

Um Modelo de Contexto para Apoiar a Acessibilidade em Ferramentas de Video Chamadas

Carlos Daniel Santana Cruz¹, Murilo Guerreiro Arouca¹, Ailton Ribeiro¹,
Maria Clara Pestana, Vaninha Vieira¹

¹Instituto de Computação, Programa de Pós-Graduação em Ciência da Computação
Universidade Federal da Bahia (UFBA)

{cdansantana, murilo.guerreiro, ailton.ribeiro, mpestana, vaninha}@ufba.br

Abstract. *Due to the advent of the Covid-19 pandemic, the use of video calls has become indispensable for the continuation of daily activities. In this scenario, there is a gap related to accessibility for the inclusion of people with disabilities. The objective of this work is to propose a tool to complement video calling platforms, by providing instructions and automating some tasks inherent to the participation of people with visual impairments. For this purpose, a context model was developed, structured from a study with users, where the main problems related to accessibility in video call platforms were highlighted. Through this approach it is possible to provide accessible information related to the aspects of their video, such as image framing and brightness level.*

Resumo. *Com o advento da pandemia de Covid-19, o uso de videochamadas tornou-se indispensável para a continuidade das atividades diárias. Nesse cenário, existe uma lacuna relacionada à acessibilidade para a inclusão de pessoas com deficiência. O objetivo deste trabalho é propor uma ferramenta para complementar as plataformas de videochamada, fornecendo instruções e automatizando algumas tarefas inerentes à participação de pessoas com deficiência visual. Para tanto, foi desenvolvido um modelo de contexto, estruturado a partir de um estudo com usuários, onde foram destacados os principais problemas relacionados à acessibilidade em plataformas de videochamada. Por meio dessa abordagem é possível disponibilizar informações acessíveis relacionadas aos aspectos de seu vídeo, como enquadramento da imagem e nível de brilho.*

1. Introdução

As vídeo conferências apresentaram um crescimento significativo no que tange a sua popularização e adoção nas últimas décadas. Contudo, devido ao advento do contexto pandêmico, decorrente da COVID-19, as vídeo conferências passaram a apresentar-se como recurso essencial para a continuidade de atividades do cotidiano em diferentes vertentes de aplicação, tais como atividades educacionais, telemedicina e atividades corporativas. Desse modo, evidencia-se a importância da utilização de vídeo chamadas para a interlocução de pessoas que não estão presentes fisicamente em um mesmo local, favorecendo a manutenção de conexões interpessoais.

Em 2018, uma pesquisa realizada pela Microsoft envolvendo cerca de 14 mil pessoas mostrou que a forma de comunicação preferida pelos entrevistados era o encontro presencial, superando métodos como *chats*, *e-mails* e ligações [Spataro 2020]. Já em

abril de 2020, dados publicados pela mesma empresa apontaram 2,7 bilhões de minutos de reunião em um dia em sua ferramenta para reuniões (Microsoft Teams¹), o que, de acordo com a empresa, representou um aumento de 200% em relação à quantidade de tempo registrada na mesma ferramenta no mês de março do mesmo ano. Outra ferramenta para vídeo conferências que apresentou um grande crescimento em vários aspectos foi o Zoom². De acordo com dados publicados, é indicado um crescimento relacionado à média diária de usuários ativos na ferramenta, passando de aproximadamente 10 milhões de pessoas em dezembro de 2019, para mais de 300 milhões em abril de 2020 [Iqbal 2020], [Novet 2020].

Neste cenário, os recursos de acessibilidade dispostos nas ferramentas de vídeo conferências, bem como nos dispositivos utilizados para realizar o acesso a essas ferramentas, desempenham papel fundamental na continuidade do processo de inclusão de pessoas com deficiência.

Embora a interação visual seja muito importante para muitas pessoas, e, devido à natureza das vídeoconferências, seja muito provável que a comunicação visual desempenhe o papel fundamental nas mesmas, deve-se lembrar que nem todos poderão ver o que está acontecendo em suas telas. Isso pode acontecer devido a vários fatores, tais como, algum dos participantes estar usando um telefone para participar da vídeoconferência e não poder prestar atenção à tela durante todo o tempo, dentre outros. Também pode acontecer de algum dos participantes ser cego ou apresentar perdas na visão, o que inviabiliza que a comunicação visual tenha efeito para esta pessoa.

Citando um exemplo, Shi et al. [Shi et al. 2019] propuseram sistemas adicionais a chamadas de vídeo, que por meio da utilização de inteligência artificial no processamento de imagens, conseguiu identificar e representar expressões corporais básicas, utilizando efeitos sonoros. O protótipo desenvolvido consistiu de duas ferramentas, atuando como um transmissor, para capturar e enviar a imagem de um dos participantes, e receptor, para analisar o vídeo recebido e produzir as informações sonoras para outro participante da vídeo conferência.

