15	4	1
Tarefa 2	20	0	2	19	1	0
Tarefa 3	20	0	2	19	0	1
Tarefa 4	19	1	4	18	0	2
Tarefa 5	20	0	1	19	1	0
Percentual do Total	95%	5%	16	90%	6%	4%

gestos em mais que 5 segundos em 54% e entre 2 e 5 segundos em 29% do total das tarefas.

Pode-se concluir que dentre todos os sensores, a *webcam* foi o sensor menos satisfatório aos usuários, um dos fatores que pode explicar a grande porcentagem de erros cometidos esta no modo em que este sensor realiza a captura gestual. Por utilizar a captura de imagem para processamento dos gestos, os usuário em muitos casos não conseguiam saber se o gesto estava sendo executado dentro do campo visão que a *Webcam* possuía, após diversas tentativas mudando a posição no espaço onde o gesto era executado e recebendo o *feedback* da interação, conseguiam achar a posição ideal para a captura gestual. Em usuário videntes este problema pode ser facilmente corrigido permitindo que ele veja a imagem sendo capturada da *webcam* no próprio navegador, porém no contexto de melhorias de acessibilidade, este sensor demonstra impor mais barreiras do que a diminuição das mesmas.

Tabela 13 – Dados quantitativos gestuais coletados por meio do sensor *Webcam*

Webcam	Interações			Tempo do gesto		
	Satisfatórias	Insatisfatórias	Erros cometidos	menos que 2s	entre 2 e 5s	mais que 5s
Tarefa 1	13	7	68	4	4	12
Tarefa 2	10	10	62	2	7	11
Tarefa 3	9	11	78	5	6	9
Tarefa 4	8	12	82	4	6	10
Tarefa 5	10	10	87	2	6	12
Percentual do Total	50%	50%	377	17%	29%	54%

Na [Tabela 14](#) os dados quantitativos gestuais coletados por meio do sensor *Myo* são apresentados. Nota-se que o total de interações no modo satisfatório é de 69%, demonstrando que os gestos empregados nas tarefas por meio do *Myo* propiciaram uma experiência satisfatória aos usuários, esta afirmação pode ser confirmada ao observar o valor de erros cometidos e de tempo gasto, 64 foi o total de erros cometidos e os usuário executaram os gestos em menos que 2 segundos em 70% das tarefas. Apesar de muito erros cometidos, a satisfação ainda é alta e um dos motivos que pode explicar esta relação está no modo que este sensor realiza a captura do gestos. Por ser um bracelete, o *Myo* reconhece o padrão muscular do braço dos usuário, devido

cada usuário possuir uma anatomia diferente, em alguns casos, o sensor não foi capaz de realizar a correta leitura do padrão muscular, após análise dos vídeos dos usuários notou-se que nos casos de muita e pouca massa muscular o sensor não funcionava corretamente. Mesmo com muitos erros os usuário, após algumas tentativas, conseguiram realizar as tarefas e dentro de um tempo bem aceitável.

Pode-se concluir que no quesito interações gestuais por meio do sensor *Myo* os usuário obtiveram um bom desempenho sendo este sensor adequado para interações gestuais no contexto deste trabalho de pesquisa.

Tabela 14 – Dados quantitativos gestuais coletados por meio do sensor *Myo*

Myo Armband	Interações			Tempo do gesto		
	Satisfatórias	Insatisfatórias	Erros cometidos	menos que 2s	entre 2 e 5s	mais que 5s
Tarefa 1	14	6	7	15	5	0
Tarefa 2	15	5	10	15	4	1
Tarefa 3	13	7	17	13	3	4
Tarefa 4	14	6	10	14	2	4
Tarefa 5	13	7	20	13	3	4
Percentual do Total	69%	31%	64	70%	17%	13%

Na [Tabela 15](#) os dados quantitativos gestuais coletados por meio do sensor *Leap Motion* são apresentados. Nota-se que o total de interações no modo satisfatório é de 99%, demonstrando que os gestos empregados nas tarefas por meio do *Leap Motion* propiciaram uma experiência extremamente satisfatória aos usuários, esta afirmação pode ser confirmada ao observar o valor de erros cometidos e de tempo gasto. Apenas 16 foi o total de erros cometidos e os usuário executaram os gestos em menos que 2 segundos em 86% das tarefas. Uma das explicações sobre o excelente resultado alcançado, está na facilidade de executar gestos com este dispositivo. Os usuários apenas deveriam posicionar uma das mãos sobre o sensor, posicionado sobre a mesa, e com simples gestos executar a tarefa desejada.

Pode-se concluir que no quesito interações gestuais por meio do sensor *Leap Motion* os usuário obtiveram um excelente desempenho sendo este sensor muito adequado para interações gestuais no contexto deste trabalho de pesquisa.

Após a discussão dos resultados obtidos da análise dos dados quantitativos sobre a perspectiva da interação gestual, a seguir serão discutidos os resultados obtidos por meio da análise dos dados quantitativos sobre a perspectiva da interação usual.

Como citado na [Subseção 3.2.3](#) chamada de Coleta de Dados, os participantes foram convidados a realizar as tarefas determinadas no protocolo de condução de teste de dois modos, gestual e usual. O modo gestual compreende o uso dos sensores para executar os gestos e assim realizar as tarefas, o modo usual compreende o uso de soluções de acessibilidade usada atualmente como teclado e leitor de tela. Isso permitiu coletar os dados de cada participante em

Tabela 15 – Dados quantitativos gestuais coletados por meio do sensor Leap Motion

Leap Motion	Interações			Tempo do gesto		
	Satisfatórias	Insatisfatórias	Erros cometidos	menos que 2s	entre 2 e 5s	mais que 5s
Tarefa 1	19	1	4	19	0	1
Tarefa 2	20	0	2	15	4	1
Tarefa 3	20	0	4	18	2	0
Tarefa 4	20	0	3	17	1	2
Tarefa 5	20	0	3	17	1	2
Percentual do Total	99%	1%	16	86%	8%	6%

duas distintas situações, uma ao qual já estavam acostumado em seu dia a dia (usual) e uma nova maneira de interação (gestual).

Na [Tabela 16](#) os dados quantitativos usuais coletados por meio do uso de teclado e leitor de tela são apresentados. Nota-se que o total de interações no modo satisfatório é de 59%. Isso demonstra que a interação empregada nas tarefas pelo modo usual não proporciona uma experiência plenamente satisfatória aos usuários, esta afirmação pode ser confirmada ao observar o valor de erros cometidos e de tempo gasto, 74 foi o total de erros cometidos e os usuários executaram os gestos entre 2 e 5 segundos em 56% e mais que 5% em 44% do total das tarefas.

Tabela 16 – Dados quantitativos coletados por meio da interação usual realizada pelo usuário (leitores de tela e teclado)

Interação Usual	Interações			Tempo da interação		
	Satisfatórias	Insatisfatórias	Erros Cometidos	menos que 2s	entre 2s e 5s	mais que 5s
Tarefa 1	14	6	7	0	12	8
Tarefa 2	12	8	16	0	12	8
Tarefa 3	10	10	11	0	8	12
Tarefa 4	10	10	22	0	11	9
Tarefa 5	13	7	18	0	13	7
Percentual do Total	59%	41%	74	0%	56%	44%

Ao comparar tais resultados apresentados na [Tabela 16](#) com as tabelas relativas a cada sensor anteriormente apresentadas, nota-se que o quesito satisfação, em relação ao usos dos sensores gestuais, temos os seguintes resultados. O sensor *Myo* apresenta uma melhora de 10%, o sensor *Smartphone* apresenta uma melhora de 36% e o sensor *Leap Motion* apresenta uma melhora de 43% em relação a interação gestual. Apenas no uso do sensor *Webcam* a interação gestual apresentou uma melhora de 9%. Isso demonstra que o uso dos sensores, em três casos dos quatro analisados, conseguiu proporcionar melhor satisfação ao usuários.

No quesito erros cometidos, o mesmo ocorre com relação a satisfação, apenas o número de erros cometidos com a *Webcam* supera os erros cometidos na interação usual, nos outros casos os erros ficam abaixo do número constatado na interação usual. Com relação ao tempo gasto,

fica ainda mais evidente a melhoria de acessibilidade que o uso dos sensores gestuais oferece, na interação usual nenhum usuário conseguiu realizar uma tarefa em menos que 2 segundos.

Apesar da comparação da interação usual com cada sensor de forma individual demonstrar bons resultados de melhoria de acessibilidade, uma análise mais assertiva pode ser realizada por meio da comparação das médias dos resultados obtidos por cada sensor com os valores obtidos na interação gestual. Como cada sensor tem sua peculiaridade, a média desse valor permite uma compreensão mais general do que a interação gestual como um todo realmente representou aos usuários.

Na [Figura 23](#), o valor relativo a interação gestual (em azul) foi obtido calculando a média de satisfação de cada sensor e depois calculando a média das médias obtidas. Nota-se que a interação gestual obteve a satisfação de 82% e a interação usual 59%, uma melhor da 23% na satisfação dos usuários com o uso da opção gestual.

Figura 23 – Comparativo de tarefas completadas com sucesso (satisfatórias) entre interação usual e gestual

Na [Figura 24](#), o valor relativo a interação gestual (em azul) foi também obtido pelo cálculo da média de erros cometidos durante o uso dos quatro sensores e a média das médias obtidas. Nota-se que a interação gestual obteve uma portagem de erros de 34% e a interação usual 65%, ou seja, 31% menos erros cometidos com o uso da opção gestual.

Na [Figura 25](#), o valor relativo a interação gestual (em azul) assim como nos casos anteriores, também foi obtido pelo cálculo da média dos tempos gastos durante o uso dos quatro sensores e a média das médias obtidas. Nota-se que a interação gestual proporcionou ao usuário um ganho de tempo de 78% para acesso as tarefas no reprodutor com menos de 2 segundos.

Portanto, com base dos dados comparativos entre interação gestual e interação usual acima discutidos, o uso dos sensores gestuais permitiu que os usuários acessassem as tarefas de

Número de erros cometidos - Gestual x Usual

Figura 24 – Comparativo de número de erros cometidos entre interação usual e gestual

Tempo do gesto por tarefa - Gestual x Usual

Figura 25 – Comparativo do tempo de cada gesto por tarefa entre interação usual e gestual

modo mais rápido, com maior nível de satisfação e com menor nível de erros cometidos para realizar as tarefas na interface do reprodutor de vídeo.

3.3.4 Sobre o feedback utilizado

Durante a realização dos testes, foi observado direta e indiretamente que no contexto de interações gestuais com vídeos na Web o *feedback* possui grande importância sendo decisivo para o usuário compreender se o gesto executado resultou na ação pretendida na interface do reprodutor de vídeo. Na interação usual, onde o usuário utiliza leitor de tela e teclado, geralmente existem três *feedbacks* disponíveis para uso, sendo estes: o áudio do vídeo, que permite ao

usuário saber se o vídeo iniciou e se o ajuste do volume está funcionando devidamente, o visual que a interface do reprodutor possui quando existe uma interação, por exemplo, a animação que o botão de volume possui para indicar aumento porém só poderá ser utilizada por usuário que possuem alguma percepção visual e a voz que o leitor de tela possui, que realiza a leitura do texto que está em seleção. Quando o leitor de tela não consegue realizar a identificação de um botão, por exemplo, o usuário fica apenas dependente do áudio do vídeo ou de alguém que possa auxiliá-lo, isso em muitos casos inviabiliza a interação e esta é caracterizada como barreira de acessibilidade.

Com o objetivo de melhorar a acessibilidade, o *Gesture4all* possui uma camada em sua arquitetura que permite adicionar três tipos de *feedbacks* durante a interação com o reprodutor, sendo este: Sonoro, Visual e Tátil como discutido na [Seção 3.1](#). Esta camada não altera o funcionamento do *feedback* visual e audível que a interface do reprodutor de vídeo possui como característica. Para verificar se o *feedback* proporcionou melhorias de acessibilidade, as variáveis: **melhor *feedback*** e ***feedback* ideal** foram coletadas e sua análise ocorre a seguir.

Após a realização dos testes, os usuários foram questionados sobre qual tipo de *feedback* mais lhe ajudou durante a interação e qual seria o *feedback* ideal para a interação, essa análise permite entender qual *feedback* é mais útil para interações gestuais com o reprodutor de vídeo e se os *feedbacks* disponibilizados são ideais às necessidades dos usuários.

Na [Tabela 17](#), a coluna Áudio (Vídeo) representa o áudio original do vídeo usado nos testes, a coluna Áudio (Voz) representa a voz que o *framework* emite dizendo qual ação foi executada no reprodutor após execução do gesto, a coluna Visual (Cor) representa a cor que o *framework* reproduz no *background* do navegador e a coluna Tátil (Vibração) representa a vibração que alguns sensores permitem usar. O melhor *feedback*, segundo os usuários, é a voz que o *framework* possui com 80% das respostas, em segundo lugar a vibração como *feedback* recebeu 13,8% das respostas dos usuários.

Tabela 17 – Dados sobre o melhor *feedback* na interação gestual

Feedback	Áudio (Vídeo)	Áudio (Voz)	Visual (Cor)	Tátil (Vibração)
Smartphone	0	15	0	5
Webcam	2	17	1	0
Myo	0	13	1	6
Leap Motion	1	19	0	0
Total	3.8%	80%	2.5%	13.8%

Diferente do modo usual onde o leitor de tela reproduz o texto em forma de voz, o *framework* apenas reproduz a voz, dizendo ao usuário qual ação foi executada, após a confirmação que a interface do reprodutor recebeu a ação correta, dessa forma o usuário sabe que o gesto capturado é reconhecido de forma correta.

Aos usuários foi questionado, após responderem qual foi o melhor *feedback* para cada

sensor, qual seria o *feedback* ideal na interação gestual. Aos usuários foram oferecidas três opções de resposta, sendo elas: todos os *feedbacks*, apenas o áudio original do vídeo ou apenas um *feedback* a sua escolha, esta última opção como resposta aberta teve o objetivo de explorar outras formas de *feedback* que pudessem ser citadas pelos usuários.

Na [Tabela 18](#), os dados sobre o *feedback* ideal é são apresentados. A resposta todos os *feedbacks* foi citada por 99% dos usuário, apenas um usuário teve preferencia por apenas a vibração como *feedback*.

Tabela 18 – Dados sobre o *feedback* ideal na interação gestual

Feedback	Todos	Somente Vibração
Smartphone	19	1
Webcam	20	-
Myo	20	0
Leap Motion	20	-
Total	99%	1%

Apesar do *feedback* Áudio (Voz) ter recebido 80% das respostas dos usuários, nota-se que a importância que quase todos os usuários entendem que quanto mais opções de *feedback* existirem mais completo e ideal a solução se apresenta. Devido este motivo, o *framework* permite ao desenvolvedor personalizar os *feedbacks* podendo alterar o som emitido a cada ação executada no reprodutor assim como o volume deste som, o comportamento das cores no *background* possibilitando alterar a cor em si, a intensidade do brilho e se ela irá piscar ou ficará presente e a intensidade e frequência da vibração. Essa flexibilidade permite atender as características de cada indivíduo para uma melhoria na acessibilidade por meio de uma experiência mais personalizada.

3.4 Considerações finais

Sobre o perfil dos usuários a grande maioria demonstrou interesse no uso cotidiano de sensores e interações gestuais, o que motiva o desenvolvimento de novas soluções envolvendo as área de interações gestuais e acessibilidade, realizar o estudo de caso com grupos diferentes de usuários e com perfis diferentes também auxiliou para a validação desta conclusão uma vez que a taxa de amostragem se apresentou heterogênea em ambos os grupos.

Outra consideração envolve o uso de programas de leitura, interações no vídeo e dificuldade enfrentadas. Usuários com maior deficiência visual sentem menos dificuldades ou barreiras em interações no vídeo por sentirem mais necessidade de possuir o conhecimento de uma tecnologia assistiva ao interagir com a Web. Já usuários com menor deficiência visual não se sentem totalmente obrigados a procurar o conhecimento de leitores de tela. Conclui-se então que usuários mesmo possuindo maior acuidade visual poderiam ter maiores ganhos de acessibilidade utilizando leitores de tela para interação com vídeos na Web.

Os resultados obtidos nesta seção também indica que reconhecer gestos realizados por sensores (**Interações com gestos em três dimensões**), aplicando a semântica estabelecida para cada gesto e possibilitar a interação diretamente na interface (**Controle de interface por meio de sensores**) e prover meios alternativos de *feedback* como tátil e sonoro, permitiu que os usuários interagissem com qualquer interface, mesmo aquelas que não possuam tal suporte. Tais estratégias adotadas pelo *Gesture4All* possibilitaram melhorias de usabilidade de três modos:

Interações com gestos em três dimensões: a primeira melhoria está em utilizar o conceito de naturalidade presente em interações em três dimensões onde o usuário pode utilizar seu corpo, algo que já utiliza desde seu nascimento, para criar uma interação mais personalizada. Realizar gestos com as mãos e os braços não depende de estímulos visuais para acontecer, diferente de pressionar um botão em uma interface Web que exige que o usuário ou localize visualmente e faça a interação com o *mouse* ou localize por meio de leitores de tela e teclado. Outro fator é que gestos podem ser mais facilmente lembrados do que procedimentos de interações convencionais (ASHBROOK; STARNER, 2010).

Controle de interface por meio de sensores: a segunda melhoria ocorre quando o *Gesture4All* provê a semântica do gesto para a interface que recebe a instrução do que deverá fazer diretamente por chamada de função, ou seja, a interface mesmo sendo criada para uso somente com o *mouse*, como exemplificado no cenário anteriormente citado, recebe as instruções do que deverá fazer (controlar o volume, por exemplo) diretamente do *framework*. No exemplo de reprodutores de vídeos criados utilizando a linguagem HTML 5, é possível realizar a chamada de função "*myVideo.play();*" em *JavaScript* para reproduzir um vídeo ou a chamada de função "*myVideo.pause();*" para pausar o vídeo, a chamada de função é executada independentemente da localização do reprodutor ou dos controles da interface do reprodutor na aplicação Web.

Feedback: a terceira melhoria está na adição de *feedbacks* auxiliares e permitindo a sua modificação. Devido reprodutores de vídeos Web geralmente focarem no *feedback* visual para seus usuários, adicionar outros meios de *feedback* pode aumentar a possibilidade do usuário em saber se sua ação obteve o resultado esperado. No contexto de pessoas com baixa visão, este item é ainda mais necessário devido estes usuários terem outros sentidos mais aguçados, como o tato (vibração) e a audição (sons adicionais).

Essas melhorias de usabilidade estão ligadas diretamente com o ganho de acessibilidade, o que será discutido no [Capítulo 4](#).

CONCLUSÕES

Consumir e compartilhar vídeos diariamente tem se tornado uma prática comum entre os usuários da Web. Fatores como redes sociais, plataformas gratuitas de compartilhamento de vídeo, mais acesso a infraestruturas de banda larga e entre outros fatores contribuem para esta prática. Com este crescente aumento, questões de acessibilidade tem se tornado cada vez mais frequentes, uma vez ao existir interesse por parte dos usuários em consumir o conteúdo do vídeo na Web, deve-se assegurar que não existam barreiras que os impeçam de serem independentes e capazes de interagir livremente com a tecnologia. Neste trabalho, buscou-se a investigação sobre melhorias de acessibilidade por meio de interfaces gestuais para usuários com deficiência visual quando estes acessam vídeos na Web.

Segundo [Petrie, Savva e Power \(2015\)](#), acessibilidade na Web pode ser definida como a possibilidade de todas as pessoas, particularmente pessoas com deficiência, poderem usar sites em uma variedade de contextos de uso, para isso os sites precisam ser projetados e desenvolvidos para apoiar a usabilidade nesses contextos. Pode-se assumir então a premissa de que a usabilidade influencia diretamente na acessibilidade na Web e, portanto, deve estar presente em soluções que pretendem melhorar o nível de acesso dos usuários na Web. A usabilidade, por sua vez, pode ser definida como: o grau em que um produto pode ser usado por usuários específicos para atingir objetivos específicos com **efetividade**, **eficiência** e **satisfação** em um contexto específico de uso ([ISO/IEC25066:2016, 2016](#)). Assim, em conformidade com a definição de [Barbosa e Silva \(2010\)](#) e [ISO/9241-210:2010 \(2010\)](#), tem-se que a efetividade consiste na capacidade com que os usuários interagem com o sistema para alcançar seus objetivos de forma correta, a eficiência corresponde aos recursos necessários para os usuários interagirem com o sistema para alcançar seus objetivos e a satisfação corresponde ao grau de aceitação e conforto usuários ao utilizar o sistema no contexto de uso para o qual foi projetado.

A partir destas definições e após análise do estudo de caso realizado, pode-se afirmar que existem fortes indícios de melhoria em acessibilidade para usuário com deficiência visual

na interação com um reprodutor de vídeo na Web por meio do ganho de efetividade, eficiência e satisfação para os usuários. Para sustentar essa afirmação, a análise das variáveis coletadas e discutidas na [Seção 3.3](#), resumidamente, indicam que: a satisfação dos usuários foi de 82% ao realizar as tarefas, apresentando uma melhora de 23% ao comparar-se a interação usual, a efetividade dos usuários foi de 34% ao realizar as tarefas e cometer menos erros, apresentando uma melhora de 31% nos erros cometidos ao comparar-se a interação usual e a eficiência dos usuários foi de 78% com o ganho de tempo ao realizar as tarefas, apresentando uma melhora de 100% para acesso as tarefas no reprodutor com menos de 2 segundos. Estes resultados validam as respostas dos usuários sobre sentir ou não algum tipo de melhoria na acessibilidade após a realização dos testes. Segundo os mesmos responderam durante a realização das entrevistas prevista no protocolo utilizado, no G1 94% e no G2 89% preferem a interação gestual do que a interação usual.

Outro indício de melhoria de acessibilidade esteve na diminuição de barreiras de acessibilidade proporcionada pelo uso de interações gestuais no contexto apresentado neste trabalho. Na interação com o reprodutor de vídeo no modo usual (teclado e leitor de tela) 41% das tarefas propostas aos usuários não foram executadas ou foram executadas de modo insatisfatório (após muitas tentativas), ou seja, para estes usuários completar a tarefa com o uso de teclado e mouse ainda é uma barreira pois impede o acesso ao conteúdo de vídeo ao qual desejam consumir. Com o uso de interações gestuais, apenas 18% das tarefas propostas aos usuários não foram executadas ou foram executadas de modo insatisfatório, ou seja, houve uma melhoria de acessibilidade de 23% com o uso de interações gestuais.

Por outro lado, apesar do ganho em acessibilidade fornecido pela interação gestual, os usuários podem estar sujeitos a novas barreiras por meio deste modo de interação. Uma destas barreiras pode ser observada no uso do sensor gestual *Myo*, no G1 72% e no G2 67% dos usuários afirmaram sentir algum tipo de dificuldade ao interagir com este sensor. Essa dificuldade ocorre por dois motivos, o primeiro consiste na necessidade de vestir o sensor no braço para captura de gestos, braços menores ou maiores que o diâmetro do sensor impossibilitam sua utilização. Outra barreira está na necessidade do controle das funções motoras para a execução dos gestos, usuários que possuem alguma deficiência motora e estão limitados a movimentos mais contidos, terão dificuldades ou não serão capazes de executar os gestos necessários para a interação gestual, caracterizando assim mais uma possível barreira.