Este trabalho propõe uma ferramenta de acessibilidade para portadores de deficiência visual que visa complementar plataformas de vídeo chamadas existentes. Com base no processamento e reconhecimento de imagens, a ferramenta fornece instruções e automatiza algumas tarefas, com o objetivo de melhorar a experiência do usuário no que se refere à sua participação em vídeo chamadas. A adaptação da ferramenta se baseia no uso de um modelo de contexto, o qual foi construído a partir de um estudo com voluntários portadores de deficiência visual.

O restante do artigo está organizado da seguinte forma: na Seção 2 são abordados conceitos básicos e trabalhos relacionados. A Seção 3 apresenta a metodologia e os resultados de um estudo preliminar com usuários reais, que buscou evidenciar problemas comuns encontrados por pessoas no que diz respeito à acessibilidade em ferramentas de vídeo chamadas. Na Seção 4 é descrito o modelo de contexto utilizado para nortear o desenvolvimento das melhorias propostas. Na Seção 5, estão detalhados alguns experimentos iniciais realizados com o protótipo. Por fim, na Seção 6 são discutidas algumas

¹<https://www.microsoft.com/pt-br/microsoft-teams/free>

²<https://zoom.us/>

considerações finais deste trabalho e trabalhos futuros.

2. Fundamentação Teórica

Esta seção apresenta alguns conceitos relevantes para este trabalho, incluindo a modelagem de sistemas sensíveis ao contexto, e acessibilidade em aplicativos focalizados em vídeo.

2.1. Modelagem de Sistemas Sensíveis ao Contexto

Desenvolvedores de software observam uma progressiva necessidade dos usuários de que os sistemas atuais sejam intuitivos, interativos, e que não demandem grande esforço para a realização de suas respectivas tarefas, de modo que estes são preteridos em comparação aos sistemas que não são desenvolvidos em observação a essas características [Vieira et al. 2009].

Em virtude da necessidade de observação de informações que permeiam o contexto do usuário, bem como da crescente dependência do suporte computacional na vida das pessoas, é demandado que os sistemas computacionais utilizem dados provenientes de contextos para auxiliar os usuários na realização de atividades [Vieira et al. 2011]. Desse modo, a observação da pluralidade de contextos pelos sistemas computacionais para auxiliar os seus usuários caracteriza o conceito de sistemas sensíveis ao contexto.

De acordo com Sagaya e Kalpana, um sistema pode ser considerado sensível ao contexto, se este utiliza informações de contexto para prover a seus usuários informações ou serviços relevantes. Esta relevância depende diretamente da tarefa que o usuário precisa desempenhar junto ao sistema [Sagaya and Y. 2016].

A modelagem de contexto é um processo essencial para o desenvolvimento de sistemas sensíveis ao contexto, e é responsável por apoiar a coleta, avaliação e disseminação de informações de contexto. Ela se torna importante, pois, a partir dela que se estrutura toda a percepção que o sistema deverá possuir do ambiente onde ele estará inserido, e do que neste ambiente é relevante para seu funcionamento [Strang and Linnhoff-Popien 2004].

Sagaya e Kalpana elencam os principais tipos de métodos utilizados para a modelagem de contexto, que vão desde a representação através de pares chave-valor, até a utilização de UML e outros métodos mais avançados [Sagaya and Y. 2016]. Em uma discussão parecida, [Strang and Linnhoff-Popien 2004] apontam que, dentre as diferentes técnicas para a modelagem de contexto, as que utilizam ontologias como meio de modelagem se destacam.

Como descrito em [Vieira et al. 2011], um modelo de contexto deve conter duas partes: i) a estrutura do contexto, ou seja, identificar que informações constituem o contexto da aplicação (ex. entidades, e elementos contextuais que caracterizam essas entidades), e ii) as regras de adaptação ao contexto, ou seja regras e padrões que definem como deve ser o comportamento do sistema, com base nas informações de contexto.

2.2. Acessibilidade em aplicativos com foco em vídeo

No que diz respeito à acessibilidade, o Brasil dispõe de vários dispositivos jurídicos que buscam dar suporte a este tema. Dentre as principais leis, está a Lei Brasileira de Inclusão

da Pessoa com Deficiência³, Lei Nº 13.146, de 6 de julho de 2015, “destinada a assegurar e a promover, em condições de igualdade, o exercício dos direitos e das liberdades fundamentais por pessoa com deficiência, visando à sua inclusão social e cidadania”.

Ainda na referida lei, no artigo 3, é definido o conceito de acessibilidade como sendo 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 público, de uso público ou privados de uso coletivo, tanto na zona urbana como na rural, por pessoa com deficiência ou com mobilidade reduzida”.

Grandes companhias voltadas ao estabelecimento de padrões internacionais no desenvolvimento e na qualidade de aplicações Web (desde sites até aplicativos móveis) abordam em suas recomendações e documentações tópicos voltados à acessibilidade em produtos digitais.