Sobre o perfil dos usuários participantes do estudo de caso, com relação ao acesso a vídeos na Web, no G1 94% e no G2 100% dos usuários acessam vídeos na Web. Este interesse pode ser confirmado ao analisar-se a frequência de acesso a vídeos, no G1 78% e no G2 56% acessam vídeos diariamente. Com relação a enfrentar barreiras ou dificuldades durante o acesso a estes vídeos, no G1 61% e no G2 100% responderam que sim. Estas análises permitem reafirmar a necessidade na busca de melhoria de acessibilidade e a importância do objetivo proposto neste trabalho.

Sobre o uso de sensores gestuais, os usuários classificaram o sensor *Leap Motion* como sensor gestual muito fácil de ser utilizado recebendo 70% das respostas, em segundo lugar o *Smartphone* com 50% e em terceiro o *Myo* com 20%. Também foi classificado pelos usuários o sensor gestual preferido após a realização dos testes, *Smartphone* ficou em primeiro lugar com 65% dos usuários, seguido do *Leap Motion* com 30% e em terceiro o *Myo* com 5%. Devido os sensores terem alcançados porcentagens de facilidade de uso elevadas, pode-se concluir que os sensores gestuais proveram maior satisfação para os usuários realizarem as tarefas no reprodutor de vídeo melhorando assim a acessibilidade dos mesmo.

Sobre o uso de gestos, os usuários no G1 classificaram como sendo fácil o uso de gestos de forma geral (sem ser necessariamente usado no contexto de vídeos na Web) em 78% das respostas, no G2 a mesma análise obteve 100% dos casos. Para o uso específico para interações com um reprodutor de vídeo na Web, os usuários no G1 classificaram como sendo fácil em 78% das respostas, ou seja, não havendo diferença com o resultado anterior e o G2 apresentou uma classificação levemente mais negativa onde 67% ao realizar a interação gestual especificamente no reprodutor de vídeo na Web. Devido os gestos também terem alcançados porcentagens de facilidade de uso elevadas, pode-se concluir que o uso de gestos proveram maior satisfação para os usuários realizarem as tarefas no reprodutor de vídeo melhorando assim a acessibilidade dos mesmo.

Sobre o *feedback*, os usuários apontaram o áudio adicional realizado pelo *framework* onde uma voz diz o nome do ação que aquele gesto executada conseguiu realizar em 80% das respostas seguido vibração com 13.8% das respostas. Ao serem questionados sobre qual seria o modelo ideal de *feedback* para o contexto dos testes ao qual participaram, 99% afirmaram que a possibilidade ter a sua disposição todos os *feedback* é o modo ideal de interação. Isso demonstra a importância que o uso de sensores gestuais possuam a opção de vibração como *feedback* existe um ganho na satisfação e percepção do usuário no acesso a vídeos na Web.

4.1 Principais contribuições

A principal contribuição deste trabalho está na obtenção dos indícios de melhoria e diminuição de barreiras de acessibilidade para usuários com deficiência visual na interação com um reprodutor de vídeo na Web, conforme discutidos anteriormente. Ademais, destacam-se outras contribuições como avanço do conhecimento do grupo de pesquisa pertencente ao laboratório Intermídia do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, que possui muito interesse na avanços de áreas como Web e Acessibilidade.

Destaca-se também o desenvolvimento de um *framework* Web (*Gesture4All*) e um protótipo de aplicação Web com suporte a interação gestual em vídeos. Por meio do *framework* foi possível a prototipação de uma página Web contendo suporte a quatro diferentes sensores gestuais e um reprodutor de vídeo. Somente a partir deste protótipo desenvolvido o estudo de

estudo de caso pode ser realizado para a obtenção dos indícios supracitados.

Durante o período deste mestrado, foram publicados os seguintes artigos científicos:

- **Artigo publicado no Webmedia 2016:** *Gestural Interaction for Accessibility of Web Videos*. Artigo que descreve um estudo de caso realizado com o grupo 1, objetivando levantar indícios sobre o uso de interações gestuais para acessibilidade de vídeos na Web (FUNES; GOULARTE, 2016).
- **Artigo publicado no Webmedia 2017** *Gestural Interaction for Accessibility of Web Videos: A Case Study Analysis with Blind and Visually Impaired Users*. Artigo que descreve um estudo de caso realizado com o grupo 2, objetivando levantar indícios sobre o uso de interações gestuais para acessibilidade de vídeos na Web (FUNES; FORTES; GOULARTE, 2017).
- **Artigo publicado no SAC 2018** *Gesture4All: a framework for 3D gestural interaction to improve accessibility of Web videos*. Artigo que descreve o desenvolvimento do *framework Gesture4All* e a condução do estudo de caso com o grupo 2, objetivando levantar indícios sobre o uso de interações gestuais para acessibilidade de vídeos na Web (FUNES *et al.*, 2018).

4.2 Limitações

Durante o período de desenvolvimento deste trabalho procurou-se seguir o rigor científico. Contudo, existem algumas limitações que derivam do escopo da proposta, metodologia utilizada para avaliação, tempo para realização, tecnologias utilizadas, entre outras variáveis.

Nesse sentido, a interface gráfica para a aplicação Web desenvolvida para viabilização dos testes no estudo de caso não teve prioridade no desenvolvimento, uma vez que optou-se por uma interface mais simples e com suporte a acessibilidade na Web. Além disso, a aplicação Web não ofereceu escolha do conteúdo do vídeo oferecido aos usuários para realização das tarefas.

Com relação a arquitetura em camadas planejada para o *framework Gesture4All*, apenas foram desenvolvidas as camadas e módulos indispensáveis para viabilização do estudo de caso. Além disso, devido objetivo do trabalho estar em melhorar a acessibilidade por meio de gestos, não foram desenvolvidos novas soluções de reconhecimento gestual ou feitas tentativas para melhorar as técnicas já presentes no estado da arte. O uso de soluções já existentes permitiu o foco na realização do estudo de caso.

4.3 Trabalhos futuros

Este trabalho apresentou uma investigação focada no levantamento de indícios sobre melhorias de acessibilidade por meio de interações gestuais em reprodutores de vídeo na Web. Uma vez constatado que tais indícios existem, novas pesquisas utilizando interações gestuais e acessibilidade podem ser realizadas.

Nesse sentido, a realização deste trabalho propiciou diversas possibilidades para realização de trabalho futuros que possam avançar o estado da arte na área de acessibilidade e interações baseadas em gestos. Entre as possibilidades de trabalhos futuros, pode-se destacar:

- **Expansão do estudo de caso:** o estudo de caso pode ser conduzido com um número maior de sujeitos e com perfis diferenciados, a fim de aumentar a confiabilidade e abrangência dos resultados destacados neste trabalho. Apesar da análise do perfil por participantes indicar abrangência de acuidade visuais, gênero e idade a repetição do estudo de caso e a comparação de resultados pode aumentar a confiabilidade dos resultados obtidos e a percepção de outras possíveis investigações antes não mapeadas.
- **Uso de novos sensores gestuais em estudo de casos:** para esta pesquisa de mestrado foram investigados os sensores gestuais já citados no estado da arte e selecionados aqueles que ofereciam diferentes modos de interação gestual. Essa escolha se deu devido o interesse em fornecer ao usuário diferentes modos de interação gestual e comparar sua experiência em cada sensor, a exploração de novos sensores gestuais poderá ampliar os indícios de melhorias em acessibilidade encontrados e obter a percepção de outras possibilidades de investigações antes não encontradas.
- **Interação com outros elementos na Web:** esta pesquisa de mestrado se limitou-se a interação com reprodutores de vídeos na Web devido o aumento do interesse do usuário em consumir esta mídia e estudos no estado da arte indicarem a presença de barreira nesse contexto. Porém, a interação com outros elementos na Web podem ser investigados para verificar se em outro contexto, não apenas o de vídeos, a melhora de acessibilidade também ocorre. Elementos como menus, reprodutores de áudio, jogos, formulários e etc, bem como a navegação de páginas também podem ser investigados para viabilização de interações gestuais e fornecer ao usuário alternativas visando a melhoria de acessibilidade.
- **Personalização dos gestos:** para que fosse possível à análise comparativa dos resultados obtidos no estudo de caso, todos os usuários executaram gestos predefinidos no desenvolvimento do protótipo utilizado no estudo. Porém, uma possibilidade levantada durante o desenvolvimento da pesquisa foi o de permitir que os usuário personalizassem os gestos de acordo com suas preferência, esta possibilidade não foi explorada devido o objetivo do trabalho se limitar no levantamento de indícios que indicassem ganhos no uso de gestos no contexto de vídeos na Web. Uma possível investigação está em permitir que os usuário

personalizem as ações no reprodutor de vídeo, essa personalização pode melhorar ainda mais os resultados de melhorias em acessibilidade.

- **Aprendizado de máquina no contexto gestual:** outra investigação levantada durante a condução da pesquisa, esteve no uso de inteligência artificial para que um sistema fosse capaz de aprender o modo que um usuário performa um gesto como tentativa de melhorar a assertividade do sistema em reconhecer os gestos realizados. Durante a pesquisa, um *dataset* contendo 7 diferentes tipos de gestos foi criado e utilizando algoritmos de Deep Learning alguns testes foram conduzidos para verificar essa possibilidade. Foi constatado que existe muito potencial de exploração dessa possibilidade porém, para ser explorado em novas pesquisas, pois demandaria maior aprofundamento do tema.
- **A continuação do desenvolvimento do *framework Gesture4All*:** ainda existem outros desafios no desenvolvimento do *framework* proposto neste trabalho de mestrado. Um deles está na continuação do desenvolvimento de sua arquitetura em camada, possibilitante a adição de novos módulos, como o de personalização ou de aprendizado gestual, possibilitando o desenvolvimento de páginas Web mais completas ao oferecerem interação gestual aos seus usuários.

Buscando incentivar a continuação dos estudos realizados neste mestrado, o *framework* desenvolvido será disponibilizado como *open source* para que futuros desenvolvedores venham a contribuir com a evolução da acessibilidade na Web.

REFERÊNCIAS

ABASCAL, J.; AIZPURUA, A.; CEARRETA, I.; GAMECHO, B.; GARAY-VITORIA, N.; n6N, R. M. Automatically generating tailored accessible user interfaces for ubiquitous services. In: **The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: ACM, 2011. (ASSETS '11), p. 187–194. ISBN 978-1-4503-0920-2. Disponível em: <<http://doi.acm.org/10.1145/2049536.2049570>>. Citado na página 59.

ABOWD, G. D.; MYNATT, E. D. Charting past, present, and future research in ubiquitous computing. **ACM Trans. Comput.-Hum. Interact.**, ACM, New York, NY, USA, v. 7, n. 1, p. 29–58, mar. 2000. ISSN 1073-0516. Disponível em: <<http://doi.acm.org/10.1145/344949.344988>>. Citado nas páginas 38, 39, 40 e 44.

AFYOUNI, I.; QAMAR, A. M.; HUSSAIN, S. O.; REHMAN, F. U.; SADIQ, B.; MURAD, A. Motion-based serious games for hand assistive rehabilitation. In: **Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion**. New York, NY, USA: ACM, 2017. (IUI '17 Companion), p. 133–136. ISBN 978-1-4503-4893-5. Disponível em: <<http://doi.acm.org/10.1145/3030024.3040977>>. Citado na página 68.

AIZPURUA, A.; ARRUE, M.; VIGO, M. Uncovering the role of expectations on perceived web accessibility. In: **Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: ACM, 2013. (ASSETS '13), p. 74:1–74:2. ISBN 978-1-4503-2405-2. Disponível em: <<http://doi.acm.org/10.1145/2513383.2513411>>. Citado nas páginas 34 e 54.

ALBINSSON, P.-A.; ZHAI, S. High precision touch screen interaction. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2003. (CHI '03), p. 105–112. ISBN 1-58113-630-7. Disponível em: <<http://doi.acm.org/10.1145/642611.642631>>. Citado na página 41.

ASHBROOK, D.; STARNER, T. Magic: A motion gesture design tool. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2010. (CHI '10), p. 2159–2168. ISBN 978-1-60558-929-9. Disponível em: <<http://doi.acm.org/10.1145/1753326.1753653>>. Citado nas páginas 35, 51 e 102.

ASUS. **Usuário realizando um gesto tridimensional com o Asus Xtion**. 2015. Disponível em: <https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/>. Citado na página 48.

BAKER, J. The dragon system—an overview. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 23, n. 1, p. 24–29, Feb 1975. ISSN 0096-3518. Citado na página 42.

BARBOSA, S.; SILVA, B. **Interação Humano-Computador**. Elsevier Brasil, 2010. ISBN 9788535211207. Disponível em: <https://books.google.com.br/books?id=qk0skwr_cewC>. Citado nas páginas 73 e 103.

- BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state markov chains. **Ann. Math. Statist.**, The Institute of Mathematical Statistics, v. 37, n. 6, p. 1554–1563, 12 1966. Disponível em: <<http://dx.doi.org/10.1214/aoms/1177699147>>. Citado na página 53.
- BONOMI, A. G.; PLASQUI, G.; GORIS, A. H. C.; WESTERTERP, K. R. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. **Journal of Applied Physiology**, American Physiological Society, v. 107, n. 3, p. 655–661, 2009. ISSN 8750-7587. Disponível em: <<http://jap.physiology.org/content/107/3/655>>. Citado na página 45.
- BORODIN, Y.; BIGHAM, J. P.; DAUSCH, G.; RAMAKRISHNAN, I. V. More than meets the eye: A survey of screen-reader browsing strategies. In: **Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)**. New York, NY, USA: ACM, 2010. (W4A '10), p. 13:1–13:10. ISBN 978-1-4503-0045-2. Disponível em: <<http://doi.acm.org/10.1145/1805986.1806005>>. Citado nas páginas 33 e 54.
- BORODIN, Y.; PUZIS, Y.; SOVIK, A.; BOUKER, J.; FENG, B.; SICOLI, R.; MELNYK, A.; MELNYK, V.; ASHOK, V.; DAUSCH, G.; RAMAKRISHNAN, I. V. Listen to everything you want to read with capti narrator. In: **Proceedings of the 11th Web for All Conference**. New York, NY, USA: ACM, 2014. (W4A '14), p. 33:1–33:2. ISBN 978-1-4503-2651-3. Disponível em: <<http://doi.acm.org/10.1145/2596695.2596728>>. Citado na página 59.
- BRAFTON. **By 2017, 74% of Web traffic will come from video**. 2015. Disponível em: <<https://www.brafton.com/news/2017-74-web-traffic-will-come-video/>>. Citado nas páginas 31 e 32.
- BROTHERTON, J.; ABOWD, G.; TRUONG, K. Supporting capture and access interfaces for informal and opportunistic meetings. **Georgia Institute of Technology Technical Report: GIT-GVU-99-06**. ACM, Atlanta, GA, USA, 1998. Disponível em: <<http://hdl.handle.net/1853/3373>>. Citado na página 38.
- BUSSINESINSIDER. **Exemplo de estrutura de uma Interface Baseada em Gestos**. 2015. Disponível em: <<http://uk.businessinsider.com/apple-gesture-control-patent-2015-1>>. Citado na página 45.
- CARTER, M.; NEWN, J.; VELLOSO, E.; VETERE, F. Remote gaze and gesture tracking on the microsoft kinect: Investigating the role of feedback. In: **Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction**. New York, NY, USA: ACM, 2015. (OzCHI '15), p. 167–176. ISBN 978-1-4503-3673-4. Disponível em: <<http://doi.acm.org/10.1145/2838739.2838778>>. Citado nas páginas 47 e 60.
- CARVALHO, J. O. F. d. O papel da interação humano-computador na inclusão digital. **Tran-sinformação**, scielo, v. 15, p. 75 – 89, 12 2003. ISSN 0103-3786. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862003000500004&nrm=iso>. Citado nas páginas 34 e 54.
- CARVALHO, J. R. **Uma abordagem para interoperabilização de dados de acelerômetros em aplicações interativas**. Tese (Tese (Mestrado em Ciências de Computação e Matemática Computacional)) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-24062013-153548/pt-br.php>>. Citado na página 45.
- CENTER, N. M. **About Netflix**. 2017. Disponível em: <<https://media.netflix.com/en/about-netflix>>. Citado na página 32.

CHAMPIN, P.-A.; ENCELLE, B.; EVANS, N. W. D.; O.-BELDAME, M.; PRIÉ, Y.; TRONCY, R. Towards collaborative annotation for video accessibility. In: **Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)**. New York, NY, USA: ACM, 2010. (W4A '10), p. 17:1–17:4. ISBN 978-1-4503-0045-2. Disponível em: <<http://doi.acm.org/10.1145/1805986.1806010>>. Citado na página 55.

CHEN, M.-Y.; KUNDU, A.; ZHOU, J. Off-line handwritten word recognition using a hidden markov model type stochastic network. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 16, n. 5, p. 481–496, May 1994. ISSN 0162-8828. Citado na página 53.

CISCO. **White paper: Cisco VNI Forecast and Methodology, 2015-2020**. 2016. Document ID:1465272001663118. Disponível em: <<http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>>. Citado na página 31.

CLICKBANK. **5 Video Marketing Trends Every Digital Marketer Needs To Know**. 2017. Disponível em: <<http://www.clickbank.com/5-video-marketing-trends-every-digital-marketer-needs-to-know/>>. Citado na página 31.

DAVIS, K. H.; BIDDULPH, R.; BALASHEK, S. Automatic recognition of spoken digits. **The journal of the acoustical society of america**, v. 24, 1959. Citado na página 42.

DESIGNBEEP. **Exemplo de apresentação gráfica do gesto a ser realizado**. 2015. Disponível em: <<http://designbeep.com/2013/09/06/10-sets-free-gesture-icons-promotedemonstrate-design-multi-touch-interfaces/>>. Citado na página 50.

DESKENG. **Usuário realizando um gesto tridimensional e visualizando o feedback da ação**. 2015. Disponível em: <<http://www.deskeng.com/de/gesture-based-design-engineering/>>. Citado na página 52.

DEY, A. K. Understanding and using context. **Personal Ubiquitous Comput.**, Springer-Verlag, London, UK, UK, v. 5, n. 1, p. 4–7, jan. 2001. ISSN 1617-4909. Disponível em: <<http://dx.doi.org/10.1007/s007790170019>>. Citado na página 39.

DIAO, W.; LIU, X.; ZHOU, Z.; ZHANG, K. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In: **Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones; Mobile Devices**. New York, NY, USA: ACM, 2014. (SPSM '14), p. 63–74. ISBN 978-1-4503-3155-5. Disponível em: <<http://doi.acm.org/10.1145/2666620.2666623>>. Citado na página 42.

DIAS, D.; BRANDÃO, A.; BRASIL, G.; TREVELIN, L. Gesture chess - interface natural de usuário na atividade motora e cognitiva. **IX WORKSHOP DE REALIDADE VIRTUAL E AUMENTADA**, v. 1, 2012. Citado na página 60.

DJEDDI, A.; DJILALI, I. A user centered ubiquitous government design framework. In: **Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication**. New York, NY, USA: ACM, 2015. (IPAC '15), p. 13:1–13:5. ISBN 978-1-4503-3458-7. Disponível em: <<http://doi.acm.org/10.1145/2816839.2816860>>. Citado na página 59.

DOUBLESHOT. **Video Marketing Trends Guide 2016/2017**. 2016. Disponível em: <<http://doubleshotvideo.co.uk/wp-content/uploads/2016/08/Video-Trends-2016-2017.pdf>>. Citado nas páginas 31 e 32.

EMAG. **Recomendações de acessibilidade para a construção e adaptação de conteúdos do governo brasileiro na Internet. eMAG, Modelo de acessibilidade do Governo Eletrônico, Versão 2.0.**, 2005. Disponível em: <<http://www.consultas.governoeletronico.gov.br/ConsultasPublicas/consultas.do;jsessionid=C13E5697FF43725DB0020A8BD15E77AB?acao=exibir&id=17>>. Citado na página 54.

EVAIN; OOL; ARGELAGUET, F.; STROCK, A.; ROUSSEL, N.; CASIEZ, G.; LÉCUYER, A. Influence of error rate on frustration of bci users. In: **Proceedings of the International Working Conference on Advanced Visual Interfaces**. New York, NY, USA: ACM, 2016. (AVI '16), p. 248–251. ISBN 978-1-4503-4131-8. Disponível em: <<http://doi.acm.org/10.1145/2909132.2909278>>. Citado na página 43.

FEDERAL, S. **lei nº 10.098, de 19 de dezembro de 2000**. 2000. Acessado em 15-01-2016. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L10098.htm>. Citado na página 53.

FREIRE, A. P.; LARA, S. M. A. de; FORTES, R. P. de M. Avaliação da acessibilidade de websites por usuários com deficiência. In: **Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems**. Porto Alegre, Brazil, Brazil: Brazilian Computer Society, 2013. (IHC '13), p. 348–351. ISBN 978-85-7669-278-2. Disponível em: <<http://dl.acm.org/citation.cfm?id=2577101.2577198>>. Citado nas páginas 33, 54, 63, 71 e 72.

FUNES, M. M.; FORTES, R. P.; GOULARTE, R. Gestural interaction for accessibility of web videos: A case study analysis with blind and visually impaired users. In: **Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web**. New York, NY, USA: ACM, 2017. (WebMedia '17), p. 501–504. ISBN 978-1-4503-5096-9. Disponível em: <<http://doi.acm.org/10.1145/3126858.3131588>>. Citado na página 106.

FUNES, M. M.; GOULARTE, R. Gestural interaction for accessibility of web videos. In: **Proceedings of the 22Nd Brazilian Symposium on Multimedia and the Web**. New York, NY, USA: ACM, 2016. (Webmedia '16), p. 343–346. ISBN 978-1-4503-4512-5. Disponível em: <<http://doi.acm.org/10.1145/2976796.2988212>>. Citado na página 106.

FUNES, M. M.; TROJAHN, T. H.; FORTES, R. P. M.; GOULARTE, R. Gesture4all: a framework for 3d gestural interaction to improve accessibility of web videos. In: ACM (Ed.). **SAC 2018: Symposium on Applied Computing , April 9–13, 2018, Pau, France**. [S.l.: s.n.], 2018. DOI: 10.1145/3167132.3167363. Citado na página 106.

GERONIMO, L.; ARAS, E.; NORRIE, M. C. Tilt-and-tap: Framework to support motion-based web interaction techniques. In: **Proceedings of the 15th International Conference on Engineering the Web in the Big Data Era - Volume 9114**. New York, NY, USA: Springer-Verlag New York, Inc., 2015. (ICWE 2015), p. 565–582. ISBN 978-3-319-19889-7. Disponível em: <http://dx.doi.org/10.1007/978-3-319-19890-3_36>. Citado na página 69.

GHIROTTI, S. E.; MORIMOTO, C. H. Um sistema de interação baseado em gestos manuais tridimensionais para ambientes virtuais. In: **Proceedings of the IX Symposium on Human Factors in Computing Systems**. Porto Alegre, Brazil, Brazil: Brazilian Computer Society, 2010. (IHC '10), p. 159–168. Disponível em: <<http://dl.acm.org/citation.cfm?id=1999593.1999611>>. Citado nas páginas 43, 44 e 52.

GIZMAG. **Usuário controlando uma aplicação por meio de sua webcam**. 2015. Disponível em: <<http://www.gizmag.com/flutter-gesture-control/22057/>>. Citado na página 49.