A *Mozilla Developer Network (MDN)*⁴, projeto que tem como missão “prover uma completa, exata e útil documentação para tudo sobre a *open Web*, sendo ele construído, suportado ou não pela Mozilla [...]” possui uma lista de requisitos de acessibilidade para desenvolvedores de aplicativos móveis, *web* ou *desktop*, requisitos estes que abrangem aspectos do desenvolvimento como cor e contraste, visibilidade de elementos, usabilidade dentre outros, conforme [MDN 2020].

Já o *World Wide Web Consortium*⁵ (W3C), consórcio que tem como missão “levar a *World Wide Web* ao seu potencial máximo, desenvolvendo protocolos e diretrizes que garantam seu crescimento a longo prazo[...]”, recomenda em sua Cartilha de Acessibilidade na Web algumas características semelhantes às indicadas pela MDN, tais como clareza em hiperlinks, áreas de toque com tamanhos que possibilite a facilidade na sua ativação, dentre outros [Campbell et al. 2018].

Em [Wahlbin et al. 2015], é descrito como uma lista de padrões de acessibilidade pode ser aplicada no desenvolvimento de conteúdo Web para dispositivos móveis, aplicativos Web para dispositivos móveis, aplicativos nativos e aplicativos híbridos, através da utilização de componentes originalmente destinados à web.

Em [Richens et al. 2019] é apresentado o “*BBC Mobile Accessibility Prototype*⁶”, um conjunto de diretrizes baseados nos requisitos do conteúdo desenvolvido pela BBC para a audiência do Reino Unido e para uso com tecnologias disponíveis no Reino Unido. Este último, apesar de não ser destinado primariamente à comunidade internacional, constitui-se em guias que podem ser referenciadas por qualquer pessoa envolvida no projeto, desenvolvimento e teste de aplicações móveis e baseadas na web.

Além de diretrizes como as mencionadas anteriormente, que, contribuem para a acessibilidade em produtos digitais, a maioria dos sistemas disponíveis atualmente possuem um ou mais recursos de acessibilidade. Dentre esses recursos, alguns dos que se destinam a prover acessibilidade para pessoas com alguma deficiência visual são:

³<https://proinclusao.ufc.br/wp-content/uploads/2018/03/a-lei-brasileira-de-inclusao.pdf>

⁴<https://developer.mozilla.org/>

⁵<https://www.w3.org/>

⁶<https://www.bbc.co.uk/accessibility/forproducts/guides/mobile/>

- **Leitores de tela** - softwares que utilizam síntese de voz para converter as informações presentes na tela para áudio;
- **Lupa ou zoom** - ferramenta que permite ao usuário ampliar partes específicas da tela, a fim de visualizar melhor as informações contidas na região;
- **Contraste e fonte** - funcionalidade que permite inverter o contraste da tela ou editar o tamanho e cor da fonte;
- **Comandos de voz** - funcionalidade que permite a utilização do celular através da voz;
- **Áudio para texto** - função que permite a conversão de voz para texto.

No entanto, mesmo com o avanço dessas tecnologias, e com a crescente difusão de conceitos abordados em diretrizes como as citadas, ainda é necessário por parte dos desenvolvedores, a atenção para com os requisitos de acessibilidade no desenvolvimento de quaisquer sistemas. O quão acessível, ou "usável", um aplicativo ou sistema pode ser considerado, está diretamente relacionado à pessoa que o está utilizando, suas características pessoais, e suas necessidades para com o sistema em questão [Hack and Forum 2020, Vera 2020].

As empresas desenvolvedoras de ferramentas para vídeo chamadas tem investido no aprimoramento de recursos de acessibilidade presentes em seus produtos. São vários os recursos disponíveis, cada plataforma oferecendo suas variações. Dentre os recursos mais comuns, presentes na maioria das plataformas, segundo [Hack and Forum 2020], destacam-se:

1. **customização do tamanho de fontes** - permite que cada pessoa possa ajustar o tamanho ou as cores das fontes utilizadas em seções das ferramentas, tais como o chat. Esse recurso é um facilitador no acesso de quem possui baixa visão, e não necessariamente utiliza leitores de tela;
2. **legendas automáticas** - através da utilização em conjunto do reconhecimento de fala e de recursos de inteligência artificial, realiza a conversão para texto do que os participantes estão falando durante a vídeo chamada. Vale ressaltar que, uma vez que esse recurso possui algumas limitações, e sua qualidade depende de vários fatores, como a qualidade e inteligibilidade do áudio de quem está falando, os resultados não são 100% precisos, sendo a descrição manual (digitada por um participante da vídeo chamada) uma alternativa a este recurso;
3. **alertas sonoros** - permite a sonorização através do leitor de telas em uso pela pessoa, de alguns eventos que ocorrem na vídeo chamada, tais como novas mensagens, chegada e saída de participantes, dentre outros;
4. **atalhos de teclado** - permite a fácil execução de funções na ferramenta através de atalhos de teclado, tais como ativar ou desativar o microfone ou a câmera;
5. **transcrição de chamadas** - esse recurso utiliza as mesmas técnicas empregadas na criação de legendas automáticas, possuindo os mesmos aspectos negativos. No entanto, o resultado deste processo é geralmente utilizado para posteriores revisões das vídeo chamadas, geralmente em ambientes empresariais;
6. **organização personalizada dos participantes** - este recurso permite que cada pessoa possa ajustar sua visualização dos outros participantes de uma vídeo chamada; e
7. **fixação de participantes na tela principal** - permite aos responsáveis pela vídeo chamada a fixação de participantes que serão visíveis a todos, independentemente