GONZALEZ, M.; MORENO, L.; MARTINEZ, P.; IGLESIAS, A. Web accessibility requirements for media players. In: **Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part IV**. Berlin, Heidelberg: Springer-Verlag, 2011. (INTERACT'11), p. 669–674. ISBN 978-3-642-23767-6. Disponível em: <<http://dl.acm.org/citation.cfm?id=2042283.2042403>>. Citado nas páginas 35 e 55.

GRIJNCU, D.; NACENTA, M. A.; KRISTENSSON, P. O. User-defined interface gestures: Dataset and analysis. In: **Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces**. New York, NY, USA: ACM, 2014. (ITS '14), p. 25–34. ISBN 978-1-4503-2587-5. Disponível em: <<http://doi.acm.org/10.1145/2669485.2669511>>. Citado na página 51.

GRILLO, F. D. N.; FORTES, R. P. de M. Tests with blind programmers using awmo: An accessible web modeling tool. In: _____. **Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I**. Cham: Springer International Publishing, 2014. p. 104–113. ISBN 978-3-319-07437-5. Disponível em: <https://doi.org/10.1007/978-3-319-07437-5_11>. Citado na página 53.

GROUPON. **Why Video Should be a Part of Your 2015 Marketing Strategy**. 2014. Disponível em: <<https://www.groupon.co.uk/merchant/blog/why-video-should-be-a-part-of-your-2014-marketing-strategy>>. Citado na página 31.

HAO, T.; XING, G.; ZHOU, G. Runbuddy: A smartphone system for running rhythm monitoring. In: **Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. New York, NY, USA: ACM, 2015. (UbiComp '15), p. 133–144. ISBN 978-1-4503-3574-4. Disponível em: <<http://doi.acm.org/10.1145/2750858.2804293>>. Citado na página 43.

HARRISON, C.; TAN, D.; MORRIS, D. Skinput: Appropriating the body as an input surface. **CHI'10**, Proceedings of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, New York, NY, p. 453–462, 2010. Citado na página 42.

HIGHQ. **2017: The year of video marketing**. 2017. Disponível em: <<https://highq.com/2017-the-year-of-video-marketing/>>. Citado na página 31.

HOGGAN, E.; NACENTA, M.; KRISTENSSON, P. O.; WILLIAMSON, J.; OULASVIRTA, A.; LEHTI, A. Multi-touch pinch gestures: Performance and ergonomics. In: **Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces**. New York, NY, USA: ACM, 2013. (ITS '13), p. 219–222. ISBN 978-1-4503-2271-3. Disponível em: <<http://doi.acm.org/10.1145/2512349.2512817>>. Citado nas páginas 50 e 60.

HONG, R.; WANG, M.; XU, M.; YAN, S.; CHUA, T.-S. Dynamic captioning: Video accessibility enhancement for hearing impairment. In: **Proceedings of the 18th ACM International Conference on Multimedia**. New York, NY, USA: ACM, 2010. (MM '10), p. 421–430. ISBN 978-1-60558-933-6. Disponível em: <<http://doi.acm.org/10.1145/1873951.1874013>>. Citado na página 55.

HONG, R.; WANG, M.; YUAN, X.-T.; XU, M.; JIANG, J.; YAN, S.; CHUA, T.-S. Video accessibility enhancement for hearing-impaired users. **ACM Trans. Multimedia Comput. Commun. Appl.**, ACM, New York, NY, USA, v. 7S, n. 1, p. 24:1–24:19, nov. 2011. ISSN 1551-6857. Disponível em: <<http://doi.acm.org/10.1145/2037676.2037681>>. Citado na página 55.

HUANG, C.-C. J.; YANG, R.; NEWMAN, M. W. The potential and challenges of inferring thermal comfort at home using commodity sensors. In: **Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. New York, NY, USA: ACM, 2015. (UbiComp '15), p. 1089–1100. ISBN 978-1-4503-3574-4. Disponível em: <<http://doi.acm.org/10.1145/2750858.2805831>>. Citado na página 39.

IBGE. **Censo demográfico**. 2012. Acessado em 10-01-2016. Disponível em: <ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religiao_Deficiencia/caracteristicas_religiao_deficiencia.pdf>. Citado na página 32.

ISHII, H.; ULLMER, B. Tangible bits: Towards seamless interfaces between people, bits and atoms. In: **Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 1997. (CHI '97), p. 234–241. ISBN 0-89791-802-9. Disponível em: <<http://doi.acm.org/10.1145/258549.258715>>. Citado na página 40.

ISO/9241-210:2010. **Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems**. 2010. Citado na página 103.

ISO/IEC25066:2016. **Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for Usability – Evaluation Report**. 2016. Citado na página 103.

JOHNSON, E. Touch display—a novel input/output device for computers. **Electronics Letters**, Institution of Engineering and Technology, v. 1, p. 219–220(1), October 1965. ISSN 0013-5194. Disponível em: <http://digital-library.theiet.org/content/journals/10.1049/el_19650200>. Citado na página 41.

KAMAL, A.; LI, Y.; LANK, E. Teaching motion gestures via recognizer feedback. In: **Proceedings of the 19th International Conference on Intelligent User Interfaces**. New York, NY, USA: ACM, 2014. (IUI '14), p. 73–82. ISBN 978-1-4503-2184-6. Disponível em: <<http://doi.acm.org/10.1145/2557500.2557521>>. Citado na página 67.

KANE, S. K.; WOBBROCK, J. O.; LADNER, R. E. Usable gestures for blind people: Understanding preference and performance. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2011. (CHI '11), p. 413–422. ISBN 978-1-4503-0228-9. Disponível em: <<http://doi.acm.org/10.1145/1978942.1979001>>. Citado nas páginas 35 e 51.

KIM, H. N. Haptic user interface design for students with visual impairments. In: **Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: ACM, 2009. (Assets '09), p. 267–268. ISBN 978-1-60558-558-1. Disponível em: <<http://doi.acm.org/10.1145/1639642.1639709>>. Citado na página 60.

KINECT. **Usuário realizando um gesto tridimensional com o Kinect**. 2015. Disponível em: <<http://www.xbox.com/pt-br/>>. Citado na página 48.

KORPELA, J.; MIYAJI, R.; MAEKAWA, T.; NOZAKI, K.; TAMAGAWA, H. Evaluating tooth brushing performance with smartphone sound data. In: **Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. New York, NY, USA: ACM, 2015. (UbiComp '15), p. 109–120. ISBN 978-1-4503-3574-4. Disponível em: <<http://doi.acm.org/10.1145/2750858.2804259>>. Citado na página 39.

KUNDU, A.; HE, Y.; BAHL, P. Recognition of handwritten word: first and second order hidden markov model based approach. In: **Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on**. [S.l.: s.n.], 1988. p. 457–462. ISSN 1063-6919. Citado na página 53.

LEAPMOTION. **Usuário realizando um gesto tridimensional com o Leap Motion**. 2015. Disponível em: <<https://www.leapmotion.com/>>. Citado na página 47.

LEE, J. C. In search of a natural gesture. **XRDS**, ACM, New York, NY, USA, v. 16, n. 4, p. 9–12, jun. 2010. ISSN 1528-4972. Disponível em: <<http://doi.acm.org/10.1145/1764848.1764853>>. Citado nas páginas 35, 40, 41, 44, 49, 50 e 52.

LEE, K.-F.; HON, H.-W.; HWANG, M.-Y.; HUANG, X. Speech recognition using hidden markov models: A cmu perspective. **Speech Commun.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 9, n. 5-6, p. 497–508, dez. 1990. ISSN 0167-6393. Disponível em: <[http://dx.doi.org/10.1016/0167-6393\(90\)90025-5](http://dx.doi.org/10.1016/0167-6393(90)90025-5)>. Citado na página 53.

LIMNA, T.; TANDAYYA, P.; SUVANVORN, N. Low-cost stereo vision system for supporting the visually impaired's walk. In: **Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology**. New York, NY, USA: ACM, 2009. (i-CREATE '09), p. 4:1–4:4. ISBN 978-1-60558-792-9. Disponível em: <<http://doi.acm.org/10.1145/1592700.1592705>>. Citado na página 59.

LIU, K.; NAKATA, K.; LI, W.; GALARRETA, D. **Information and Knowledge Management in Complex Systems: 16th IFIP WG 8.1 International Conference on Informatics and Semiotics in Organisations**. [S.l.]: Springer Publishing Company, Incorporated, 2015. ISBN 331916273X, 9783319162737. Citado na página 35.

LOVATO, S.; PIPER, A. M. "siri, is this you?": Understanding young children's interactions with voice input systems. In: **Proceedings of the 14th International Conference on Interaction Design and Children**. New York, NY, USA: ACM, 2015. (IDC '15), p. 335–338. ISBN 978-1-4503-3590-4. Disponível em: <<http://doi.acm.org/10.1145/2771839.2771910>>. Citado na página 42.

MACKENZIE, I. **Human-Computer Interaction: An Empirical Research Perspective**. Elsevier Science, 2012. ISBN 9780124071650. Disponível em: <<https://books.google.com.br/books?id=k0kBgYCaokAC>>. Citado nas páginas 63, 76 e 79.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: **In 5-th Berkeley Symposium on Mathematical Statistics and Probability**. [S.l.: s.n.], 1967. p. 281–297. Citado na página 52.

MANDERS, C.; FARBIZ, F.; YIN, T. K.; MIAOLONG, Y.; CHONG, B.; GUAN, C. G. Interacting with 3d objects in a virtual environment using an intuitive gesture system. In: **Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and**

Its Applications in Industry. New York, NY, USA: ACM, 2008. (VRCAI '08), p. 5:1–5:5. ISBN 978-1-60558-335-8. Disponível em: <<http://doi.acm.org/10.1145/1477862.1477869>>. Citado nas páginas 51 e 60.

MANN, S. Natural interfaces for musical expression: Physiphones and a physics-based organology. In: **Proceedings of the 7th International Conference on New Interfaces for Musical Expression**. New York, NY, USA: ACM, 2007. (NIME '07), p. 118–123. Disponível em: <<http://doi.acm.org/10.1145/1279740.1279761>>. Citado nas páginas 40 e 41.

MANN, S.; FUNG, J.; MONCRIEFF, E. Eyetap technology for wireless electronic news gathering. **SIGMOBILE Mob. Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 3, n. 4, p. 19–26, out. 1999. ISSN 1559-1662. Disponível em: <<http://doi.acm.org/10.1145/584039.584044>>. Citado nas páginas 40 e 41.

MAPARI, R. B.; KHARAT, G. American static signs recognition using leap motion sensor. In: **Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies**. New York, NY, USA: ACM, 2016. (ICTCS '16), p. 67:1–67:5. ISBN 978-1-4503-3962-9. Disponível em: <<http://doi.acm.org/10.1145/2905055.2905125>>. Citado na página 68.

MATTOS, F. A. M. d.; CHAGAS, G. J. d. N. Desafios para a inclusão digital no brasil. **Perspectivas em Tecnologia da Informação**, scielo, v. 13, p. 67 – 94, 04 2008. ISSN 1413-9936. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362008000100006&nrm=iso>. Citado nas páginas 34 e 54.

MCCULLOUGH, M.; XU, H.; MICHELSON, J.; JACKOSKI, M.; PEASE, W.; COBB, W.; KALESCKY, W.; LADD, J.; WILLIAMS, B. Myo arm: Swinging to explore a ve. In: **Proceedings of the ACM SIGGRAPH Symposium on Applied Perception**. New York, NY, USA: ACM, 2015. (SAP '15), p. 107–113. ISBN 978-1-4503-3812-7. Disponível em: <<http://doi.acm.org/10.1145/2804408.2804416>>. Citado na página 46.

MITRA, S.; ACHARYA, T. Gesture recognition: A survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 37, n. 3, p. 311–324, May 2007. ISSN 1094-6977. Citado nas páginas 50 e 60.