de quem esteja falando. Este recurso geralmente é aplicado quando há uma pessoa responsável por fazer a tradução para a língua de sinais.

Sejam as vídeo chamadas para fins de trabalho ou para manter contato com outras pessoas, a utilização desses recursos tem o benefício de torná-las mais acessíveis e inclusivas para todos os participantes. A seguir estão algumas dicas de como contribuir para melhorar a experiência de todos os participantes de uma vídeo chamada, de acordo com [Hack and Forum 2020]:

- **diminua o ruído de fundo** - por exemplo, desligando aparelhos de som, silenciando televisores ou indo para uma sala silenciosa caso haja outras pessoas no ambiente onde você está;
- **ao usar legendas automáticas** - falar claramente mantendo um ritmo natural. Caso alguma pessoa presente na vídeo chamada dependa deste recurso, também é importante solicitar a todos os participantes que utilizem o recurso;
- **se possível, utilizar um microfone externo** - isto por que a utilização de um microfone externo tem o benefício de proporcionar uma melhor qualidade de áudio para todos os participantes;
- **posicionar-se próximo ao microfone** - isto ajuda tecnologias de legendas automáticas a captarem suas palavras mais claramente;
- **manter e solicitar a todos que mantenham o microfone silenciado** - enquanto não estiverem falando - isto ajuda a diminuir ruídos indesejáveis na vídeo chamada.
- **certificar-se da clareza no vídeo** -isto ajuda a quem utiliza a leitura labial como forma de entender o que os outros estão falando;
- **em caso de apresentações de slides** - se possível, disponibilizar acesso aos arquivos para todos os presentes, assim, eles poderão ler o conteúdo da apresentação independentemente da qualidade da vídeo chamada.

3. Estudo Exploratório sobre Acessibilidade em Vídeo Chamadas

Foram realizadas entrevistas *online* com 10 participantes voluntários, portadores de deficiência visual, com o objetivo de elucidar como aspectos da acessibilidade presentes nos aplicativos de vídeo chamadas estão sendo utilizados, percebidos e avaliados por este grupo de pessoas. Com este estudo, buscou-se levantar possíveis lacunas percebidas pelos participantes no que se refere aos recursos de acessibilidade, atualmente implementados pelas ferramentas utilizadas por eles para a realização de vídeo chamadas.

Para o recrutamento dos participantes, foram realizados convites em grupos em redes sociais. Os participantes possuíam idade entre 20 e 59 anos, sendo que 5 participantes identificaram-se como pertencentes ao gênero feminino, e 5 como pertencentes ao gênero masculino. Todos os participantes identificaram-se como sendo Pessoa com Deficiência (PcD). Dentre os tipos de deficiência, todos indicaram possuir deficiência visual.

A entrevista foi dividida em duas etapas. A primeira, com perguntas mais amplas, que abrangiam quaisquer dispositivos e ferramentas de vídeo chamadas que os participantes tivessem acesso. A segunda etapa foi focada exclusivamente em aspectos da ferramenta para vídeo chamadas Zoom. Sendo que nesta etapa, foram entrevistados apenas participantes que já a utilizaram.

Foram perguntadas a frequência que cada participante utilizava os recursos de câmera de algum dispositivo digital e a frequência que cada um participava de reuniões *online*. A fim de quantificar estas respostas, foi solicitado que os participantes indicassem as frequências em uma escala de Likert com indicações de 1 a 5. Os participantes indicaram os dispositivos que mais utilizavam para participar de vídeo chamadas, sendo os *smartphones*, os *laptops*, e os computadores de mesa, na ordem de preferência da maioria. Com relação a atividades do dia a dia envolvendo a câmera de dispositivos, a utilização da câmera para reconhecer textos e objetos foi indicada por 6 participantes. Tirar fotos foi indicada por 5 dos 9 participantes, e apenas 3 deles indicaram que utilizavam a câmera de dispositivos para gravar vídeos.