MORELLI, T.; FOLMER, E. Real-time sensory substitution to enable players who are blind to play video games using whole body gestures. In: **Proceedings of the 6th International Conference on Foundations of Digital Games**. New York, NY, USA: ACM, 2011. (FDG '11), p. 147–153. ISBN 978-1-4503-0804-5. Disponível em: <<http://doi.acm.org/10.1145/2159365.2159385>>. Citado na página 59.

MORIMOTO, C. H.; AMIR, A.; FLICKNER, M. Free head motion eye gaze tracking without calibration. In: **CHI '02 Extended Abstracts on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2002. (CHI EA '02), p. 586–587. ISBN 1-58113-454-1. Disponível em: <<http://doi.acm.org/10.1145/506443.506496>>. Citado na página 42.

MORO, L.; PEREIRA, R.; MORTIMER, E.; ARAUJO, A.; SA, E.; SANTOS, E.; SILVA, P.; MARTINS, R.; QUADROS, A.; AZEVEDO, L. Gestos ancorados em um terceiro modo semiótico: como auxiliam na criação de modelos tridimensionais. In: **Atas do IX Encontro Nacional de Pesquisa em Educação em Ciências**. [S.l.]: ENPEC, 2013. (IHC '10). Citado na página 43.

MOSER, C.; TSCHELIGI, M. Physics-based gaming: Exploring touch vs. mid-air gesture input. In: **Proceedings of the 14th International Conference on Interaction Design and Children**. New York, NY, USA: ACM, 2015. (IDC '15), p. 291–294. ISBN 978-1-4503-3590-4. Disponível em: <http://doi.acm.org/10.1145/2771839.2771899>. Citado nas páginas 47 e 60.

MULLING, T.; SATHIYANARAYANAN, M. Characteristics of hand gesture navigation: A case study using a wearable device (myo). In: **Proceedings of the 2015 British HCI Conference**. New York, NY, USA: ACM, 2015. (British HCI '15), p. 283–284. ISBN 978-1-4503-3643-7. Disponível em: <http://doi.acm.org/10.1145/2783446.2783612>. Citado nas páginas 46, 60 e 68.

NESPOULOUS, P. P. J.-L.; LECOURS, A. **The Biological foundations of gestures: Motor and semiotic aspects**. [S.l.]: Erlbaum Associates, 1986. Citado na página 49.

NETWORK, G. O. **Usuário realizando um gesto tridimensional com o Wii Remote**. 2015. Disponível em: <http://gameofficenetwork.com/2015/07/24/o-uso-do-nintendo-wii-na-reabilitacao-fisioterapeutica-neurolgica/>. Citado na página 46.

NGUYEN, T. T. M.; PHAM, N. H.; DONG, V. T.; NGUYEN, V. S.; TRAN, T. T. H. A fully automatic hand gesture recognition system for human-robot interaction. In: **Proceedings of the Second Symposium on Information and Communication Technology**. New York, NY, USA: ACM, 2011. (SoICT '11), p. 112–119. ISBN 978-1-4503-0880-9. Disponível em: <http://doi.acm.org/10.1145/2069216.2069241>. Citado nas páginas 49 e 60.

NORMAN, D. A. Why interfaces don't work. In: **The art of human-computer interface design**. [S.l.: s.n.], 1990. p. 209–219. Citado nas páginas 40 e 60.

O'DONOVAN, K.; AYER, S. Real-time joint angle measurement using the shimmer wireless sensor platform. In: **Proceedings of the First ACM Workshop on Mobile Systems, Applications, and Services for Healthcare**. New York, NY, USA: ACM, 2011. (mHealthSys '11), p. 7:1–7:2. ISBN 978-1-4503-0684-3. Disponível em: <http://doi.acm.org/10.1145/2064942.2064953>. Citado na página 45.

OLIVEIRA, L. M.; MORITA, M. E. Introdução aos modelos escondidos de markov (hmm). **Simpósio do Programa de Pós-Graduação em Informática Aplicada da PUC-Pr**, Curitiba, Paraná, 2010. Citado na página 53.

PERRENOUD, C.; PHAN, K. Emergence of web technology: An implementation of web accessibility design in organizations. In: **Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET '12:**. [S.l.: s.n.], 2012. p. 633–645. Citado nas páginas 55, 56 e 61.

PETRIE, H.; SAVVA, A.; POWER, C. Towards a unified definition of web accessibility. In: **Proceedings of the 12th Web for All Conference**. New York, NY, USA: ACM, 2015. (W4A '15), p. 35:1–35:13. ISBN 978-1-4503-3342-9. Disponível em: <http://doi.acm.org/10.1145/2745555.2746653>. Citado na página 103.

POTTER, L. E.; ARAULLO, J.; CARTER, L. The leap motion controller: A view on sign language. In: **Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration**. New York, NY, USA: ACM, 2013. (OzCHI '13), p. 175–178. ISBN 978-1-4503-2525-7. Disponível em: <http://doi.acm.org/10.1145/2541016.2541072>. Citado nas páginas 47 e 60.

QAMAR, A. M.; AFYOUNI, I.; RAHMAN, M. A.; REHMAN, F. U.; HUSSAIN, D.; BASALAMAH, S.; LBATH, A. A gis-based serious game interface for therapy monitoring. In: **Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems**. New York, NY, USA: ACM, 2014. (SIGSPATIAL '14), p. 589–592. ISBN 978-1-4503-3131-9. Disponível em: <<http://doi.acm.org/10.1145/2666310.2666376>>. Citado na página 59.

RABINER, L. R.; WILPON, J. G.; SOONG, F. K. High performance connected digit recognition using hidden markov models. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 37, n. 8, p. 1214–1225, Aug 1989. ISSN 0096-3518. Citado na página 53.

REELNREEL. **10 video marketing predictions for 2015**. 2015. Disponível em: <<http://www.reelnreel.com/10-video-marketing-predictions-2015>>. Citado na página 31.

REVELLE, G.; ZUCKERMAN, O.; DRUIN, A.; BOLAS, M. Tangible user interfaces for children. In: **CHI '05 Extended Abstracts on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2005. (CHI EA '05), p. 2051–2052. ISBN 1-59593-002-7. Disponível em: <<http://doi.acm.org/10.1145/1056808.1057095>>. Citado na página 51.

RICO, J.; BREWSTER, S. Gestures all around us: User differences in social acceptability perceptions of gesture based interfaces. In: **Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services**. New York, NY, USA: ACM, 2009. (MobileHCI '09), p. 64:1–64:2. ISBN 978-1-60558-281-8. Disponível em: <<http://doi.acm.org/10.1145/1613858.1613936>>. Citado na página 55.

ROSAS-VILLENA, J. M.; RAMOS, B.; GOULARTE, R.; FORTES, R. P. M. Facilitas: an accessible video player. In: 19TH BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB (WEBMEDIA). **XII Workshop de Ferramentas e Aplicações (WFA)**. Salvador (Bahia) - Brazil, 2013. I, p. 49–52. Citado na página 55.

_____. Universal access in human-computer interaction. access to today's technologies: 9th international conference, uahci 2015, held as part of hci international 2015, los angeles, ca, usa, august 2-7, 2015, proceedings, part i. In: _____. Cham: Springer International Publishing, 2015. cap. Video Accessibility on the Most Accessed Websites - A Case Study Regarding Visual Disabilities, p. 231–241. ISBN 978-3-319-20678-3. Disponível em: <http://dx.doi.org/10.1007/978-3-319-20678-3_23>. Citado nas páginas 32, 33, 35 e 55.

SCHILIT, B.; ADAMS, N.; WANT, R. Context-aware computing applications. **IEEE Workshop on mobile computing systems and applications**, IEEE, Los Alamitos, v. 1, p. 85–90, 1992. Citado na página 40.

SCHLOMER, T.; POPPINGA, B.; HENZE, N.; BOLL, S. Gesture recognition with a wii controller. In: **Proceedings of the 2Nd International Conference on Tangible and Embedded Interaction**. New York, NY, USA: ACM, 2008. (TEI '08), p. 11–14. ISBN 978-1-60558-004-3. Disponível em: <<http://doi.acm.org/10.1145/1347390.1347395>>. Citado nas páginas 52 e 53.

SCHWALLER, M.; LALANNE, D.; KHALED, O. A. Pygmi: Creation and evaluation of a portable gestural interface. In: **Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries**. New York, NY, USA: ACM, 2010. (NordiCHI '10), p. 773–776. ISBN 978-1-60558-934-3. Disponível em: <<http://doi.acm.org/10.1145/1868914.1869026>>. Citado na página 67.

- SRIJEYANTHAN, K.; JAYALATH, C. Parallel processing on real-time gesture recognition system. In: **Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry**. New York, NY, USA: ACM, 2011. (VRCAI '11), p. 447–452. ISBN 978-1-4503-1060-4. Disponível em: <<http://doi.acm.org/10.1145/2087756.2087842>>. Citado na página 49.
- STURMAN, D. J.; ZELTZER, D. A design method for “whole-hand” human-computer interaction. **ACM Trans. Inf. Syst.**, ACM, New York, NY, USA, v. 11, n. 3, p. 219–238, jul. 1993. ISSN 1046-8188. Disponível em: <<http://doi.acm.org/10.1145/159161.159159>>. Citado nas páginas 51 e 60.
- SUBBU, K. P.; GOZICK, B.; DANTU, R. Locateme: Magnetic-fields-based indoor localization using smartphones. **ACM Trans. Intell. Syst. Technol.**, ACM, New York, NY, USA, v. 4, n. 4, p. 73:1–73:27, out. 2013. ISSN 2157-6904. Disponível em: <<http://doi.acm.org/10.1145/2508037.2508054>>. Citado na página 58.
- SUN, X.; LU, Z.; HU, W.; CAO, G. Symdetector: Detecting sound-related respiratory symptoms using smartphones. In: **Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing**. New York, NY, USA: ACM, 2015. (UbiComp '15), p. 97–108. ISBN 978-1-4503-3574-4. Disponível em: <<http://doi.acm.org/10.1145/2750858.2805826>>. Citado na página 38.
- TANAKA, E. H.; ROCHA, H. V. D. Evaluation of web accessibility tools. In: **Proceedings of the 10th Brazilian Symposium on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction**. Porto Alegre, Brazil, Brazil: Brazilian Computer Society, 2011. (IHC+CLIH '11), p. 272–279. ISBN 978-85-7669-257-7. Disponível em: <<http://dl.acm.org/citation.cfm?id=2254436.2254483>>. Citado na página 33.
- TAVARES, J. a.; BARBOSA, J.; COSTA, C.; YAMIN, A.; REAL, R. A smart wheelchair based on ubiquitous computing. In: **Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments**. New York, NY, USA: ACM, 2013. (PETRA '13), p. 1:1–1:4. ISBN 978-1-4503-1973-7. Disponível em: <<http://doi.acm.org/10.1145/2504335.2504336>>. Citado na página 58.
- TECMUNDO. **Interação do usuário com um monitor e smartphone**. 2015. Disponível em: <<http://www.tecmundo.com.br/touchscreen>>. Citado na página 51.
- THALMIC. **Usuário realizando um gesto tridimensional com o Myo Armband**. 2015. Disponível em: <<https://www.myo.com/>>. Citado na página 46.
- THATCHER, J.; WADDELL, C. **Constructing Accessible Websites**. [S.l.]: APress L. P., 2003. ISBN 1590591488. Citado na página 32.
- TRUONG, K. N.; ABOWD, G. D.; BROTHERTON, J. A. Who, what, when, where, how: Design issues of capture & access applications. In: **Proceedings of the 3rd International Conference on Ubiquitous Computing**. London, UK, UK: Springer-Verlag, 2001. (UbiComp '01), p. 209–224. ISBN 3-540-42614-0. Disponível em: <<http://dl.acm.org/citation.cfm?id=647987.741319>>. Citado na página 38.
- VERTANEN, K.; KRISTENSSON, P. O. On the benefits of confidence visualization in speech recognition. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. New York, NY, USA: ACM, 2008. (CHI '08), p. 1497–1500. ISBN 978-1-60558-011-1. Disponível em: <<http://doi.acm.org/10.1145/1357054.1357288>>. Citado na página 42.

- W3C. **W3c accessibility**. 2012. Acessado em 15-01-2016. Disponível em: <<http://www.w3.org/WAI/intro/accessibility.php>>. Citado nas páginas 34, 35 e 54.
- WACG. **W3c accessibility**. 2012. Acessado em 24-02-2016. Disponível em: <<https://www.w3.org/2008/06/video-notes#q1>>. Citado na página 55.
- WANG, B. Machine learning techniques for brain signal analysis with applications on seizure detection and brain-computer interfaces. **AI Matters**, ACM, New York, NY, USA, v. 2, n. 3, p. 18–19, jun. 2016. ISSN 2372-3483. Disponível em: <<http://doi.acm.org/10.1145/2911172.2911178>>. Citado na página 42.
- WANT, R.; HOPPER, A.; FALÇÃO, V.; GIBBONS, J. The active badge location system. **ACM Trans. Inf. Syst.**, ACM, New York, NY, USA, v. 10, n. 1, p. 91–102, jan. 1992. ISSN 1046-8188. Disponível em: <<http://doi.acm.org/10.1145/128756.128759>>. Citado na página 39.
- WAZLAWICK, R. **Metodologia de Pesquisa para Ciência da Computação**. [S.l.]: Elsevier Brasil, 2017. ISBN 9788535277838. Citado na página 37.
- WCAG. **Web Content Accessibility Guidelines (WCAG) 2.0**. 2008. Disponível em: <<https://www.w3.org/TR/WCAG20/>>. Citado nas páginas 32 e 74.
- WEISER, M. Some computer science issues in ubiquitous computing. **Commun. ACM**, ACM, New York, NY, USA, v. 36, n. 7, p. 75–84, jul. 1993. ISSN 0001-0782. Disponível em: <<http://doi.acm.org/10.1145/159544.159617>>. Citado nas páginas 37 e 60.
- WHO. **World Helth Organization Media Center - Visual impairment and blindness, Fact Sheet N°282**. 2014. Acessado em 10-01-2016. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs282/en/>>. Citado na página 32.
- WILLIAMS, R. C. Finger tracking and gesture interfacing using the nintendo® wiimote. In: **Proceedings of the 48th Annual Southeast Regional Conference**. New York, NY, USA: ACM, 2010. (ACM SE '10), p. 11:1–11:4. ISBN 978-1-4503-0064-3. Disponível em: <<http://doi.acm.org/10.1145/1900008.1900025>>. Citado na página 60.
- YAIRI, I. E.; NAOE, K.; IWASAWA, Y.; FUKUSHIMA, Y. Do multi-touch screens help visually impaired people to recognize graphics. In: **The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility**. New York, NY, USA: ACM, 2011. (ASSETS '11), p. 237–238. ISBN 978-1-4503-0920-2. Disponível em: <<http://doi.acm.org/10.1145/2049536.2049585>>. Citado na página 42.
- YANG, G.-Z.; DEMPÈRE-MARCO, L.; HU, X.-P.; ROWE, A. Visual search: psychophysical models and practical applications. **Image and Vision Computing**, v. 20, n. 4, p. 291 – 305, 2002. ISSN 0262-8856. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0262885602000227>>. Citado na página 42.
- YOUTUBE. **Estatísticas do Youtube**. 2015. Disponível em: <<https://www.youtube.com/yt/press/statistics.html>>. Citado na página 32.

GLOSSÁRIO

Framework: é uma abstração que une códigos comuns entre vários projetos de *software* provendo uma funcionalidade genérica. *Frameworks* são projetados com a intenção de facilitar o desenvolvimento de *software*, habilitando designers e programadores a gastarem mais tempo determinando as exigências do *software* do que com detalhes de baixo nível do sistema.

Padrões de projeto: ou *Design Pattern*, descreve uma solução geral reutilizável para um problema recorrente no desenvolvimento de sistemas de *software* orientados a objetos. Não é um código final, é uma descrição ou modelo de como resolver o problema do qual trata, que pode ser usada em muitas situações diferentes.

Template: é um documento sem conteúdo, com apenas a apresentação visual (apenas cabeçalhos por exemplo) e instruções sobre onde e qual tipo de conteúdo deve entrar a cada parcela da apresentação.

Web: Sinônimo mais conhecido de *World Wide Web* (WWW). É a interface gráfica da Internet que torna os serviços disponíveis totalmente transparentes para o usuário e ainda possibilita a manipulação multimídia da informação.

WYSIWYG: “What You See Is What You Get” ou “O que você vê é o que você obtém”. Recurso tem por objetivo permitir que um documento, enquanto manipulado na tela, tenha a mesma aparência de sua utilização, usualmente sendo considerada final. Isso facilita para o desenvolvedor que pode trabalhar visualizando a aparência do documento sem precisar salvar em vários momentos e abrir em um *software* separado de visualização.

ANEXO A - QUESTIONÁRIO GRUPO 1

Entrevista Estruturada

Questionário desenvolvido como roteiro para a entrevista estruturada utilizado para levantamento de perfil do estudo de caso com os usuários do Grupo 1.

1. Gênero: 1() Masculino 2() Feminino
2. Idade:
3. Descreva sua acuidade visual:
 - a) () Visão Normal (Snellen: 20/12 a 20/25 / Decimal: 1,5 a 0,8 / Bifocais comuns)
 - b) () Próxima do normal (Snellen: 20/30 a 20/60 / Decimal: 0,6 a 0,3 / Bifocais mais fortes lupas e baixo poder)
 - c) () Baixa visão moderada (Snellen: 20/80 a 20/15 / Decimal: 0,25 a 0,12 / Lentes esferoprismáticos, lupas mais fortes)
 - d) () Baixa visão severa (Snellen: 20/200 a 20/400 / Decimal: 0,10 a 0,05 / Lentes esféricas, lupas de mesa de alto poder)
 - e) () Baixa visão profunda (Snellen: 20/500 a 20/1000 / Decimal: 0,04 a 0,02 / Lupa telescópico, Magnificação, Bengala)
 - f) () Próximo à cegueira (Snellen: 20/1200 a 20/2500 / Decimal: 0,015 a 0,008 / Magnificação, Livros falados, Braile, saída de voz)
 - g) () Cegueira total (Snellen: SPL / Decimal: SPL / Aparelhos de SA)
4. Você utiliza recursos de acessibilidade em seu dia-a-dia? 1() Não 2 () Sim. Se sim, qual(ais)?
5. Qual o seu nível de escolaridade?

- a) ☐ Fundamental incompleto
 - b) ☐ Fundamental completo
 - c) ☐ Ensino médio incompleto
 - d) ☐ Ensino médio completo
 - e) ☐ Superior incompleto
 - f) ☐ Superior completo
 - g) ☐ Pós graduação incompleto
 - h) ☐ Pós graduação completo
 - i) ☐ Outros
6. Há quanto tempo você utiliza o computador? 1(☐) Menos de 1 ano 2(☐) Entre 1 e 5 anos 3(☐) Mais de 5 anos
7. Com qual frequência você navega na Web (Internet)? 1(☐) Diariamente 2(☐) Algumas vezes na semana 3(☐) Algumas vezes no mês
8. Qual(is) sites mais acessa em seu dia-a-dia? 1(☐) Redes Sociais 2(☐) Sites com Vídeos 3(☐) Notícias 4(☐) Outros:
9. Como costuma ser seu comportamento na Web? 1(☐) Acessa várias vezes ao dia 2(☐) Reserva um momento do dia para o uso 3(☐) Outro:
10. Você costuma acessar vídeos na Web? 1(☐) Sim 2(☐) Não. Se sim, com qual frequência? 1(☐) Diariamente 2(☐) Algumas vezes na semana 3(☐) Algumas vezes no mês
11. Qual o dispositivo que costuma acessar vídeos? 1(☐) Computador 2(☐) Celular 3(☐) Tablet 4(☐) Outros
12. Você utiliza algum programa de leitura de telas ou dispositivo para acessar vídeos? 1(☐) Não 2(☐) Sim, Qual aparelho: (☐) Computador (☐) Mobile
13. Em qual(is) sites costuma acessar vídeos na Internet? 1(☐) Youtube 2(☐) Vimeo 3(☐) Site de notícias (Uol, folha, globo, etc.) 4(☐) Redes sociais (Facebook, twitter, etc.) 5(☐) Outros
14. Quais conteúdos de vídeos você possui maior interesse? 1(☐) Notícias 2(☐) Entretenimento (música, filmes, novelas) 3(☐) Informações 4(☐) Educação 5(☐) Outros
15. Você sente dificuldades (barreiras) ou inseguranças ao acessar vídeos na Internet? 1(☐) Não 2(☐) Sim, qual(is)
16. Durante a reprodução do vídeo você costuma fazer interações com o reprodutor? 1(☐) Não, Porque: 2(☐) Sim, Porque:

17. Quais opções você mais utiliza durante a reprodução do vídeo?
- a) ☐ Pause
 - b) ☐ Aumentar Volume
 - c) ☐ Diminuir Volume
 - d) ☐ Avançar
 - e) ☐ Retroceder
 - f) ☐ Outro
18. Você gostaria de conhecer novas formas de interação com vídeo na Internet? 1(☐) Sim 2 (☐) Não

Entrevista Semi-Estruturada

Questionário desenvolvido como roteiro para a entrevista semi-estruturada utilizado para coleta das opiniões após a prática das tarefas do estudo de caso dos usuários do Grupo 1.

1. Como foi interagir com um novo sensor? 1(☐) Muito fácil 2(☐) Fácil 3(☐) Nem fácil, nem difícil 4(☐) Difícil 5(☐) Muito difícil. Comentários:
2. Qual nível de facilidade você teve para interagir e acessar as funções do reprodutor de vídeo por meio de gestos? 1(☐) Muito fácil 2(☐) Fácil 3(☐) Nem fácil, nem difícil 4(☐) Difícil 5(☐) Muito difícil. Comentários:
3. Qual nível de facilidade você teve ao acessar as funções do reprodutor de vídeo por meio de gestos no Myo? 1(☐) Muito fácil 2(☐) Fácil 3(☐) Nem fácil, nem difícil 4(☐) Difícil 5(☐) Muito difícil. Comentário:
4. Qual tarefa você teve maior dificuldade para realizar no modo usual? 1(☐) Todas 2(☐) Nenhuma 3(☐) Play 4(☐) Pause 5(☐) Aumentar Volume 6(☐) Diminuir Volume 7(☐) Avançar 8(☐) Retroceder. Comentários:
5. Qual tarefa você teve maior dificuldade para realizar com no Myo? 1(☐) Todas 2(☐) Nenhuma 3(☐) Play 4(☐) Pause 5(☐) Aumentar Volume 6(☐) Diminuir Volume 7(☐) Avançar 8(☐) Retroceder. Comentários:
6. Qual tarefa você teve maior facilidade para realizar com no Myo? 1(☐) Todas 2(☐) Nenhuma 3(☐) Play 4(☐) Pause 5(☐) Aumentar Volume 6(☐) Diminuir Volume 7(☐) Avançar 8(☐) Retroceder. Comentários:
7. Feedback: Qual feedback mais lhe ajudou para saber se o gesto foi reconhecido no Myo? 1(☐) O áudio do vídeo 2(☐) A voz 3(☐) Cores 4(☐) Vibração. Comentários:

8. Feedback: Em sua opinião como seria o feedback ideal no Myo? 1() Todos os 4 2() Apenas o áudio original do vídeo 3() Somente:
9. Em sua opinião alguma barreira que você descreveu anteriormente foi ultrapassada?
- a) Sim, qual(is):
 - b) Não, motivo(s):
 - c) Não sei, motivo:
 - d) Comentários:
10. Usaria os gestos para acesso de vídeo na Web no seu dia-a-dia?
- a) () Sim
 - b) () Não
 - c) () Não sei, motivo:
 - d) Comentários:
11. Qual método foi o melhor, do modo usual ou por uso de gestos?
- a) () Usual
 - b) () Gestual
 - c) () Não sei, motivo:
 - d) Comentários:
12. Você gostaria de personalizar o gesto para cada uma das tarefas?
- a) () Usual
 - b) () Não
 - c) () Não sei, motivo:
 - d) Comentários:

ANEXO B - QUESTIONÁRIO GRUPO 2

Entrevista Estruturada

Questionário desenvolvido como roteiro para a entrevista estruturada utilizado para levantamento de perfil do estudo de caso com os usuários do Grupo 2.