Durante a entrevista, os participantes indicaram, em uma escala de 0 a 5, o quão importante consideravam a qualidade do vídeo de outros participantes a serem exibidos para eles, e a qualidade de seu próprio vídeo a ser exibido para outros participantes de uma vídeo chamada. Com relação à qualidade do vídeo de outras pessoas a ser exibido, cinco participantes indicaram 0 na escala, dois participantes indicaram 1 na escala, e dois participantes indicaram 3 na escala. Em relação à qualidade do próprio vídeo a ser exibido a outras pessoas, observou-se uma diversidade nos níveis de importância indicados. Conforme pode-se verificar na Figura 1, a maioria dos participantes indicaram os dois maiores níveis da escala.

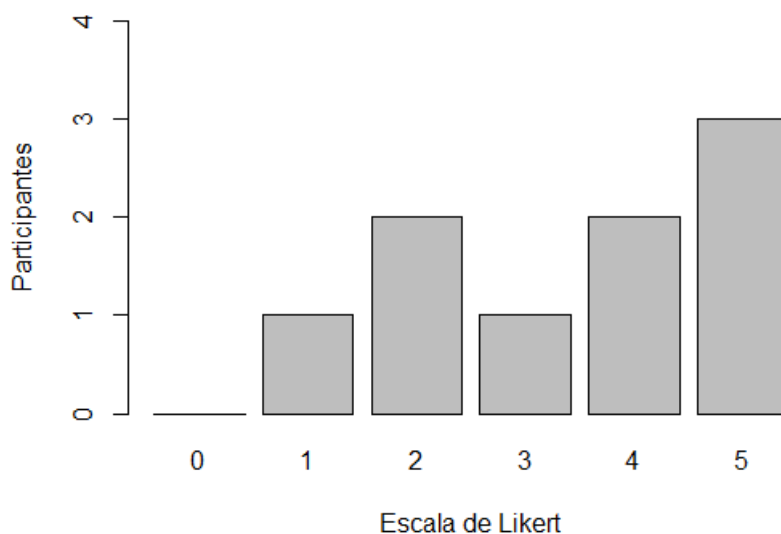


Figura 1. Nível de importância da qualidade do vídeo exibido pelos participantes a outras pessoas em uma videoconferência.

Foi perguntado se, ao utilizar a câmera de seus dispositivos, algum dos entrevistados fazia uso de alguma assistência. Para esta pergunta, P1 respondeu "Eu utilizo a identificação de rostos da câmera do celular", P3 indicou que utiliza "Leitor de tela no celular que ajuda a focalizar o objeto. Mas se for para vídeo chamada costumo pedir para alguém que enxerga dizer se a qualidade está boa".

No fim da primeira parte das entrevistas, foi perguntado se, caso a pessoa já tivesse participado de reuniões online, em algum momento ele(a) precisou ou sentiu falta de algum auxílio durante a reunião. Para esta pergunta, 4 participantes indicaram que já

precisaram ou sentiram falta de auxílio para acessar funcionalidades do sistema (tais como chat, recurso de "levantar a mão", dentre outros), 3 indicaram a necessidade de auxílio para focalizar a imagem da câmera e 2 para selecionar a câmera a ser usada durante a vídeo chamada. Apenas 1 participante indicou ter tido dificuldades para ajustar o microfone.

Para a segunda etapa da entrevista, foram realizadas perguntas relacionadas a aspectos de acessibilidade da ferramenta Zoom. Os participantes avaliaram, em uma escala de Likert de 0 a 5, os seguintes aspectos:

- Navegação (entre seções em uma tela, entre telas diferentes, dentre outros);
- Identificação de botões e suas respectivas ações;
- Retorno sonoro ou visual de ações realizadas (ligar ou desligar microfone ou câmera, enviar uma mensagem, dentre outros);
- Customização de aspectos da reunião (opções para convidados, opções de privacidade, dentre outros);
- Customização do seu perfil;
- Customização de recursos de acessibilidade (contraste, alertas, legendas, dentre outros).

Todos os respondentes indicaram um número maior ou igual a 3 na escala de Likert, exceto no último aspecto (Customização de recursos de acessibilidade), onde um respondente indicou o número 2 na escala. Assim, evidencia-se a relevância dos aspectos elucidados acima, bem como a lacuna existente em relação a acessibilidade nestes aspectos.

A próxima seção descreve uma proposta de ferramenta para apoiar a acessibilidade em ferramentas de vídeo chamadas.

4. Descrição da Ferramenta Proposta

Com base nos resultados obtidos no estudo com os usuários descrito anteriormente, foi elaborada uma proposta de ferramenta que pode ser implementada não somente na plataforma Zoom, como também em qualquer ferramenta de vídeo chamada, uma vez que a solução proposta se encaixa na maioria dos cenários que envolvem vídeo chamadas.

4.1. Modelo de Contexto

O modelo de contexto proposto está estruturado seguindo as recomendações abordadas em [Vieira et al. 2011]. Na Tabela 1, pode-se verificar os dados relevantes para a composição do modelo de contexto. Primeiramente, define-se possíveis contextos e o que o sistema deve realizar. Desse modo, são descritas as entidades relevantes para a proposta juntamente com os elementos contextuais a serem utilizados. Em nosso modelo constam quatro entidades: Pessoa, Dispositivo, Vídeo e Áudio. Em relação aos elementos contextuais, por exemplo, para a Entidade Pessoa, constam o elemento de contexto perfil da pessoa, e sua prioridade na vídeo chamada.

Na Tabela 2 estão dispostos os comportamentos da ferramenta proposta, juntamente com as regras de adaptação ao contexto, descritas na Tabela 3. Por exemplo, quando é identificado o contexto (3) *Foco da câmera muda*, a seguinte adaptação comportamental do sistema é ativada: *o sistema notifica a pessoa sobre a perda ou aquisição de foco na imagem*. Para essa adaptação, as seguintes regras são ativadas (Tabela 3): *Caso o*

Tabela 1. Modelo de contexto: identificação das entidades e elementos contextuais

| Entidade | Elementos Contextuais Relevantes |
|--------------------|---|
| Pessoa | Perfil; prioridade na vídeo chamada |
| Dispositivo em uso | Microfone; câmera; flash para a câmera; recursos de acessibilidade |
| Vídeo | Imagem de fundo; imagem de rosto; quantidade de pessoas identificadas |
| Áudio | Som ambiente; registro de voz |

vídeo sendo capturado perca o foco da pessoa, o dispositivo emite um alerta sonoro para a pessoa. Quando uma pessoa for focalizada novamente, o sistema emite outro alerta informando da focalização.

Tabela 2. Adaptações de Comportamento do Sistema

| Contexto | Adaptação comportamental do sistema |
|---|--|
| 1. Pessoa acessa a tela principal da ferramenta | O sistema identifica se algum recurso de acessibilidade no dispositivo está ativo. |
| 2. Pessoa inicia ou ingressa em uma vídeo chamada | O sistema analisa características do vídeo do usuário e o notifica. |
| 3. Foco da câmera muda | O sistema notifica a pessoa sobre a perda ou aquisição de foco na imagem. |
| 4. Luminosidade do vídeo muda | O sistema notifica a pessoa sobre a alteração na luminosidade do vídeo. |
| 5. Dispositivo possui Flash para a câmera em uso | O sistema ativa o flash da câmera em uso caso o ambiente esteja muito escuro. |
| 6. Pessoa permanece microfone ativo | O sistema detecta se o usuário está com o microfone ativo mas não está falando. |
| 7. Mais de um participante fala ao mesmo tempo | O sistema desativa o microfone de outras pessoas com base em prioridades. |
| 8. Sistema aprimora a qualidade do áudio | O sistema usa processamento de sinal digital para aprimorar o áudio da pessoa. |

4.2. Adaptações ao Contexto Desenvolvidas

Após a elaboração do modelo de contexto, realizamos alguns testes com recursos de acessibilidade para acessar ferramentas de vídeo chamadas, disponíveis em sistemas existentes. A seguir, descrevemos esses testes, seguindo os contextos e regras de comportamentos definidos na Tabela 3.

Para o Contexto 1 (*Pessoa acessa a tela principal da ferramenta*), foi planejada uma função a ser implementada no sistema. Essa função consulta junto ao dispositivo em uso a presença de algum recurso de acessibilidade. Caso encontre alguma ferramenta sendo utilizada, as funcionalidades previstas são acionadas. A consulta pode ser realizada de várias maneiras, dependendo do dispositivo, do sistema operacional e da linguagem de programação em questão.

Para a maioria dos sistemas operacionais presentes nos dispositivos atualmente, existem métodos para consultar se há alguma ferramenta de acessibilidade sendo execu-

tada no momento. No MacOs, é possível consultar arquivos de configuração atualizados pelo sistema, ou chaves disponíveis no registro do sistema. No Windows, a depender da linguagem de programação escolhida pelo desenvolvedor, existirão várias maneiras de verificar se algum recurso de acessibilidade está sendo utilizado. Uma maneira se dá através da consulta por executáveis de aplicações específicas junto aos programas atualmente em execução. Já nos sistemas Android e Ios, existem recursos específicos de cada plataforma para detectar o estado dos recursos de acessibilidade presentes no dispositivo.

Tabela 3. Regras de Adaptação de acordo com o comportamento do sistema

| Contexto | Regra |
|---|--|
| 1. Pessoa acessa a tela principal da ferramenta | O sistema consulta se algum recurso de acessibilidade está ativo no dispositivo em uso. Caso positivo, ele ativa as funcionalidades do assistente. |
| 2. Pessoa inicia ou ingressa em uma vídeo chamada | O sistema consulta junto ao dispositivo em uso a presença de câmeras, flashes para as mesmas, e microfones. Após isto, se tudo estiver em funcionamento, o sistema analisa o vídeo sendo capturado, e emite informações tais como luminosidade, foco da imagem, e quantidade de pessoas. Caso contrário, emite alerta informando a ausência dos itens não encontrados. |
| 3. Foco da câmera muda | Caso o vídeo sendo capturado perca o foco da pessoa, o dispositivo emite um alerta sonoro para a pessoa. Quando uma pessoa for focalizada novamente, o sistema emite outro alerta informando da focalização. |
| 4. Luminosidade do vídeo muda | Se, durante a captura de vídeo, a luminosidade mudar abruptamente, o sistema sinaliza para a pessoa, o alerta varia de acordo com a luminosidade percebida. |
| 5. Dispositivo possui Flash para a câmera em uso | Se, quando o sistema perceber a falta de luminosidade no vídeo, o dispositivo em uso possuir flash para a câmera sendo utilizada, ele deve sugerir ao usuário que ative o flash. |
| 6. Pessoa permanece com o microfone ativo | Se a pessoa permanecer por uma determinada quantidade de tempo com o microfone ativado, mas a voz da pessoa não for identificada, o sistema deve desativar o microfone da pessoa. |
| 7. Mais de um participante fala ao mesmo tempo | Se mais de um participante falar ao mesmo tempo durante a vídeo chamada, o sistema deve desativar os microfones dos participantes com base na prioridade estabelecida pelo responsável pela vídeo chamada. |
| 8. Sistema aprimora a qualidade do áudio | Com base no registro da voz da pessoa e no som captado pelo microfone, o sistema deve suprimir o som ambiente, ressaltando a voz da pessoa. Caso não seja possível, o sistema deve notificar à pessoa sobre a dificuldade em isolar a voz dela do som ambiente. |

Para o Contexto 2 (*Pessoa inicia ou ingressa em uma vídeo chamada*), a primeira parte da funcionalidade se caracteriza da mesma forma que o Contexto 1. Após a aquisição das informações necessárias para a continuidade, o sistema analisa o vídeo sendo capturado, e emite informações tais como luminosidade, foco da imagem, e quantidade de pessoas. O objetivo seria permitir à pessoa que vai ingressar em uma reunião ajustar aspectos de seu vídeo antes de ingressar na mesma.

As regras e contextos restantes destinam-se a ser acionadas enquanto a pessoa estiver em uma vídeo chamada. No que se refere à análise do áudio e do vídeo, são utilizadas técnicas de reconhecimento e processamento de imagem e técnicas de processamento de sinais digitais para obter os resultados desejados.

Com relação ao Contexto 3 (*Foco da câmera muda*), Contexto 4 (*Luminosidade*

do vídeo muda) e Contexto 5 (*Dispositivo possui Flash para a câmera em uso*), o retorno sonoro se dá através da verbalização (com o leitor de telas) das condições previstas em cada situação: número de rostos focalizados ou presentes no vídeo atualmente, ausência de rostos, intensidade da luminosidade percebida. Referente à luminosidade no vídeo, o retorno sonoro se dá através de tons musicais, sendo tons mais graves para quando a luminosidade estiver mais fraca e tons mais agudos para quando estiver mais brilhante.

Uma vez que a má qualidade de áudio impacta negativamente na clareza da comunicação, as regras e Contexto 6 (*Pessoa permanece com o microfone ativo*), Contexto 7 (*Mais de um participante fala ao mesmo tempo*), e Contexto 8 (*Sistema aprimora a qualidade do áudio*), destinam-se a contribuir com uma melhor qualidade no que se refere a este aspecto da vídeo chamada. Os 3 contextos complementam-se mutuamente, sendo o Contexto 8 responsável pela qualidade da mensagem sonora que a pessoa deseje transmitir, o Contexto 6 responsável por evitar que a pessoa transmita conteúdo sonoro indesejado, e o Contexto 7 responsável por tentar estabelecer uma organização na comunicação.

4.3. Estudo com Usuários

Como forma de verificar os comportamentos contextuais definidos no modelo de contexto, foi realizado e registrado um estudo de usuários com ferramentas externas à plataforma de vídeo chamadas.

O estudo se destinou a reproduzir os comportamentos esperados descritos na modelagem. Para tanto, foi realizada uma chamada de vídeo, e durante esta, utilizou-se tecnologias externas à plataforma, sendo a principal delas o reconhecimento facial.

Com base nisto, sempre que a ferramenta responsável por aplicar o reconhecimento facial identificava alguma das características esperadas (um rosto, mudanças de luminosidade) no vídeo sendo capturado, ela realizava a sinalização sonora através do leitor de telas presente no dispositivo, ou de sinais sonoros pré-gravados.

O estudo de usuários foi realizado com o auxílio de ferramentas existentes atualmente, externas à plataforma, uma vez que no momento da escrita deste trabalho, não foi possível implementar as funcionalidades planejadas diretamente na ferramenta Zoom. Os experimentos foram registrados em áudio e vídeo e estão disponíveis pelo link⁷.

Para desenvolver o protótipo, foi utilizado o framework para desenvolvimento mobile Ionic, em sua versão 4. Utilizou-se componentes próprios do framework que viabilizam o acesso e a manipulação de componentes dos dispositivos, tais como a câmera (para a captura de vídeo) e os alto falantes (para a reprodução de efeitos sonoros).

O framework permite a utilização da linguagem Java Script, e devido aos aplicativos construídos com a versão 4 do framework terem como base a utilização do Web View, foi possível utilizar bibliotecas como o WebAudio (para a reprodução dos efeitos sonoros), e a biblioteca face-api.js para a detecção de rostos.

Enquanto o protótipo estava em execução no dispositivo, era feita constantemente a verificação no conteúdo captado pela câmera do dispositivo, e assim que uma condição era encontrada (um rosto, por exemplo), o sistema reproduzia efeitos sonoros anterior-

⁷<https://encurtador.com.br/bdtLO>

mente gravados, para comunicar ao usuário o evento que fora detectado.

5. Considerações Finais e Trabalhos Futuros

Neste trabalho foi apresentado um modelo de contexto para apoiar a adição de recursos de acessibilidade a ferramentas de vídeo chamadas. As melhorias sugeridas visam contribuir para a melhora da experiência de pessoas que utilizem recursos de acessibilidade para utilizar estas ferramentas.

As sugestões aqui apresentadas basearam-se nas respostas obtidas através das perguntas descritas, juntamente com percepções feitas pelo autor durante estudos exploratórios com a ferramenta Zoom. No entanto, é necessário ressaltar que as sugestões não puderam ser implementadas diretamente na ferramenta em questão, uma vez que as opções disponíveis na ferramenta para implementações por terceiros não atendiam aos requisitos do modelo proposto.

Um outro aspecto importante é a quantidade de participantes envolvidos, o que indica que as soluções propostas não devem ser consideradas como válidas para todas as pessoas que utilizem recursos de acessibilidade no acesso ao Zoom. Por tanto, para solucionar os aspectos citados anteriormente, faz-se necessário um estudo e elaboração de estratégias que possam despertar o interesse do público alvo, de modo a refinar as propostas aqui descritas. Por fim, a implementação direta na plataforma, através das ferramentas de desenvolvimento disponibilizadas para terceiros também deve ser alvo de estudos, uma vez que os testes realizados para este trabalho se basearam em ferramentas externas à plataforma.

Referências

- Campbell, A., Kirkpatrick, A., Connor, J. O., and Cooper, M. (2018). Web content accessibility guidelines (WCAG) 2.1. W3C recommendation, W3C. <https://www.w3.org/TR/2018/REC-WCAG21-20180605/>.
- Hack, T. B. and Forum, B. D. (2020). Best video conferencing apps and software for accessibility.
- Iqbal, M. (2020). Zoom revenue and usage statistics (2020).
- MDN (2020). Accessibility.
- Novet, J. (2020). Zoom has added more videoconferencing users this year than in all of 2019 thanks to coronavirus, bernstein says.
- Richens, E. P., Williams, G. F., Knight, J., Mathews, M., and Nancarrow, R. (2019). Mobile accessibility guidelines.
- Sagaya, P. and Y., K. (2016). A review on context modelling techniques in context aware computing. 8:429–433.
- Shi, L., Tomlinson, B., Tang, J., Cutrell, E., McDuff, D., Venolia, G., Johns, P., and Rowan, K. (2019). Accessible video calling: Enabling nonvisual perception of visual conversation cues. In *Computer-Supported Cooperative Work 2019*. ACM, ACM.
- Spataro, J. (2020). Remote work trend report: meetings.

- Strang, T. and Linnhoff-Popien, C. (2004). A context modeling survey. In *First International Workshop on Advanced Context Modelling, Reasoning And Management at UbiComp 2004, Nottingham, England, September 7, 2004*. LIDO-Berichtsjahr=2004, Potential plagiarism at <http://dl.comsoc.org/cocoon/comsoc/servlets/GetPublication?id=9014634>.
- Vera, C. L. (2020). Which video conferencing tools are most accessible?
- Vieira, V., Tedesco, P., and Salgado, A. C. (2009). Modelos e processos para o desenvolvimento de sistemas sensíveis ao contexto.
- Vieira, V., Tedesco, P., and Salgado, A. C. (2011). Designing context-sensitive systems: An integrated approach. *Expert Systems with Applications*, 38(2):1119 – 1138. Intelligent Collaboration and Design.
- Wahlbin, K., Spellman, J. F., and Patch, K. (2015). Mobile accessibility: How WCAG 2.0 and other W3C/wai guidelines apply to mobile. W3C working draft, W3C. <https://www.w3.org/TR/2015/WD-mobile-accessibility-mapping-20150226/>.