1. Gênero: 1() Masculino 2() Feminino
2. Idade:
3. Descreva sua acuidade visual:
 - a) () Visão Normal (Snellen: 20/12 a 20/25 / Decimal: 1,5 a 0,8 / Bifocais comuns)
 - b) () Próxima do normal (Snellen: 20/30 a 20/60 / Decimal: 0,6 a 0,3 / Bifocais mais fortes lupas e baixo poder)
 - c) () Baixa visão moderada (Snellen: 20/80 a 20/15 / Decimal: 0,25 a 0,12 / Lentes esferoprismáticos, lupas mais fortes)
 - d) () Baixa visão severa (Snellen: 20/200 a 20/400 / Decimal: 0,10 a 0,05 / Lentes esféricas, lupas de mesa de alto poder)
 - e) () Baixa visão profunda (Snellen: 20/500 a 20/1000 / Decimal: 0,04 a 0,02 / Lupa telescópico, Magnificação, Bengala)
 - f) () Próximo à cegueira (Snellen: 20/1200 a 20/2500 / Decimal: 0,015 a 0,008 / Magnificação, Livros falados, Braile, saída de voz)
 - g) () Cegueira total (Snellen: SPL / Decimal: SPL / Aparelhos de SA)
4. Você utiliza recursos de acessibilidade em seu dia-a-dia? 1() Não 2 () Sim. Se sim, qual(ais)?
5. Qual o seu nível de escolaridade?

- a) ☐ Fundamental incompleto
 - b) ☐ Fundamental completo
 - c) ☐ Ensino médio incompleto
 - d) ☐ Ensino médio completo
 - e) ☐ Superior incompleto
 - f) ☐ Superior completo
 - g) ☐ Pós graduação incompleto
 - h) ☐ Pós graduação completo
 - i) ☐ Outros
6. Há quanto tempo você utiliza o computador? 1(☐) Menos de 1 ano 2(☐) Entre 1 e 5 anos 3(☐) Mais de 5 anos
7. Com qual frequência você navega na Web (Internet)? 1(☐) Diariamente 2(☐) Algumas vezes na semana 3(☐) Algumas vezes no mês
8. Qual(is) sites mais acessa em seu dia-a-dia? 1(☐) Redes Sociais 2(☐) Sites com Vídeos 3(☐) Notícias 4(☐) Outros:
9. Como costuma ser seu comportamento na Web? 1(☐) Acessa várias vezes ao dia 2(☐) Reserva um momento do dia para o uso 3(☐) Outro:
10. Você costuma acessar vídeos na Web? 1(☐) Sim 2(☐) Não. Se sim, com qual frequência? 1(☐) Diariamente 2(☐) Algumas vezes na semana 3(☐) Algumas vezes no mês
11. Qual o dispositivo que costuma acessar vídeos? 1(☐) Computador 2(☐) Celular 3(☐) Tablet 4(☐) Outros
12. Você utiliza algum programa de leitura de telas ou dispositivo para acessar vídeos? 1(☐) Não 2(☐) Sim, Qual aparelho: (☐) Computador (☐) Mobile
13. Em qual(is) sites costuma acessar vídeos na Internet? 1(☐) Youtube 2(☐) Vimeo 3(☐) Site de notícias (Uol, folha, globo, etc.) 4(☐) Redes sociais (Facebook, twitter, etc.) 5(☐) Outros
14. Quais conteúdos de vídeos você possui maior interesse? 1(☐) Notícias 2(☐) Entretenimento (música, filmes, novelas) 3(☐) Informações 4(☐) Educação 5(☐) Outros
15. Você sente dificuldades (barreiras) ou inseguranças ao acessar vídeos na Internet? 1(☐) Não 2(☐) Sim, qual(is)
16. Durante a reprodução do vídeo você costuma fazer interações com o reprodutor? 1(☐) Não, Porque: 2(☐) Sim, Porque:

17. Quais opções você mais utiliza durante a reprodução do vídeo?
- a) ☐ Pause
 - b) ☐ Aumentar Volume
 - c) ☐ Diminuir Volume
 - d) ☐ Avançar
 - e) ☐ Retroceder
 - f) ☐ Outro
18. Você gostaria de conhecer novas formas de interação com vídeo na Internet? 1(☐) Sim 2(☐) Não

Entrevista Semi-Estruturada

Questionário desenvolvido como roteiro para a entrevista semi-estruturada utilizado para coleta das opiniões após a prática das tarefas do estudo de caso dos usuários do Grupo 2.

1. Perguntas sobre a fase de aprendizado

- a) Dos sensores apresentados qual já teve contato anteriormente? 1(☐) Smartphone 2(☐) Myo 3(☐) Leap Motion 4(☐) Câmera. Comentários:
- b) Dos sensores apresentados qual você mais gostou? 1(☐) Smartphone 2(☐) Myo 3(☐) Leap Motion 4(☐) Câmera. Comentários:
- c) Como foi interagir com um novo sensor? 1(☐) Muito fácil 2(☐) Fácil 3(☐) Nem fácil, nem difícil 4(☐) Difícil 5(☐) Muito difícil. Comentários:
- d) Qual nível de facilidade você teve para interagir e acessar as funções do reprodutor de vídeo por meio de gestos? 1(☐) Muito fácil 2(☐) Fácil 3(☐) Nem fácil, nem difícil 4(☐) Difícil 5(☐) Muito difícil. Comentários:

2. Perguntas sobre os sensores

- a) Smartphone:
 - i. Qual nível de facilidade você teve ao acessar as funções do reprodutor de vídeo por meio de gestos no Smartphone? 1(☐) Muito fácil 2(☐) Fácil 3(☐) Nem fácil, nem difícil 4(☐) Difícil 5(☐) Muito difícil. Comentários:
 - ii. Qual tarefa você teve maior dificuldade para realizar com o Smartphone? 1(☐) Todas 2(☐) Nenhuma 3(☐) Play 4(☐) Pause 5(☐) Aumentar Volume 6(☐) Diminuir Volume 7(☐) Avançar 8(☐) Retroceder. Comentários:
 - iii. Qual tarefa você teve maior facilidade para realizar com o Smartphone? 1(☐) Todas 2(☐) Nenhuma 3(☐) Play 4(☐) Pause 5(☐) Aumentar Volume 6(☐) Diminuir Volume 7(☐) Avançar 8(☐) Retroceder. Comentários:

- iv. Feedback: Qual feedback mais lhe ajudou para saber se o gesto foi reconhecido no Smartphone? 1() O áudio do vídeo 2() A voz 3() Cores 4() Vibração. Comentários:
- v. Feedback: Em sua opinião como seria o feedback ideal no Smartphone? 1() Todos os 4 2() Apenas o áudio original do vídeo 3() Somente:
- b) Webcam:
- i. Qual nível de facilidade você teve ao acessar as funções do reprodutor de vídeo por meio de gestos na Webcam? 1() Muito fácil 2() Fácil 3() Nem fácil, nem difícil 4() Difícil 5() Muito difícil. Comentários:
- ii. Qual tarefa você teve maior dificuldade para realizar com a Webcam? 1() Todas 2() Nenhuma 3() Play 4() Pause 5() Aumentar Volume 6() Diminuir Volume 7() Avançar 8() Retroceder. Comentários:
- iii. Qual tarefa você teve maior facilidade para realizar com a Webcam? 1() Todas 2() Nenhuma 3() Play 4() Pause 5() Aumentar Volume 6() Diminuir Volume 7() Avançar 8() Retroceder. Comentários:
- iv. Feedback: Qual feedback mais lhe ajudou para saber se o gesto foi reconhecido na Webcam? 1() O áudio do vídeo 2() A voz 3() Cores. Comentários:
- v. Feedback: Em sua opinião como seria o feedback ideal na Webcam? 1() Todos os 3 2() Apenas o áudio original do vídeo 3() Somente:
- c) Myo:
- i. Qual nível de facilidade você teve ao acessar as funções do reprodutor de vídeo por meio de gestos no Myo? 1() Muito fácil 2() Fácil 3() Nem fácil, nem difícil 4() Difícil 5() Muito difícil. Comentário:
- ii. Qual tarefa você teve maior dificuldade para realizar com no Myo? 1() Todas 2() Nenhuma 3() Play 4() Pause 5() Aumentar Volume 6() Diminuir Volume 7() Avançar 8() Retroceder. Comentários:
- iii. Qual tarefa você teve maior facilidade para realizar com no Myo? 1() Todas 2() Nenhuma 3() Play 4() Pause 5() Aumentar Volume 6() Diminuir Volume 7() Avançar 8() Retroceder. Comentários:
- iv. Feedback: Qual feedback mais lhe ajudou para saber se o gesto foi reconhecido no Myo? 1() O áudio do vídeo 2() A voz 3() Cores 4() Vibração. Comentários:
- v. Feedback: Em sua opinião como seria o feedback ideal no Myo? 1() Todos os 4 2() Apenas o áudio original do vídeo 3() Somente:
- d) Leap Motion:
- i. Qual nível de facilidade você teve ao acessar as funções do reprodutor de vídeo por meio de gestos no Leap Motion? 1() Muito fácil 2() Fácil 3() Nem fácil, nem difícil 4() Difícil 5() Muito difícil. Comentários:

- ii. Qual tarefa você teve maior dificuldade para realizar com no Leap Motion? 1() Todas 2() Nenhuma 3() Play 4() Pause 5() Aumentar Volume 6() Diminuir Volume 7() Avançar 8() Retroceder. Comentários:
- iii. Qual tarefa você teve maior facilidade para realizar com no Leap Motion? 1() Todas 2() Nenhuma 3() Play 4() Pause 5() Aumentar Volume 6() Diminuir Volume 7() Avançar 8() Retroceder. Comentários:
- iv. Feedback: Qual feedback mais lhe ajudou para saber se o gesto foi reconhecido no Leap Motion? 1() O áudio do vídeo 2() A voz 3() Cores 4() Vibração. Comentários:
- v. Feedback: Em sua opinião como seria o feedback ideal no Leap Motion? 1() Todos os 4 2() Apenas o áudio original do vídeo 3() Somente:

3. Perguntas finais

- a) Em sua opinião alguma barreira que você descreveu anteriormente foi ultrapassada?
 - i. Sim, qual(is):
 - ii. Não, motivo(s):
 - iii. Não sei, motivo:
 - iv. Comentários:
- b) Usaria os gestos para acesso de vídeo na Web no seu dia-a-dia?
 - i. () Sim
 - ii. () Não
 - iii. () Não sei, motivo:
 - iv. Comentários:
- c) Qual método foi o melhor, do modo usual ou por uso de gestos?
 - i. () Usual
 - ii. () Gestual
 - iii. () Não sei, motivo:
 - iv. Comentários:
- d) Você gostaria de personalizar o gesto para cada uma das tarefas?
 - i. () Usual
 - ii. () Não
 - iii. () Não sei, motivo:
 - iv. Comentários: