

Universidade Federal da Paraíba
Centro de Informática
Programa de Pós-Graduação em Informática

**Acessibilidade em Cinemas Digitais: Uma Proposta
de Geração e Distribuição de Libras e Audiodescrição**

Leonardo de Araújo Domingues

João Pessoa, Paraíba, Brasil, 13 de Outubro de 2015

Universidade Federal da Paraíba
Centro de Informática
Programa de Pós-Graduação em Informática

**Acessibilidade em Cinemas Digitais: Uma Proposta
de Geração e Distribuição de Libras e Audiodescrição**

Leonardo de Araújo Domingues

Dissertação submetida à Coordenação do Curso de Pós-Graduação
em Informática da Universidade Federal da Paraíba como parte dos
requisitos necessários para obtenção do grau de Mestre em Informá-
tica.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Computação Distribuída

Dr. Tiago Maritan Ugulino de Araújo
(Orientador)

João Pessoa, Paraíba, Brasil

©Leonardo de Araújo Domingues, 13 de Outubro de 2015

D671a Domingues, Leonardo de Araújo.
Acessibilidade em cinemas digitais: uma proposta de
geração e distribuição de libras e autodescrição / Leonardo de
Araújo Domingues.- João Pessoa, 2015.
108f. : il.
Orientador: Tiago Maritan Ugolino de Araújo
Dissertação (Mestrado) - UFPB/CI
1. Informática. 2. Cinema digital - acessibilidade.
4. Deficiência auditiva - Libras. 5. Deficiência visual -
audiodescrição. 6. Aplicações multimídia.

UFPB/BC

CDU: 004(043)

Ata da Sessão Pública de Defesa de Dissertação de
Mestrado de **LEONARDO DE ARAUJO
DOMINGUES**, candidato ao título de Mestre em
Informática na Área de Sistemas de Computação,
realizada em 13 de outubro de 2015.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Aos treze dias do mês de outubro do ano de dois mil e quinze, às dezesseis horas, no Centro de Informática - Universidade Federal da Paraíba (unidade Mangabeira), reuniram-se os membros da Banca Examinadora constituída para julgar o Trabalho Final do Sr. **Leonardo de Araújo Domingues** vinculado a esta Universidade sob a matrícula 2013116132, candidato ao grau de Mestre em Informática, na área de "Sistemas de Computação", na linha de pesquisa "Computação Distribuída", do Programa de Pós-Graduação em Informática, da Universidade Federal da Paraíba. A comissão examinadora foi composta pelos professores doutores: Tiago Maritan Ugulino de Araújo (PPGI-UEPB), Orientador e Presidente da Banca, Guido Lemos de Souza Filho (PPGI-UEPB), Examinador Interno, Rostand Edson Oliveira Costa (UECG), Examinador Interno e Daniela Francisco Brauner (UFPel) Examinador Externo à Instituição. Dando início aos trabalhos, o professor Presidente da Banca cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que o mesmo fizesse, oralmente, a exposição do trabalho de dissertação intitulado "Acessibilidade em Cinemas Digitais: Uma Proposta de Geração e Distribuição de Libras e Audiodescrição". Concluída a exposição, o candidato foi arguido pela Banca Examinadora que emitiu o seguinte parecer: "aprovado". Assim sendo, eu, Nadja Rayssa Soares de Almeida, Auxiliar Administrativo, Secretária do Programa de Pós Graduação em Informática - PPGI, lavrei a presente ata que vai assinada por mim e pelos membros da Banca Examinadora. João Pessoa, 13 de outubro de 2015.

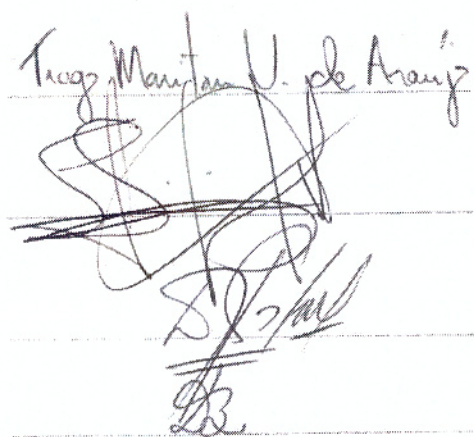

Nadja Rayssa Soares de Almeida

Profº Drº Tiago Maritan Ugulino de Araújo
Orientador (PPGI-UEPB)

Profº Drº Guido Lemos de Souza Filho
Examinador Interno (PPGI-UEPB)

Profº Drº Rostand Edson Oliveira Costa
Examinador Interno (UECG)

Profº Drº Daniela Francisco Brauner
Examinador Externo à Instituição (UFRN)



*À minha família, pelo apoio incondicional,
pelo incentivo à realização dos meus objetivos,
e pela amizade sem igual.
Sem eles a realização deste sonho não seria factível.*

Agradecimentos

O agradecimento é uma arte valiosa que muitas vezes está oculta sob a aparência do óbvio. Portanto, para perceber essa arte é preciso ir além das palavras e das ideias, é preciso compreender a essência da alma.

O maior exemplo, os ensinamentos mais valiosos, o amor incondicional percebido mesmo no mais singelo olhar, a coragem e a persistência, posso encontrar na pessoa mais ilustre que tive a honra de conhecer, minha mãe Elizabeth. Para sempre minha melhor amiga. Agradeço pelo ensinamento de que o caráter do homem se revela em sua integridade. Saiba que enquanto eu tiver a oportunidade de viver, sempre irei dedicar meus esforços para vê-la feliz. Te amo, mãe!

A doçura e a gentileza nunca foram tão evidentes quando olhando em seus olhos, marcados pela trajetória da vida, percebi que não há distinção de época ou idade para amar alguém ou alguma coisa. Essa é você, minha avó Terezinha, meu anjo, sempre serena e de coração puro. A plenitude da sua fé em Deus revela a pureza da sua alma, e isso figura entre as chaves secretas para a minha formação.

Durante esse processo de formação, em muitos momentos estive diante da fraqueza. A decisão de esmorecer ou reagir estava completamente sob meu domínio, porém, nunca houve êxito para a fraqueza, pois, reagir em qualquer circunstância implica na persistência do propósito da vida, e o propósito da minha vida é ser um exemplo para você, minha irmã Beatriz. Espero que enxergue em mim a honestidade, a perseverança e o desejo de ser feliz. Seu sorriso é combustível para o meu sucesso!

Habitado com a ausência da figura do pai educador, encontrei muitas vezes em você, meu irmão Rafael, o papel daquele que contribui com a experiência, com a construção do conhecimento e com a formação dos valores éticos e morais de um homem de bem. Sou eternamente grato por cada palavra de encorajamento, pelos seus conselhos e por acreditar que tenho potencial suficiente para definir uma meta e lutar até alcançá-la. Aprendi bastante coisa com você.

De modo muito especial, agradeço a você, Antônio Gonçalves. Seu apoio foi essencial para a conquista deste objetivo, pois, fico tranquilo ao saber que está cuidando

muito bem de duas das minhas jóias mais preciosas, que são minha mãe e minha avó. Agradeço também por não medir esforços em ajudar quando mais preciso. Você é um amigo sem igual, um amigo daqueles que devemos zelar para todo sempre.

Lembro quando era apenas um garoto que sonhava em ser um jogador de futebol. Lembro disso com bastante saudade, pois essa é a melhor fase da vida, quando a maior preocupação é o que queremos ser quando crescer. E por falar e crescer, agradeço a você, meu tio Carlos Henrique (*in memoriam*), por todas as vezes em que sentávamos na calçada e você falava sobre a importância de estudar, de conquistar uma formação de nível superior, de ser respeitado e de fazer a diferença. Seus conselhos foram valiosos para mais esta conquista.

Um homem de caráter deve ser sempre fiel aos seus valores. Esse foi um dos ensinamentos mais importantes que já tive. O mentor desse e de outros ensinamentos tão valiosos chama-se Josué Soares (*in memoriam*), meu eterno grande amigo. Tive a honra de conhecê-lo, e de receber de você o verdadeiro amor de pai. Você é um ídolo para mim. Aprendi tanta coisa com você que o espaço e as palavras são insuficientes para expressar todas elas. Serei eternamente grato pelo seu exemplo, e irei honificar com meu compromisso.

No sentido mais amplo, o papel do orientador consiste em usar do seu conhecimento para potencializar a capacidade de um indivíduo para que ele escolha e assuma a direção do seu próprio destino. Dessa forma, agradeço imensamente ao meu orientador, o Prof. Dr. Tiago Maritan. Sem dúvida, um exemplo de educador. Sempre dedicado e disposto a contribuir com a construção do conhecimento, para que assim possamos viver em um mundo melhor. Agradeço todas as críticas e cobranças, com certeza foram essenciais para minha formação. Enfim, sou muito grato por tudo.

Agradeço também pela oportunidade de fazer parte da família dos guerreiros do Laboratório de Aplicações de Vídeo Digital, o LAViD. Grande parte dos conhecimentos adquiridos vieram das experiências vividas em projetos desenvolvidos por esse laboratório. Juntos, tivemos momentos tensos e momentos de conquistas importantes. De modo especial, quero expressar minha gratidão aos meus colegas, amigos e realizadores do LAViD nas pessoas de Guido Lemos, Danilo Assis, Erickson Silva, Jaelson Jr, Fernando Brito, Felipe Alves, Ícaro Magalhães, Mateus Pires, Manuella Aschoff, Virgínia

Campos, Carlos Hacks, Wesnydy Ribeiro, Renan Soares, David Guilherme, Claudio-mar Araújo, Ezequiel Silva, Victor Igor, Andrea Fernandes, Rostand Costa, Daniela Brauner, dentre outros colegas pelos momentos de fraternidade e pelas valiosas ideias e sugestões que contribuíram de alguma forma para a realização deste trabalho.

Por fim, agradeço a todos os meus familiares, amigos e colegas que, mesmo distante ou até mesmo que não tenham sido citados aqui explicitamente, sempre torceram por mim e me deram grande força para continuar nessa jornada.

*“Saber muito não lhe torna inteligente.
A inteligência se traduz na forma que você recolhe,
julga, maneja e, sobretudo,
onde e como aplica esta informação.”*

(Carl Sagan)

Resumo

Pessoas com deficiência auditiva ou visual enfrentam diversos entraves para participar da vida em sociedade, para se comunicar com outras pessoas e ter acesso à informação. A participação dessas pessoas em ambientes como cinemas, teatros, escolas, enquanto navegam na Internet, entre outros, é muitas vezes limitada devido a ausência de soluções acessíveis que viabilizem sua participação e interação. Para os surdos, por exemplo, os recursos de acessibilidade normalmente oferecidos nas salas de cinema se limitam a legendas no formato texto. Esta é a solução predominante no cenário atual. Porém, em contrapartida, existem algumas pesquisas que encontram-se em desenvolvimento para tentar resolver esses problemas, como a utilização de dispositivos de segunda tela para projetar os conteúdos com Língua de Sinais. Para os cegos, o problema de acessibilidade é ainda mais agravante, pois, atualmente, existem poucas salas de cinema que disponibilizam os recursos da Audiodescrição. Normalmente, os cegos contam apenas com a ajuda de familiares ou amigos para narrar os acontecimentos durante a exibição do filme. Do ponto de vista da acessibilidade, essas abordagens são consideradas ineficientes, uma vez que os surdos, por exemplo, têm bastante dificuldade de ler e escrever na língua oral do seu país. Portanto, para minimizar esses problemas, este trabalho apresenta uma solução computacional capaz de gerar automaticamente a partir das legendas, trilhas de vídeo com janelas de Libras para usuários surdos, e a partir do roteiro do filme, gerar Audiodescrição para usuários cegos e distribuir esses conteúdos para usuários usando dispositivos de segunda tela. Experimentos com um protótipo mostraram que a solução tem potencial para gerar de forma eficiente conteúdos inteligíveis para esses usuários, melhorando o seu acesso a esse tipo de tecnologia.

Palavras-chave: Cinema Digital, Acessibilidade, Deficiência Auditiva, Libras, Deficiência Visual, Audiodescrição, Aplicações Multimídia.

Abstract

People with hearing or visually impairment face many difficulties to participate in the social life, to communicate with other people and to access information. Their participation in cinemas, theaters, educational environment, among others, is often limited due to the lack of solutions that enable their interaction and participation. For the deaf, for example, the accessibility resources normally provided at the cinema room is limited to subtitle text format. This is the predominant solution on the current scenario. However, on the other hand, there are some studies that are in progress to try to solve these problems, such as the use of second screen devices to project content with sign language. For the blind, the accessibility problem is further aggravated because currently there are few cinema rooms that offer the resources of audio description. Normally, the blind rely solely on the help of family or friends to narrate the events during the exhibition of the film. The accessibility point of view, these approaches are considered inefficient, since the deaf, for example, have enough trouble reading and writing in the spoken language of their country. Therefore, to minimize these problems, this paper presents a computational solution capable of automatically generating from the subtitles, video tracks with Brazilian Sign Language (Libras) windows for deaf users, and from the movie script, produce audio description for blind users and distribute such content for users using the second screen devices. Experiments using a prototype showed that the solution has the potential to generate efficiently contents intelligible for this users, increasing your access to this technology type.

Keywords: Digital Cinema, Accessibility, Hearing Impairment, Libras, Visually Impairment, Audio Description, Multimedia Applications.

Conteúdo

1	Introdução	1
1.1	Motivação	3
1.2	Objetivos	5
1.3	Organização da Dissertação	7
2	Fundamentação Teórica	8
2.1	Deficiência Auditiva	9
2.2	Língua de Sinais	11
2.2.1	Língua Brasileira de Sinais	11
2.3	Deficiência Visual	14
2.3.1	Audiodescrição	16
2.4	Fogo Player	17
2.5	Considerações	20
3	Trabalhos Relacionados	21
3.1	Acessibilidade para Surdos em Segunda Tela	22
3.1.1	Interatividade dos Surdos em Sala de Aula	23
3.1.2	<i>Sony Glasses</i>	24
3.1.3	<i>Wireless Captioning</i>	25
3.1.4	<i>New Wireless Captioning System Debuts at Dallas</i>	27
3.2	Acessibilidade para Deficientes Visuais	28
3.2.1	<i>Collaborative Annotation for Video Accessibility</i>	28
3.2.2	<i>Describing Online Videos with Text-to-Speech Narration</i>	31
3.2.3	Whatscine	32

3.3	Considerações	34
4	Solução Proposta	37
4.1	Sistema de Acessibilidade	38
4.1.1	Servidor de Acessibilidade	39
4.1.2	Arquitetura do Sistema de Libras	40
4.1.3	Arquitetura do Sistema de Audiodescrição	43
4.1.4	Distribuidor	45
4.2	Protocolo de Comunicação	50
4.3	Considerações	53
5	Resultados e Discussões	55
5.1	Testes Computacionais	55
5.1.1	Ambiente de Testes	56
5.1.2	Análise dos Resultados	59
5.1.3	Análise do Tráfego de Rede	66
5.1.4	Discussão	68
5.2	Avaliação dos Conteúdos de Libras	69
5.3	Avaliação dos Conteúdos de Audiodescrição	74
6	Considerações Finais	78
6.1	Perspectivas de Trabalhos Futuros	79
	Referências Bibliográficas	85
A	Questionário aplicado no teste de inteligibilidade com os surdos	86
B	Artigo Completo Publicado no Webmedia'14	90
C	Prêmio de Melhor Artigo na Trilha Principal do Webmedia'14	91
D	Artigo Completo Publicado no <i>Journal of Information and Data Management</i> - JIDM	92

Lista de Figuras

2.1	Alfabeto da Língua Brasileira de Sinais.	13
2.2	Alfabeto do Sistema Braille.	15
2.3	Visão esquemática dos subsistemas de <i>Streamer</i> e <i>Player</i> . (Fonte: Aquino Júnior et al. (2013))	17
2.4	Operação de distribuição da arquitetura do Fogo Player. (Fonte: Aquino Júnior et al. (2013))	18
3.1	Visão arquitetural do sistema de geração de conteúdos acessíveis proposto por Tambascia et al. (2012).	23
3.2	Cenário de uso da tecnologia desenvolvida pela Sony (2012).	25
3.3	Exibição das legendas durante um jogo de Futebol Americano no <i>Dallas Cowboys Stadium</i>	27
3.4	Tecnologia formada por um editor de script, um player de vídeo, um repositório de metadados, e um formato baseado em texto para trocar os scripts de Audiodescrição entre os componentes (Fonte: Kobayashi et al. (2010)).	31
3.5	Arquitetura do Whatscine — fonte: (WHATSCINE, 2015).	33
4.1	Arquitetura do Sistema de Acessibilidade.	38
4.2	Visão arquitetural do sistema de geração de Libras.	41
4.3	Arquitetura do sistema de geração de Audiodescrição.	44
4.4	Organização do componente de distribuição.	45
4.5	Diagrama de sincronização entre cliente e servidor.	48
4.6	Estrutura das mensagens usadas na comunicação entre o sistema da solução proposta e o Fogo Player.	50

4.7	Cenário de comunicação entre a solução proposta e o Fogo Player. . . .	52
5.1	Vídeo com cronômetro usado no teste.	57
5.2	Momento da captura dos quadros: (a) 1459; (b) 14799; (c) 19813; (d) 42231.	58
5.3	(a) Momento em que o dispositivo 5 interrompe a exibição do conteúdo; (b) depois de solicitar novamente seu registro na aplicação, aguarda o processo de sincronismo; (c) reinicia a exibição do vídeo.	62
5.4	Variação de atraso mínimo, médio e máximo de cada dispositivo.	63
5.5	Dispersão do nível de sincronismo de cada dispositivo.	64
5.6	Dispersão geral do nível de sincronismo entre os dispositivos.	66
5.7	Análise do tráfego de rede durante o teste.	67
5.8	Informações estatísticas sobre o tráfego de rede após 13 minutos.	67
5.9	Informações estatísticas sobre o tráfego de rede após 27 minutos.	68
5.10	Percentual de acertos no teste de inteligibilidade.	71
5.11	Análise de dispersão dos resultados entre as abordagens investigadas. .	72
5.12	Resultado da avaliação do grau de dificuldade em usar um dispositivo móvel.	73
5.13	Resultado do nível de dificuldade para acompanhar as duas telas ao mesmo tempo.	73
5.14	Resultado do nível de compreensão dos conteúdos apresentados.	75
5.15	Nível de facilidade para entender os filmes utilizados no experimento. .	75
5.16	Benefícios da Audiodescrição gerada pela solução na compreensão dos filmes.	76

Lista de Tabelas

3.1	Efeitos das inserções de pausas (ENCELLE; BELDAME; PRIÉ, 2013). . .	30
3.2	Classificação das principais características dos trabalhos relacionados. .	36
4.1	Especificação dos tipos de mensagens utilizadas na comunicação entre a solução proposta e o Fogo Player.	51
5.1	Captura dos quadros referente a cada amostra do teste.	60
5.2	Atraso dos dispositivos em relação a cada amostra de quadro do servidor	61
5.3	Cálculos das medidas de dispersão.	65

Capítulo 1

Introdução

Avanços recentes das tecnologias de vídeo digital, assim como o desenvolvimento de câmeras de alta resolução, projetores digitais, modernos padrões de compressão de imagens, redes de computadores de alta velocidade, dispositivos com alta capacidade de processamento e armazenamento, entre outros, têm direcionado o desenvolvimento da era do cinema digital. Com esses recursos, atualmente, é possível capturar, editar e exibir vídeos digitais em ultra alta definição (do inglês, *Ultra High Definition* – UHD) com ou sem estereoscopia, devido as vantagens inerentes das mídias digitais sobre as analógicas sem perda de qualidade de imagem.

Além de proporcionar conteúdos audiovisuais de qualidade, a digitalização do cinema também abre espaço para o desenvolvimento e a inclusão de novos serviços e aplicações. Esses novos recursos trazem benefícios aos espectadores, assim também como expandem as funções do sistema, permitindo que os usuários participem de experiências inovadoras, tenham acesso à informação, interajam com os conteúdos audiovisuais, entre outros. Porém, diante dessa evolução tecnológica, surge o questionamento se os surdos e os cegos têm as mesmas condições de acesso à esses conteúdos? Ou seja, quais os recursos disponíveis para esses deficientes nos cinemas?

Atualmente, nos cinemas, os recursos de acessibilidade oferecidos para os usuários surdos são as legendas em formato de texto, que normalmente são sobrepostas a tela do filme, e que são compartilhadas igualmente com todos os usuários do ambiente. Essa é a solução mais usada no cenário atual. Porém, a forma natural de comunicação dos surdos é através das línguas de sinais, que no Brasil é denominada como Língua Brasileira de

Sinais (Libras). Então, para tentar encurtar as barreiras do acesso à informação dos surdos, não apenas no Brasil, existem algumas pesquisas na literatura científica cujo objetivo é direcionar o desenvolvimento de tecnologias e recursos para esses usuários em diversos ambientes (ARAÚJO et al., 2014), (FERREIRA et al., 2011), (BUTTUSSI; CHITTARO; COPPO, 2007), (KATO et al., 2010), (ARAÚJO, 2012), (TAMBASCIA et al., 2012), (SONY, 2012), (WEST et al., 2009), (WHATSCINE, 2015).

Devido os surdos se comunicarem naturalmente através das línguas de sinais, a utilização de legendas em formato de texto é considerada ineficiente, uma vez que, para os surdos, as línguas baseadas em sons — como o português do Brasil — acaba sendo uma segunda língua. As línguas de sinais possuem uma estrutura gramatical própria, composta pelos diversos níveis linguísticos, como morfologia, sintaxe e semântica. As línguas de sinais se caracterizam pela combinação de movimentos utilizando principalmente os membros superiores do corpo — mãos, braços, cabeça e tronco — para expressar diversos conceitos, entre eles, de caráter descritivo, emotivo, literal, metafórico, concreto e abstrato. Portanto, a ausência de conteúdos com língua de sinais nos cinemas torna-se um entrave no acesso à informação dos surdos.

Para os usuários com deficiência visual, o problema é ainda mais agravante, pois, além de não poder contar com as legendas, o número de salas de cinema que disponibilizam o recurso da Audiodescrição é bastante pequeno. Normalmente, os cegos contam apenas com a ajuda de familiares ou amigos para narrar os acontecimentos durante a exibição de um filme. A Audiodescrição foi concebida com o propósito de facultar à pessoas cegas a oportunidade de ter acesso à informação percebida apenas por meio da visão. Para Motta (2008), o recurso da Audiodescrição é um instrumento de inclusão cultural que irá contribuir para a formação crítica e para a educação da pessoa com deficiência visual, preparando-a para o exercício pleno de sua cidadania.

Dessa forma, em virtude das necessidades dos cegos, algumas pesquisas estão sendo desenvolvidas com o objetivo de oferecer à pessoas cegas, condições de acesso à informação em diferentes ambientes, como em vídeos disponíveis na internet, em salas de cinema, entre outros contextos (ENCELLE; BELDAME; PRIÉ, 2013), (KOBAYASHI et al., 2010), (WHATSCINE, 2015).

Nesse contexto, este trabalho apresenta uma solução cujo objetivo é gerar automa-

ticamente, conteúdos acessíveis para deficientes auditivos e deficientes visuais em salas de cinema, isto é, oferecer condições para que esses usuários possam ter acesso à trilhas de vídeo com janelas de Libras e Audiodescrição, respectivamente. A ideia é explorar a versatilidade dos dispositivos de segunda tela (*smartphones*, *tablets*, dentre outros) para disponibilizar os conteúdos acessíveis gerados pela solução, e então, dessa forma, que esta seja uma alternativa para a construção de sessões de cinema inclusivas.

Uma vez que o entretenimento de uma sessão de cinema é uma experiência coletiva, a apresentação dos conteúdos acessíveis na primeira tela pode incomodar as pessoas que não possuem deficiências sensoriais. Dessa forma, a exibição dos conteúdos em um dispositivo de segunda tela pode ser, portanto, uma alternativa viável para que todos os usuários de uma sala de cinema compartilhem o mesmo espaço e tenham acesso à apresentação do conteúdo de maneira compatível com as suas necessidades.

1.1 Motivação

O desenvolvimento inclusivo tem como objetivo promover a igualdade de oportunidades e direitos acessíveis a todos, através do projeto e execução de ações políticas voltadas para o desenvolvimento socioeconômico e humano, independente de classe social, gênero, raça e das condições físicas, mentais ou sensoriais (BIELER, 2005). Ele possui um conjunto de princípios básicos que norteiam o desenvolvimento de uma sociedade ativa e participativa, que dispõe de recursos para acessar informações, obter conhecimentos e meios para se comunicar com outras pessoas.

A acessibilidade é um recurso que possibilita não apenas a igualdade, mas também assegura os direitos das pessoas com deficiência. Por meio da acessibilidade, é possível melhorar a qualidade de vida dessas pessoas, principalmente com relação aos fatores perceptivos, cognitivos e motores (GALLAHUE; OZMUN, 2005). Em outras palavras, a acessibilidade consiste em desenvolver mecanismos, ferramentas e técnicas que possibilitem às pessoas com deficiência o acesso à espaços físicos, à informações, aos produtos e aos serviços, independente do tipo de deficiência apresentada.

Com base na literatura científica, existem alguns trabalhos que foram desenvolvidos com o propósito de oferecer recursos acessíveis para ajudar na comunicação e no acesso

à informação das pessoas com deficiência auditiva e visual em diversos cenários, como, por exemplo, no meio educacional, em teatros e cinemas, na Internet, dentre outros ambientes (TAMBASCIA et al., 2012), (WEST et al., 2009), (SONY, 2012), (KOBAYASHI et al., 2010), (ENCELLE; BELDAME; PRIÉ, 2013), (WHATSCINE, 2015).

Em Tambascia et al. (2012), o objetivo do trabalho é viabilizar a inclusão dos surdos em ambientes educacionais. Sua proposta consiste em equipar o professor com um microfone e os alunos surdos com dispositivos móveis. Dessa forma, todo o conteúdo transmitido pelo professor, será traduzido para a gramática da língua de sinais. A partir dessa tradução, será utilizado um dicionário de sinais para gerar um fluxo vídeo acessível (representados por um avatar 2D), que será transmitido juntamente com o texto para os dispositivos móveis dos alunos surdos.

Para West et al. (2009) e Sony (2012), o propósito é viabilizar a participação dos surdos nas salas de cinema. Para isso, foram utilizados óculos especiais com microdisplays apropriados para exibir as legendas do filme. Uma característica interessante dessa abordagem é que, embora posicionado próximo ao olho do usuário, o microdisplay utiliza lentes especiais que permitem ter a sensação de que o conteúdo está flutuando vários metros de distância. Com isso, segundo os pesquisadores, é possível que o usuário tenha uma visualização suave de ambas as imagens, ou seja, filme e legendas.

A partir da pesquisa de Encelle, Beldame e Prié (2013), foi criado o projeto ACAV (*Collaborative Annotation for Video Accessibility*). A ideia do projeto consiste em inserir janelas de interrupção – uma espécie de *pseudo* pausas – nos intervalos do vídeo onde há silêncio (isto é, inexistência de diálogos ou elementos sonoros importantes), para que, dessa forma, possa ser introduzido o recurso da Audiodescrição, fundamental para que os deficientes visuais tenham acesso ao conteúdo do filme. Então, preocupados em não prejudicar a compreensão dos conteúdos, devido a inserção de interrupções, os pesquisadores realizaram alguns experimentos. Os resultados mostraram que as pausas artificiais foram bem aceitas pelos deficientes visuais.

Em virtude da grande quantidade de vídeos que trafegam na Internet, Kobayashi et al. (2010) propuseram uma plataforma Web para inserir Audiodescrição nos vídeos disponíveis na rede. A abordagem utiliza a tecnologia *Text-to-Speech* (TTS) para gerar conteúdos com voz sintética e *scripts* de Audiodescrição em formato de texto, onde esses

scripts contém as descrições das imagens e o tempo de apresentação de cada descrição. Os resultados obtidos através de experimentos mostraram que os conteúdos gerados pelo sistema foram considerados satisfatórios.

Concebida pela *Universidad Carlos III de Madrid*, Espanha, a tecnologia do Whatscine (2015) tem o objetivo de oferecer conteúdos acessíveis para surdos e para cegos nas salas de cinema. A ideia do Whatscine é tornar as salas de cinema acessíveis, de modo que as pessoas com dificuldade de enxergar possam ouvir a Audiodescrição do filme sem interferências no áudio principal ou incômodos para resto do público, e que as pessoas com problemas auditivos tenham acesso às legendas ou à tradução em língua de sinais por meio de óculos especiais ou *smartphones*. Até onde foi encontrado, o Whatscine (2015) utiliza apenas conteúdos previamente gerados e, além disso, não foram encontrados resultados que mostrassem a eficiência da ferramenta com relação a distribuição dos conteúdos.

Apesar da contribuição científica, os trabalhos desenvolvidos por Tambascia et al. (2012), West et al. (2009) e Sony (2012), por usar legendas em formato de texto, não endereçam as necessidades da maioria das pessoas surdas, uma vez que esses usuários possuem muita dificuldade para ler e escrever na língua oral do seu país (WAUTERS, 2005), (MENEZES; CAVALCANTE, 2008), (IBGE, 2000). Do mesmo modo, as soluções apresentadas em (KOBAYASHI et al., 2010), (ENCELLE; BELDAME; PRIÉ, 2013) e (WHATSCINE, 2015) possuem problemas que, para a disponibilização dos conteúdos acessíveis, precisam inserir pausas no vídeo principal, ou não fazem distribuição para multidispositivos, ou então necessitam que os conteúdos sejam previamente gerados, ou seja, não oferecem suporte à geração desses conteúdos.

1.2 Objetivos

Este trabalho tem como objetivo geral o desenvolvimento de uma solução computacional prática e eficiente, capaz de gerar automaticamente trilhas alternativas de vídeo com língua de sinais em Libras e Audiodescrição em salas de cinema. A ideia é que os usuários com deficiência auditiva ou com deficiência visual tenham acesso aos conteúdos através de dispositivos de segunda tela, como *smartphones*, *tablets*, óculos equipados

com microdisplay, dentre outros. Então, também como resultado dessa pesquisa, é esperado que esses usuários tenham acesso à conteúdos intelegíveis, e que, dessa forma, tenham condições de participar igualmente das sessões de cinema.

Para alcançar o objetivo geral deste trabalho, foram definidos os objetivos específicos. Esses objetivos foram divididos e são elencados a seguir:

- a) Desenvolvimento de um protocolo de comunicação entre a solução proposta e o player de cinema, de modo que possibilite obter os arquivos de legenda e do roteiro do filme, assim como informações temporais, utilizadas para controlar o sincronismo entre os servidores de aplicação;
- b) Definição da arquitetura da solução proposta. O desenho arquitetural precisa ilustrar a organização dos componentes que fazem a comunicação com o player de cinema e o relacionamento entre os componentes internos do sistema;
- c) Implementação dos componentes de software definidos no desenho arquitetural da solução. E então, dessa forma, construir um protótipo do sistema para a realização de experimentos computacionais e com usuário;
- d) Elaboração de questionários para a aplicação dos testes de inteligibilidade. Esses questionários precisam direcionar a investigação para o nível de compreensão dos deficientes auditivos e dos deficientes visuais com relação aos conteúdos gerados pela solução proposta;
- e) Execução de testes computacionais com o protótipo da solução. O objetivo desses testes é validar a estratégia de distribuição de conteúdos, ou seja, medir o nível de sincronismo entre os dispositivos móveis e o servidor de distribuição desses conteúdos;
- f) Execução de testes de inteligibilidade. O objetivo desses testes é avaliar o nível de compreensão dos conteúdos gerados pela solução. Então, dessa forma, devem ser aplicados os questionários para os dois grupos de usuários, ou seja, para os deficientes auditivos e para os deficientes visuais.

1.3 Organização da Dissertação

A estrutura deste trabalho está organizada da seguinte forma. O primeiro Capítulo contextualiza o problema de pesquisa, ressaltando as principais dificuldades que os surdos e os cegos enfrentam para poder se comunicar e ter acesso à informação especialmente em ambientes de entretenimento como os cinemas e teatros, no âmbito educacional e nos conteúdos disponíveis na Internet.

O segundo Capítulo, por sua vez, apresenta a fundamentação teórica dos assuntos inerentes a linha de pesquisa abordada neste trabalho. Neste Capítulo, alguns conceitos-chave, como, por exemplo, acessibilidade, deficiência auditiva, deficiência visual, Língua Brasileira de Sinais (Libras), Audiodescrição e o player de cinema — denominado Fogo Player — são apresentados e discutidos em detalhes.

O terceiro Capítulo apresenta os principais trabalhos correlatos encontrados. Este Capítulo foi dividido em duas partes. A primeira apresenta as contribuições desenvolvidas para pessoas com problemas de audição. A segunda parte deste Capítulo discute os trabalhos e as tecnologias desenvolvidas para dar capacidade e autonomia às pessoas com deficiência visual em ambientes como cinema, nas escolas e na Internet.

O quarto Capítulo apresenta a solução proposta neste trabalho. Este Capítulo ilustra a arquitetura do sistema de acessibilidade, constituída pelos dois módulos responsáveis pela geração dos conteúdos acessíveis, o módulo *Libras* e o módulo *Audiodescrição*. Além disso, este Capítulo também apresenta a especificação do protocolo de comunicação entre o player de cinema e o servidor de comunicação da solução proposta desenvolvida neste trabalho.

O quinto Capítulo apresenta os resultados obtidos após a realização de experimentos com o protótipo da solução. O objetivo deste Capítulo é mostrar os resultados referentes ao nível de compreensão e de satisfação dos usuários surdos e dos usuários cegos com relação aos conteúdos gerados. Além disso, este Capítulo também apresenta os resultados dos testes computacionais, cujo objetivo é medir o nível de sincronismo entre os dispositivos móveis e o servidor de distribuição.

Por fim, o sexto Capítulo apresenta as considerações finais: as contribuições, as dificuldades no desenvolvimento da pesquisa e as propostas de trabalhos futuros.

Capítulo 2

Fundamentação Teórica

O objetivo deste Capítulo é apresentar os principais conceitos e definições a respeito dos termos abordados neste trabalho. Para isso, inicialmente, são expostas as características e propriedades fundamentais que compreendem a deficiência auditiva (Seção 2.1). Ainda nesse contexto, são apresentados os principais conceitos relacionados à Línguas de Sinais (Seção 2.2), e especialmente à Língua Brasileira de Sinais – LIBRAS (Seção 2.2.1), principal recurso usado pelos surdos brasileiros como forma de comunicação e expressão.

Em seguida, são apresentados alguns aspectos da deficiência visual (Seção 2.3), onde são caracterizadas as alterações graves ou total das funções elementares da visão. Para as pessoas que sofrem com a deficiência visual, a audiodescrição é um recurso muito importante para obter o acesso à informação de modo independente. Portanto, por essa razão, as técnicas e os principais conceitos deste mecanismo também são apresentados e discutidos neste Capítulo, na Seção 2.3.1.

Por fim, este Capítulo apresenta a tecnologia denominada Fogo Player na Seção 2.4, que consiste basicamente em um sistema de reprodução de conteúdos multimídia em salas de cinema. Então, no contexto deste trabalho, o Fogo Player é responsável também por integrar a solução proposta como componente de acessibilidade, tornando, dessa forma, as salas de cinema acessíveis à surdos e à cegos.

2.1 Deficiência Auditiva

De acordo com a Febraban (2010), a deficiência auditiva pode ser compreendida como aquela que, por motivo de perda ou anomalia congênita ou adquirida, parcial ou total, de estrutura ou função da audição, pode resultar em restrições da capacidade de comunicação, de interpretação sobre as condições de segurança e de orientação.

Segundo o Decreto de Lei nº 5.296, de 2 de Dezembro de 2004 (BRASIL, 2004), da constituição federal brasileira, a deficiência auditiva consiste na perda bilateral, parcial ou total, de quarenta e um decibéis (41 dB) ou mais, aferida por audiograma nas frequências de 500Hz, 1.000Hz, 2.000Hz e 3.000Hz. A insuficiência do aparelho auditivo pode ser compreendida, de maneira geral, como o contraste que existe entre a performance do indivíduo e a habilidade normal de percepção sonora de acordo com os padrões estabelecidos pela *American National Standards Institute* (ANSI 1989).

Em muitos casos, confundida ou mal interpretada, a deficiência auditiva não implica que o indivíduo perdeu completamente a capacidade de percepção dos sons. Embora existam aqueles que são completamente surdos, há também casos em que a lesão do aparelho auditivo não compromete a audição literalmente. Nesses casos, é bastante comum afirmar que o indivíduo possui perda auditiva funcional.

Para a comissão organizadora da Brasilmedia (2012), a perda auditiva funcional pode ser classificada da seguinte forma:

- *Leve*: o indivíduo é incapaz de perceber a emissão de sons reproduzidos com frequência abaixo de 30 dB. Nessas condições, participar de discursos, por exemplo, pode resultar em situações frustrantes, principalmente se o ambiente dispor de muitos ruídos;
- *Moderada*: nesta categoria, sons abaixo de 50 dB tendem a não serem perceptíveis. Dependendo da ocasião, para que não ocorram perdas de informações e possíveis circunstâncias de desconforto, é recomendada a utilização de um aparelho ou prótese auditiva;
- *Severa*: o indivíduo não é capaz de perceber ruídos sonoros abaixo de 80 dB. Em algumas situações, os aparelhos e próteses auditivas podem ser úteis, no

entanto, nem sempre são eficazes. Alguns indivíduos com perda auditiva severa se comunicam através de uma língua baseada em gestos, e outros recorrem às técnicas de leitura labial;

- *Profunda*: ausência da capacidade de ouvir, ou a incapacidade de ouvir sons abaixo de cerca de 95 dB. Nesse caso, é fundamental a utilização das línguas baseadas em gestos ou habilidades de leitura labial para a comunicação e acesso à informação.

Na perspectiva do Conselho Federal e Regional de Fonoaudiologia (CFRF, 2009), a classificação do tipo de perda auditiva deve levar em consideração a comparação entre os limiares da via aérea e da via óssea de cada orelha. Dessa forma, é imprescindível a realização de um estudo clínico aprofundado dos limiares tonais por via aérea e óssea para uma conclusão ou diagnóstico mais preciso. A classificação dos tipos de deficiência auditiva e das características inerentes a cada uma delas pode dividida da seguinte maneira:

- *Perda auditiva condutiva*: avaria localizada na região externa ou intermediária do composto auditivo. Quando ocorrem lesões nessas regiões, os ossículos (martelo, bigorna e estribo) podem não vibrar adequadamente, impedindo a detecção das informações sonoras;
- *Perda auditiva neurosensorial ou sensório-neural*: apresenta problemas na região interna da estrutura auditiva ou, no sistema auditivo periférico, cientificamente denominado patologia coclear e retrococlear. Nesse caso, os ossos da parte interna podem vibrar perfeitamente, mas os nervos auditivos responsáveis são incapazes de transmitir informações sonoras para o cérebro;
- *Perda auditiva mista*: abrange tanto a perda condutiva quanto a neurosensorial. Esta é comumente diagnosticada perda auditiva por apresentar problemas nas regiões externas, intermediária e interna do composto auditivo. O tratamento para esses casos pode ser feito através de procedimentos cirúrgicos ou, com o auxílio de aparelhos ou próteses auditivas.

2.2 Língua de Sinais

A comunicação é um mecanismo fundamental para a vida das pessoas. Usamos este recurso em grande parte das atividades que realizamos, seja em conversas com os amigos, assistindo a programas de entretenimento, em gestos de desaprovação no trânsito, navegando na internet, dentre outros. Em virtude disso, podemos assumir que a comunicação é um fator imprescindível para a sobrevivência das pessoas.

Para que seja efetiva, é preciso que a comunicação seja clara, consistente e livre de obstruções, pois havendo ruídos no canal de transmissão (meio que interliga o emissor e o receptor), a informação emitida pode ser comprometida, perdida ou até mesmo impossibilitada. Dessa forma, é importante que a comunicação esteja fundamentada nos quatro pilares da educação pela UNESCO (2010), como: aprender a conhecer; aprender a fazer; aprender a conviver; e aprender a ser. Para o Passadori (2012), a melhoria do processo educacional será possível apenas se a capacidade de comunicação dos elementos que compõem a rede da Educação (família, escola, mídias e sociedade) estiverem desenvolvidas e preparadas para cumprir seu papel.

No mundo, cada país possui um idioma no qual as pessoas podem se comunicar, transmitir e ter acesso à informação. Do mesmo modo, em cada região, as pessoas com deficiência auditiva utilizam uma língua baseada em gestos, denominada Língua de Sinais, para obter o mesmo fim. A combinação de movimentos utilizando principalmente os membros superiores do corpo (isto é, mãos, braços, cabeça e tronco) permite que o emissor expresse diversos conceitos, como, por exemplo, de caráter descritivo, emotivo, literal, metafórico, concreto e abstrato, e que o receptor perceba e compreenda esses conceitos utilizando o sistema visual (SILVA, 2012). Na Subseção 2.2.1 é apresentada a Língua Brasileira de Sinais, que é o principal mecanismo de comunicação utilizado pelos surdos brasileiros.

2.2.1 Língua Brasileira de Sinais

Segundo a Lei Brasileira No 10.436, de 24 de Abril de 2002 (BRASIL, 2002a), a língua de sinais oficial do Brasil, e utilizada pela maioria dos surdos brasileiros, é a Língua Brasileira de Sinais - Libras. A Libras, assim como as outras línguas de sinais, possui

uma estrutura gramatical própria, composta pelos diversos níveis linguísticos como, morfologia, sintaxe e semântica. Assim como as línguas orais, a Libras também possui itens léxicos, denominados sinais.

Os sinais em Libras são compostos por fonemas, que são as unidades básicas da comunicação baseada nessa estrutura. Segundo Buttussi, Chittaro e Coppo (2007), um sinal consiste, e é, unicamente identificado por cinco fonemas:

1. *Configuração da mão*: representa a posição dos dedos e seus movimentos. Um sinal pode ser produzido pela mão predominante (mão direita para os destros) ou pelas duas mãos. Uma configuração de mão pode diferenciar-se das demais pela extensão (lugar e número de dedos estendidos), pela contração (mão aberta ou fechada), e pelo contato ou divergência dos dedos. Segundo Felipe e Monteiro (2005), em Libras existem atualmente sessenta possíveis configurações de mão, dentre as quais vinte e seis são usadas para representar as letras do alfabeto (ver Figura 2.1).
2. *Ponto de Articulação*: representa a parte do corpo do emissor onde os sinais são realizados (ou começam a ser realizados). Em Libras, esse ponto de articulação pode estar localizado em alguma parte do corpo ou estar localizado em um espaço neutro vertical (do meio do corpo até a cabeça) e horizontal (à frente do emissor). Por exemplo, os sinais TRABALHAR, BRINCAR, e CONSERTAR são executados no espaço neutro, enquanto que os sinais ESQUECER, APRENDER e PENSAR são feitos na região próximo da testa.
3. *Movimento*: um sinal pode conter movimentos ou não. O movimento em um sinal representa o deslocamento da mão no espaço durante a execução do sinal. A tarefa de caracterizar um sinal não é fácil, uma vez que a mão é um objeto assimétrico e que os eixos podem se deslocar simultaneamente. Segundo Fusco (2004), no entanto, a maioria dos sinais podem ser divididos em pequenos segmentos de movimentos, onde cada um está relacionado a um eixo. Nesse caso, por exemplo, os sinais em Libras: AJOELHAR e EM-PÉ, não possuem movimento.
4. *Orientação*: a orientação de um sinal representa a direção ou orientação do movimento. Os sinais que possuem algum tipo de movimento são executados em

uma determinada direção. Em Libras, a inversão da direção pode representar a ideia de oposto, contrário ou modificar a concordância do tipo número-pessoal do sinal.

5. *Expressões não manuais*: muitos sinais podem requerer características adicionais para expressar sentimentos de alegria, de tristeza, uma pergunta ou uma exclamação, e necessitam de um traço diferenciador, como uma expressão facial e/ou corporal, para poder expressar essas características. Dessa forma, as expressões não manuais podem assumir tanto uma função léxica (expressar alegria e tristeza, por exemplo) quanto uma função sintática (expressar uma pergunta ou uma exclamação, por exemplo). A inclusão de uma expressão facial e/ou corporal no discurso pode alterar o significado de um sinal, assim como sua ausência pode produzir um sinal sem sentido.

Conforme Buttussi, Chittaro e Coppo (2007), a partir da combinação desses cinco fonemas formam-se os sinais, que permitem expressar diferentes ideias ou aspectos de uma sentença. A comunicação em Libras ou em qualquer outra língua de sinais, portanto, nada mais é do que a combinação desses elementos para formar os sinais e para formar sentenças em um determinado contexto.

Figura 2.1: Alfabeto da Língua Brasileira de Sinais.



Da mesma forma que acontece com as línguas de modalidade oral, as línguas de sinais possuem um conjunto de sinais para representar o alfabeto. Em Libras, por exemplo, para cada sinal do alfabeto existe um gesto que é representado com as mãos. O alfabeto manual das línguas de sinais teve origem pela necessidade de representar as letras de forma visual, e era usado principalmente para ensinar pessoas surdas a ler e escrever (Silva, et al. 2007). A Figura 2.1 ilustra as configurações de cada sinal que constituem o alfabeto da Língua Brasileira de Sinais (Libras).

2.3 Deficiência Visual

De acordo com a *International Classification of Diseases* (ICD-10), comissão integrante da *World Health Organization*, responsável pelo diagnóstico padrão da epidemiologia, gestão de saúde e efeitos clínicos, existem quatro níveis em que a visão humana pode ser classificada: visão normal; deficiência visual moderada; deficiência visual severa; e cegueira, que é quando o indivíduo não possui capacidade alguma de enxergar.

Para Nicolaiewsky e Correa (2009), a deficiência visual pode ser dividida em duas categorias: cegueira e visão reduzida. Quando um indivíduo é diagnosticado com a cegueira, significa que, em termos pedagógicos, sua deficiência torna-se um entrave no processo educacional, como aprender a ler e a escrever, por exemplo. Por outro lado, a formação educacional daqueles que possuem visão reduzida é algo mais tangível, porém, mesmo assim, necessitam de algumas adaptações, como, por exemplo, o aumento no tamanho das letras, ou o uso de lentes especiais de aumento.

A acuidade visual é a capacidade que um indivíduo tem de perceber estímulos minúsculos de um sentido ou do grau de sensibilidade. De acordo com Carvalho (1994), quando a acuidade visual de um indivíduo está abaixo de 10%, e a percepção do campo visual está abaixo de 20 graus, este pode ser considerado cego. Por outro lado, em situações em que o indivíduo tem acuidade visual entre 10% e 30%, e percepção do campo visual entre 20 e 50 graus, e que não há tratamento clínico (lentes ou óculos convencionais, por exemplo) ou cirúrgico, a deficiência pode ser classificada como visão reduzida.

É importante ressaltar que a deficiência visual é um problema que afeta o mundo

inteiro. Segundo a *World Health Organization* (ORGANIZATION, 2014), a estimativa é que, no mundo, existam cerca de 285 milhões de pessoas com algum nível de deficiência visual, onde, dentre esses, 39 milhões não enxergam completamente nada, isto é, são cegos, e os outros 246 milhões possuem baixa capacidade de visão, ou visão reduzida.

Com o propósito de promover a inclusão social das pessoas com deficiência visual, foi reconhecido por meio da Portaria nº 2.678 de 24 de Setembro de 2002 (BRASIL, 2002b), o Sistema Braille, que compreende a Língua Portuguesa do Brasil em todas as suas modalidades e aplicações. Criado por Louis Braille no ano de 1827, o Sistema Braille é constituído por 63 sinais formados por pontos em alto relevo a partir do conjunto matricial $\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$ (123456). Dessa forma, a pessoa com deficiência visual pode fazer a leitura de textos por meio do tato. A Figura 2.2 mostra a representação do alfabeto no Sistema Braille.

Figura 2.2: Alfabeto do Sistema Braille.

$\begin{smallmatrix} \bullet & \\ & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet \\ & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet \\ & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet & \\ & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet & \\ & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \end{smallmatrix}$
A	B	C	D	E	F	G	H	I	J
1	2	3	4	5	6	7	8	9	0
$\begin{smallmatrix} \bullet & \\ \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet & \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$
K	L	M	N	O	P	Q	R	S	T
$\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet & \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$
U	V	X	Y	Z	Ç	É	Á	È	Ú
$\begin{smallmatrix} \bullet & \\ \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & & \bullet & \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$	$\begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{smallmatrix}$
Â	Ê	Î	Ô	Ù	À	Ï	Ü	Ö	Ò/W

O Sistema Braille permitiu que os cegos tivessem mais oportunidades em adquirir conhecimento intelectual, como, filosofia, psicologia, teologia, matemáticas, história, literatura, direito, dentre outros. Porém, apesar dos benefícios, o Braille tem sido pouco usado pela comunidade de deficientes visuais. Um dos fatores que contribuem para isso é a escassez de materiais em Braille disponíveis nas bibliotecas, que podem

eventualmente não corresponder às necessidades dos potenciais utilizadores.

Com o advento das tecnologias assistivas, ao invés do Braille, os cegos tendem a recorrer preferencialmente aos textos introduzidos no computador, que a partir de comandos do teclado ou por reconhecimento de voz, facilmente leem todo o conteúdo do material usando uma voz sintética ou uma voz de um profissional narrador (ou descritor) que fora gravada previamente.

Na Seção 2.3.1, será apresentado um recurso bastante usado pelos deficientes visuais para a obtenção de informações que só podem ser captadas com o mecanismo da visão, a Audiodescrição.

2.3.1 Audiodescrição

A Audiodescrição é um recurso de acessibilidade que foi criado com o objetivo de proporcionar às pessoas cegas a oportunidade de ter acesso às informações que só podem ser percebidas por meio da visão. De acordo com Cintas (2005), a Audiodescrição consiste na transformação de imagens visuais em palavras, que então podem ser faladas nos momentos onde não há diálogo ou informações sonoras relevantes nos programas audiovisuais ou de performances ao vivo.

Para o profissional responsável pela narração dos conteúdos visuais, denominado audiodescritor, sua opinião não deve, em hipótese alguma, fazer parte das descrições, devendo este apenas descrever as informações daquilo que está sendo visto, de forma clara e coerente. Segundo Pozzobon (2010), o papel do audiodescritor é ser uma ponte entre aquele evento e o sujeito cliente do serviço, devendo dar a este os subsídios necessários e pertinentes à compreensão do evento.

Para o Ministério das Comunicações, órgão do poder Executivo brasileiro encarregado das políticas de radiodifusão, a Audiodescrição pode ser compreendida como:

“A narração, em língua portuguesa, integrada ao som original da obra audiovisual, contendo descrições de sons e elementos visuais e quaisquer informações adicionais que sejam relevantes para possibilitar a melhor compreensão desta por pessoas com deficiência visual e intelectual” (BRASIL, 2010).

Além de atender as necessidades das pessoas com deficiência visual, a Audiodes-

criação também pode ser usada para beneficiar pessoas com deficiência intelectual e dislexia, e ainda, os idosos, pois:

“A audiodescrição é um instrumento de inclusão cultural que irá contribuir para a formação crítica e para a educação da pessoa com deficiência visual, preparando-a para o exercício pleno de sua cidadania. Além disso, os benefícios de tal recurso estendem-se, também, às pessoas com deficiência intelectual, pessoas idosas e pessoas com dislexia” Motta (2008).

Dessa forma, pode-se perceber que a Audiodescrição é um recurso que permite não apenas o acesso à informação, mas, sobretudo, a promoção da igualdade social para as pessoas com deficiência, nesse caso, especialmente visual.

2.4 Fogo Player

Segundo Aquino Júnior et al. (2013), o Fogo Player é uma solução computacional que tem como objetivo viabilizar a reprodução e transmissão de vídeos com resolução 4K (4096 x 2160 pixels), isto é, quatro vezes o tamanho de um vídeo em *Ultra High Definition* (UHD). Para isso, o Fogo Player utiliza uma arquitetura escalável, flexível e de baixo custo de desenvolvimento. A arquitetura do Fogo Player é apresentada na Figura 2.3.

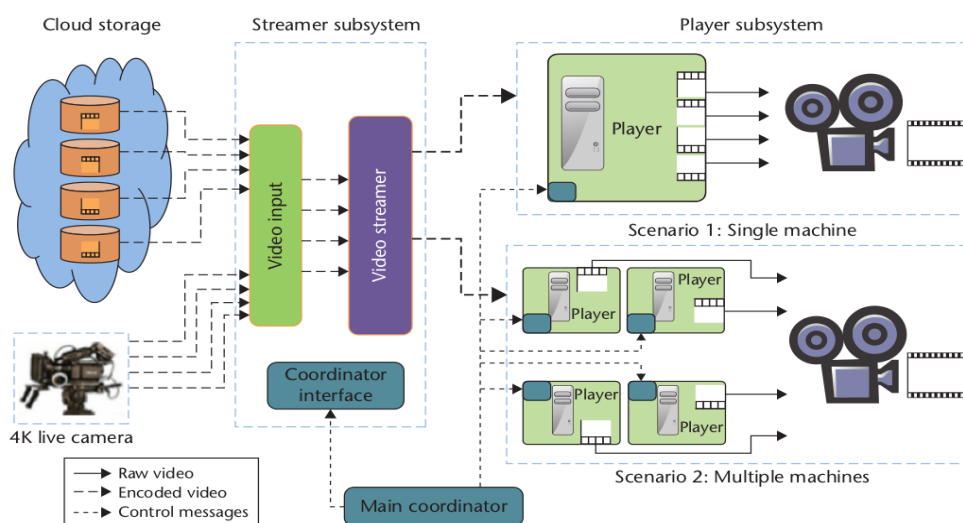


Figura 2.3: Visão esquemática dos subsistemas de *Streamer* e *Player*. (Fonte: Aquino Júnior et al. (2013))

A possibilidade de utilizar mais de um distribuidor de conteúdos, por exemplo, é uma característica que torna o Fogo Player uma solução viável para o cenário de aplicações multimídia distribuídas. Através desse recurso, é possível que um vídeo de alta resolução seja dividido espacialmente em quadrantes e que cada um desses quadrantes seja transmitido simultaneamente por fontes de distribuição diferentes.

De acordo com a Figura 2.3, a arquitetura de alto nível do Fogo Player é constituída por quatro subsistemas principais: Distribuição, Pré-processamento, Streamer e Player. Cada um desses subsistemas possui um conjunto de componentes que são responsáveis por uma série de requisitos inerentes ao sistema. A Figura 2.4 mostra, de modo simplificado, as operações realizadas pelo subsistema de Distribuição.

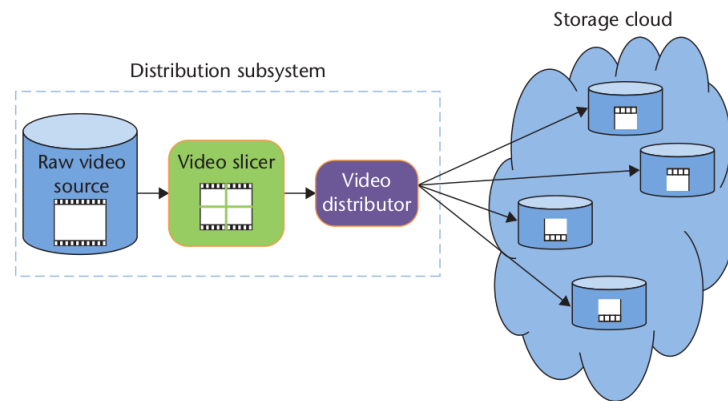


Figura 2.4: Operação de distribuição da arquitetura do Fogo Player. (Fonte: Aquino Júnior et al. (2013))

O subsistema de Distribuição (ilustrado na Figura 2.4) é responsável por receber o conteúdo do vídeo em formato original (isto é, sem nenhuma formatação), dividi-lo espacialmente em quadrantes e, em seguida, transmitir cada um desses quadrantes para um repositório de armazenamento intermediário, disponível em uma infraestrutura de nuvem. Uma vez que todos os quadrantes se encontram no repositório, o subsistema de Pré-Processamento irá codificar simultaneamente cada um deles usando padrões de codificação tal como H.264 ou JPEG 2000.

O subsistema de Streamer é responsável por representar uma fonte de conteúdos multimídia abstrata, cujo propósito principal é encapsular o processo de aquisição e transmissão desses conteúdos (ver Figura 2.3).

O subsistema de Player oferece suporte para dois cenários: *single and multiple*

machine. No primeiro (*single machine*), é necessário dispor de apenas uma máquina para receber e reproduzir os conteúdos transmitidos pelo subsistema de Streamer. No entanto, para casos em que os vídeos transmitidos são de alta resolução, esse cenário pode não ser eficaz, pois exigirá bastante desempenho da máquina receptora. Em contrapartida, o segundo cenário (i.e. *Multiple machines*) possibilita a utilização de várias máquinas receptoras (ou player). Com essa abordagem, é possível que cada uma das máquinas disponíveis se responsabilize por apenas um quadrante do vídeo original, transmitido por uma das fontes de distribuição.

Em linhas gerais, a responsabilidade do subsistema *Player* consiste em receber os conteúdos do vídeo, decodificar esses conteúdos, criar uma camada de abstração entre os dispositivos do modo gráfico e, em seguida, disponibilizar uma interface de coordenação para receber eventuais mensagens de controle, essenciais para a sincronização e exibição dos conteúdos no ambiente de projeção.

A coordenação dos componentes do Fogo Player é feita pelo *Main Coordinator* (ver Figura 2.3). O *Main Coordinator* é responsável, dentre outros, pela preservação da referência temporal entre o subsistema *Streamer* e as demais máquinas que compõem o subsistema *Player*. Dessa forma, quando os subsistemas estão executando em máquinas que estão em redes diferentes, o componente *Main Coordinator* utiliza um sistema de sincronização auxiliar para manter o relógio interno das máquinas participantes sincronizados. Além disso, outro papel fundamental do *Main Coordinator* é o atendimento às requisições externas, ou seja, ele funciona como uma interface entre os clientes e as funções disponíveis pelo Fogo Player.

O Fogo Player foi desenvolvido pelo Núcleo de Pesquisa e Extensão do Laboratório de Aplicações de Vídeo Digital (NPE/LAVID), na Universidade Federal da Paraíba (UFPB). Então, como o Fogo Player foi concebido para atuar em plataformas de Cinema Digital, e a solução proposta neste trabalho está inserida no mesmo contexto e, além disso, foi desenvolvida dentro do mesmo laboratório, a integração dos dois sistemas se tornou algo favorável. Dessa forma, a criação do protocolo de comunicação (ver Seção 4.2) entre os sistemas se deu sem grandes complicações.

Maiores informações sobre o Fogo Player podem ser encontradas em Aquino Júnior et al. (2013).

2.5 Considerações

Os surdos se comunicam naturalmente através das línguas de sinais, e outros ainda contam com a habilidade de leitura orofacial (ou leitura labial), que consiste na compreensão de mensagens por meio da observação dos lábios da pessoa falante. No Brasil, a língua de sinais utilizada pelos surdos é a Língua Brasileira de Sinais (Libras). Dessa forma, por meio da Libras, os surdos conseguem se comunicar e ter acesso à informação em qualquer lugar, desde que este recurso esteja disponível.

Para os cegos, a principal dificuldade está na percepção de imagens, ou seja, na capacidade de enxergar tudo aquilo que está ao nosso redor. Para Nicolaiewsky e Correa (2009), a deficiência visual pode ser classificada de duas formas: a cegueira, onde o indivíduo não enxerga completamente nada; e a visão reduzida, onde o deficiente é capaz de distinguir alguns objetos, formas ou tons de cores. Para os indivíduos com cegueira, a Audiodescrição surge como um recurso acessível capaz de descrever por meio de palavras tudo aquilo que só podemos perceber com o auxílio da visão.

No contexto dos cinemas digitais, os recursos de acessibilidade disponíveis para os surdos são limitados às legendas no formato de texto, e para os cegos, o problema é ainda maior, pois, contam apenas com o áudio original do filme. Dessa forma, a solução apresentada neste trabalho utilizou o sistema do Fogo Player para obter os conteúdos das legendas e do roteiro do filme para gerar e distribuir recursos de acessibilidade em Libras e Audiodescrição para os surdos e os cegos, respectivamente.

O Capítulo 3 apresenta alguns trabalhos encontrados na literatura científica com o propósito de oferecer recursos de acessibilidade para deficientes auditivos e visuais em diversos ambientes.

Capítulo 3

Trabalhos Relacionados

Na comunidade científica, existem diversas contribuições tecnológicas cujo objetivo consiste em oferecer recursos de acessibilidade que venham a facilitar a comunicação e o acesso à informação de pessoas com deficiência auditiva ou com deficiência visual. Portanto, nesse contexto, este Capítulo apresenta algumas contribuições consideradas relevantes para o tema abordado neste trabalho.

Como o público investigado nesta pesquisa é formado por duas categorias de usuários, isto é, surdos e cegos, as contribuições apresentadas neste Capítulo estão divididas de acordo com o tipo de usuário. Dessa forma, inicialmente, na Seção 3.1, são apresentados e discutidos os trabalhos que têm como objetivo oferecer recursos de acessibilidade em segunda tela para os surdos em diferentes contextos, como, por exemplo, em salas de aula, nos cinemas, teatros, estádios de futebol, dentre outros ambientes. Em seguida, na Seção 3.2, são apresentados os trabalhos voltados para dar condições de acesso à informação a usuários cegos em diferentes ambientes.

Por fim, a Seção 3.3 deste Capítulo apresenta as considerações finais, onde, além de recapitular os principais pontos chave discutidos, também é feita uma análise comparativa entre as principais contribuições investigadas na literatura científica e a solução proposta neste trabalho.

3.1 Acessibilidade para Surdos em Segunda Tela

Estudos empíricos têm mostrado que pessoas com deficiência auditiva, seja ela parcial ou total, encontram muitos obstáculos para se comunicar e ter acesso à informação em ambientes de participação coletiva, como, por exemplo, nos cinemas, teatros, centros de ensino, dentre outros (TAMBASCIA et al., 2012), (KATO et al., 2010), (BUTTUSSI; CHITTARO; COPPO, 2007), (SONY, 2012), (WEST et al., 2009). Esses problemas ocorrem principalmente pela falta de soluções que atendam, efetivamente, as necessidades desse público.

Atualmente, de acordo com a literatura, esse problema pode ser resolvido de três de formas. A primeira consiste na transcrição dos diálogos que compõem a exibição de filmes, programas de TV e jogos eletrônicos, por exemplo, acompanhando o mesmo em sobreposição. Esse mecanismo é geralmente denominado de legenda (TAMBASCIA et al., 2012), (WEST et al., 2009), (SONY, 2012). Todavia, Lima (2006) ressalta que os surdos têm grandes dificuldades para ler e escrever na língua oral do seu país.

A segunda alternativa é utilizar um intérprete de línguas de sinais para traduzir o conteúdo apresentado. Embora essa seja a abordagem preferida dos surdos, em geral, possui um custo operacional associado bastante elevado, pois é preciso que o intérprete esteja disponível em tempo integral. Além do mais, essa estratégia é inviável em alguns cenários, como, por exemplo, no sistema hipertextual que opera através da Internet - a Web, dada a natureza dinâmica dos seus conteúdos (Souza Filho; ARAÚJO, 2011).

A última estratégia utiliza recursos computacionais para traduzir textos da língua oral para a língua de sinais de forma automática (TAMBASCIA et al., 2012), (FERREIRA et al., 2011), (Souza Filho; ARAÚJO, 2011). Com essa estratégia, é possível gerar vídeos acessíveis em língua de sinais, onde os sinais normalmente são representados por um avatar 3D. Apesar da semelhança com a primeira abordagem, essa não é a primeira alternativa das pessoas surdas, pois, a sinalização feita pelos avatares não é tão natural quanto os sinais realizados por um intérprete humano (SMITH; MORRISSEY; SOMERS, 2010), (HURDICH, 2008), (LIMA, 2005).

As Subseções 3.1.1, 3.1.2, 3.1.3 e 3.1.4 apresentam alguns trabalhos presentes na literatura científica que foram selecionados de forma *ad hoc*, isto é, utilizando ape-

nas uma análise subjetiva como critério de avaliação e classificação, mas que estão diretamente relacionados aos objetivos definidos neste trabalho.

3.1.1 Interatividade dos Surdos em Sala de Aula

No contexto da educação, é preciso adequar a metodologia de ensino e os materiais didáticos para permitir o aprendizado e facilitar a interação das pessoas com deficiência auditiva dentro das salas de aula (TAMBASCIA et al., 2012). De acordo com Carvalho (2004), independente de classe, gênero ou raça, todos têm o direito de aprender em escolas comuns, e que o processo pedagógico atenda a todos, sem discriminação. Então, baseado nessa perspectiva, Tambascia et al. (2012) desenvolveu uma solução tecnológica voltada para o contexto da educação, onde a proposta é traduzir a fala do professor e gerar conteúdos acessíveis em Língua Brasileira de Sinais (Libras).

De acordo com (TAMBASCIA et al., 2012), a ideia é equipar o professor com um *headset* para que, dessa forma, seja possível captar os sinais de audio através do microfone. A partir disso, os próximos passos consistem em utilizar tecnologias de reconhecimento de voz — como o ARS (*Automatic Speech Recognition*) — para transformar a voz digital em formato de texto; utilizar um conjunto de regras de tradução para adaptar o texto à gramática de Libras; consultar um dicionário de sinais em Libras e, por fim, a medida em que os conteúdos acessíveis são gerados, distribuí-los dentro da sala de aula para os dispositivos móveis dos alunos surdos. A Figura 3.1 ilustra a arquitetura da solução proposta por Tambascia et al. (2012).

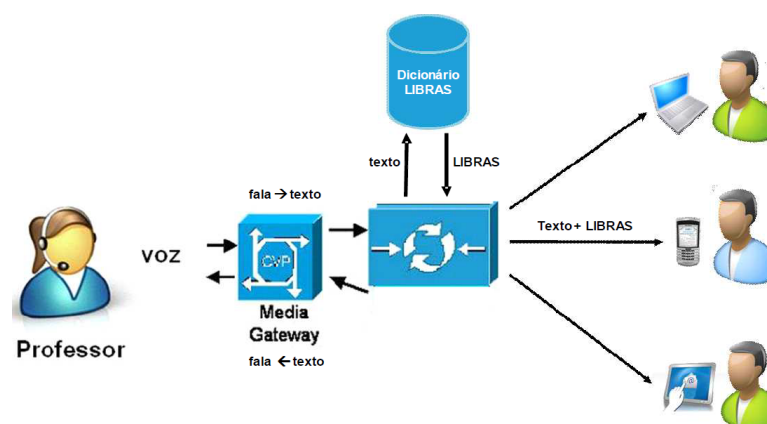


Figura 3.1: Visão arquitetural do sistema de geração de conteúdos acessíveis proposto por Tambascia et al. (2012).

Os sinais em Libras gerados por essa solução são representados por um avatar 2D, isto é, são exibidos em um espaço bidimensional de natureza plana. Dessa forma, embora a contribuição do trabalho de Tambascia et al. (2012) seja pertinente, existem algumas limitações inerentes. Uma delas é que os sinais feitos usando um avatar 2D inviabiliza a reprodução de alguns parâmetros específicos da língua de sinais, por exemplo, duas configurações de mão podem ser confundidas devido à oclusão nos dedos (BUTTUSI; CHITTARO; COPPO, 2007).

Além do problema em utilizar um avatar 2D, outras limitações encontradas no trabalho de Buttussi, Chittaro e Coppo (2007) são: necessidade das salas de aula terem tratamento acústico para reduzir o ruído ou qualquer fator que comprometa o reconhecimento da fala do professor; treinamento dos professores e alunos para que falem pausadamente, com boa dicção e bom português, e evitar falas simultâneas; soluções de tradução do Português para Libras; dentre outras.

3.1.2 *Sony Glasses*

Após a transição dos filmes produzidos em películas de 35 mm, em meados da primeira década do século XXI, o Cinema Digital encontra-se em plena expansão, com mudanças importantes do ponto de vista social e cultural. Essas mudanças trouxeram um aumento da qualidade dos conteúdos audiovisuais produzidos, inclusão de novas tecnologias, dentre outros benefícios e oportunidades.

Contudo, apesar da evolução tecnológica, os cinemas ainda não apresentam soluções técnicas viáveis para promover a acessibilidade à pessoas com deficiências sensoriais ou cognitivas. Sendo assim, com o propósito de reverter (ou amenizar) essa situação, os engenheiros da empresa Sony (2012) criaram uma solução que consiste em disponibilizar o recurso das legendas ou reproduzir conteúdos audiodescritivos em um dispositivo de segunda tela acoplado em óculos especiais.

A tecnologia utilizada nos óculos para a reprodução do conteúdo de áudio é denominada pela empresa como STW-C140GI, e a tecnologia para transmissão dos dados como STWA-C101. A possibilidade de reproduzir áudio é uma característica importante, pois permite que pessoas com problemas na visão tenham acesso à conteúdos audiodescritos, viabilizando, dessa forma, sua inclusão nos cinemas.

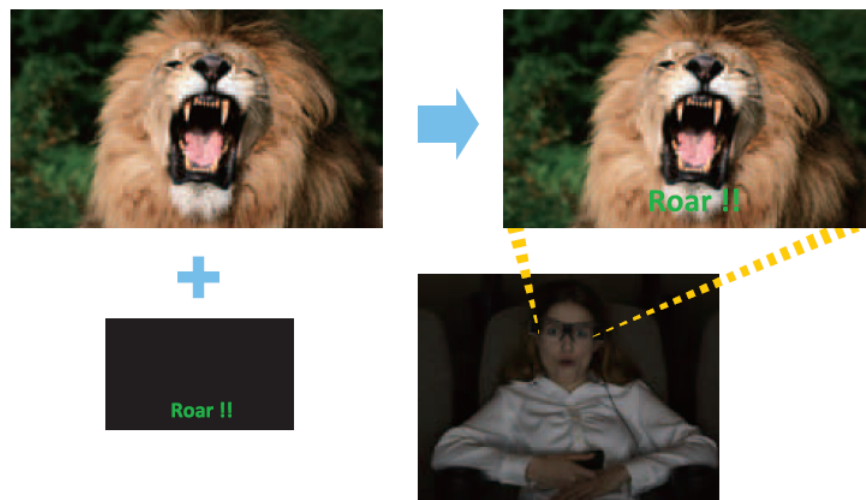


Figura 3.2: Cenário de uso da tecnologia desenvolvida pela Sony (2012).

Segundo os engenheiros da Sony, suas tecnologias exclusivas de holografia permitem que as legendas sejam projetadas de modo que pareçam estar flutuando a vários metros de distância em relação ao ponto de visão do usuário, ou seja, o usuário tem a sensação de enxergar o texto da legenda em terceira dimensão, próximo a tela do filme. Essa abordagem evita que o usuário precise desviar o foco durante a exibição dos conteúdos. A Figura 3.2 ilustra a ideia da solução desenvolvida pela (SONY, 2012).

Apesar dos recursos tecnológicos avançados, como a holografia, a solução desenvolvida pela Sony (2012) é considerada inviável para os usuários surdos, uma vez que, de acordo com (WAUTERS, 2005), esses usuários possuem muita dificuldade para ler e escrever na língua oral do seu país. Portanto, devido os usuários surdos utilizarem uma língua em que se utiliza de gestos, sinais e expressões faciais e corporais, ao invés de sons na comunicação, o ideal é utilizar conteúdos com essas características, isto é, as línguas de sinais.

3.1.3 *Wireless Captioning*

West et al. (2009), pesquisadores do *Georgia Tech Research Institute* (GTRI), desenvolveram uma solução para oferecer recursos de acessibilidade para pessoas com problemas de audição em ambientes de entretenimento como cinemas, centros esportivos,

teatros, também em ambientes institucionais de ensino, em reuniões governamentais, em museus, igrejas, dentre outros.

Basicamente, sua proposta consiste em disponibilizar para o usuário, através de um microdisplay acoplado em óculos especiais, as legendas dos conteúdos apresentados. O objetivo em usar esse tipo de dispositivo, segundo o pesquisador, é a oportunidade de criar uma espécie de Assistente de Comunicação (AC).

Em um ambiente de utilização, para viabilizar a distribuição dos conteúdos acessíveis, é necessário que esses AC estejam conectados à uma rede Wireless, devidamente configurada para operar sob o protocolo 802.11b. Com isso, é possível atender uma grande quantidade de dispositivos conectados simultaneamente sem comprometer o entendimento do conteúdo apresentado.

De acordo com West et al. (2009), embora posicionado muito próximo da região ocular do usuário, o microdisplay utiliza lentes que fazem com que a legenda pareça estar flutuando vários metros de distância em relação ao ponto de percepção. Esse recurso permite ao usuário uma experiência agradável, com uma visualização suave de ambas as imagens, isto é, da apresentação audiovisual e das legendas, além de não tirar a capacidade de percepção do mundo ao seu redor.

A arquitetura da solução proposta por West et al. (2009) é composta essencialmente por três componentes. Uma descrição breve da função de cada um deles é apresentado conforme a seguir:

- *Transmissor*: componente responsável pela distribuição das legendas, que podem ser pré-gravadas, produzidas em tempo de reprodução ou geradas utilizando técnicas de reconhecimento de voz;
- *Receptor*: dispositivos móveis (smartphones, laptops, PDA) equipados com uma interface de rede e suporte ao protocolo 802.11b para receber o conteúdo das legendas;
- *Display*: interface de exibição das legendas, que de acordo com a especificação do GTRI, pode ser a própria tela do dispositivo móvel ou um microdisplay conectado à um dispositivo.

Os recursos tecnológicos da solução apresentada por West et al. (2009) é factível, porém, para os usuários surdos, essa abordagem é considerada inadequada, pois, a forma natural de acesso à informação desses usuários é através da língua de sinais. De acordo com WAUTERS (2005), esses usuários possuem muita dificuldade para ler e escrever na língua oral do seu país. Portanto, a forma ideal de oferecer conteúdos acessíveis para os surdos é através das línguas de sinais.

3.1.4 *New Wireless Captioning System Debuts at Dallas*

A solução desenvolvida por West et al. (2009), apresentada na Seção 3.1.3, foi demonstrada em Janeiro de 2010. O evento ocorreu no *Dallas Cowboys Stadium*, localizado na cidade de Arlington, estado do Texas nos Estados Unidos.

Segundo o diretor do departamento de pesquisas do *Georgia Institute of Technology* (GTI), Kirk Englehardt, e o gerente das comunicações da *Georgia Tech Research News & Publications*, John Toon, os dispositivos denominados Assistente de Comunicação foram adaptados para funcionar de acordo com os sistemas do estádio. Dessa forma, segundo o diretor, foi possível transmitir as legendas via *broadcast* para cerca de 700 usuários, simultaneamente, conectados em uma rede Wireless 802.11b. A Figura 3.3 mostra o uso da solução de West et al. (2009) durante o evento esportivo.



Figura 3.3: Exibição das legendas durante um jogo de Futebol Americano no *Dallas Cowboys Stadium*.

O diretor do *Dallas Cowboys Stadium*, Scott Purcel, afirmou que estava bastante satisfeito em disponibilizar um serviço que, utilizando apenas um dispositivo móvel,

poderia oferecer condições de acesso à informação para os usuários com deficiência auditiva parcial ou total.

A integração entre o sistema do estádio e a solução desenvolvida por West et al. (2009) se mostrou eficiente, de acordo com o diretor do departamento de pesquisas do GTI, onde aproximadamente 700 usuários puderam receber as legendas de forma simultânea. No entanto, o problema de acessibilidade ainda é recorrente, pois, o uso de legendas em formato de texto não endereça as necessidades da maioria dos surdos, uma vez que esses usuários utilizam as línguas de sinais como principal forma de comunicação e acesso à informação em qualquer ambiente e contexto.

3.2 Acessibilidade para Deficientes Visuais

Com o advento das tecnologias emergentes, tem sido cada vez mais habitual para os seus usuários, a difusão e o compartilhamento de informações. Normalmente, essas informações são representadas de forma visual, ou seja, através de imagens. Para muitos especialistas, os diálogos e os sinais sonoros utilizados nos cinemas e nos programas de TV, por exemplo, funcionam como uma espécie de complemento, reiterando aquilo que está sendo exibido na tela. Dessa forma, a visão pode ser considerada como principal meio de acesso à informação.

Tendo em vista a dificuldade de acesso aos conteúdos visuais, é necessário, portanto, desenvolver novos métodos e técnicas para promover a acessibilidade para as pessoas com deficiência visual, de modo que possibilitem a independência social, a participação plena e efetiva na sociedade e a igualdade de oportunidades.

Na literatura, existem alguns trabalhos que foram desenvolvidos para tentar reduzir esses problemas. As Subseções 3.2.1, 3.2.2 e 3.2.3 apresentam algumas contribuições importantes, classificadas por meio de análise subjetiva, isto é, levando em consideração seu objetivo de pesquisa e sua relevância com o tema deste trabalho.

3.2.1 *Collaborative Annotation for Video Accessibility*

A quantidade de conteúdos multimídia que trafegam através da Web tem aumentado significativamente nos últimos anos. De acordo com a Cisco (2014), uma projeção feita

para medir a intensidade do tráfego na Internet entre o ano de 2013 a 2018, mostrou que os vídeos são responsáveis por cerca de 29% dos conteúdos que circulam na rede. No entanto, as soluções que possibilitam as pessoas com deficiência visual de ter acesso à esses conteúdos de vídeo ainda são modestas.

A *Web Accessibility Initiative* (WAI), fomentada pela *World Wide Web Consortium* (W3C) para prover recursos de acessibilidade na Web para pessoas com deficiências sensoriais e cognitivas, tem realizado um esforço expressivo na definição de padrões de acessibilidade para as páginas da Web, porém, ainda são limitadas as soluções acessíveis para os conteúdos de vídeo, especialmente para os deficientes visuais.

Para tentar reduzir esses problemas, Encelle, Beldame e Prié (2013) desenvolveram o projeto ACAV (*Collaborative Annotation for Video Accessibility*). Esse projeto utiliza uma abordagem que consiste em inserir janelas de interrupção - uma espécie de *pseudo* pausas - nos intervalos do vídeo onde há silêncio (isto é, inexistência de diálogos ou elementos sonoros importantes), para que, dessa forma, os conteúdos audiodescritivos possam ser introduzidos.

Para Encelle, Beldame e Prié (2013), existia uma preocupação em utilizar essa abordagem, pois, devido a inserção das interrupções, a compreensão dos deficientes visuais com relação ao vídeo poderia ser comprometida. Então, para investigar o impacto dessa técnica, os pesquisadores optaram por realizar um experimento. No experimento, foi utilizada a animação intitulada *Tong*, que possui classificação livre e tempo de duração de 8 minutos e 59 segundos.

Inicialmente, foram geradas as audiodescrições e identificados os intervalos de silêncio. Com isso, foram observados 26 blocos de audiodescrição que não se encaixavam nos respectivos intervalos encontrados. O motivo disso é que os blocos tinham um tempo de duração maior do que o intervalo disponível. Nessa situação, com o uso da técnica desenvolvida, tiveram que ser introduzidas 26 pausas no vídeo utilizado. A duração das pausas foram classificadas em três categorias: (DP1) a duração possui 0,5 vezes do intervalo de silêncio; (DP2) a duração da pausa tem o mesmo tamanho do intervalo; (DP3) o tempo de duração possui 1,5 vezes do intervalo de silêncio.

O experimento contou com a participação de 18 voluntários cegos, sendo 9 homens e 9 mulheres, com idades entre 12 e 69 anos. Os participantes foram divididos em três

Tabela 3.1: Efeitos das inserções de pausas (ENCELLE; BELDAME; PRIÉ, 2013).

Questão	Concorda plenamente	Concorda	Não concorda	Tanto faz	Discorda completamente
Pausas percebidas	9	7	0	1	1
Pausas úteis	6	5	4	1	2
Voz sintética boa	10	4	2	2	0
Pausas perturbadas	2	3	5	4	4

grupos de seis, onde cada grupo ficou responsável por avaliar uma das três técnicas de duração de pausas. Após a etapa de configuração do experimento, os participantes assistiram duas vezes o vídeo correspondente ao seu grupo e, logo em seguida, foram convidados a responder um conjunto de perguntas. A Tabela 3.1 mostra o resultado dos efeitos causados pelas inserções das pausas no vídeo.

De acordo com a Tabela 3.1, 50% dos participantes concordaram com as pausas introduzidas, enquanto apenas 11,1%, mesmo tendo certo incômodo, concordaram com a abordagem. Outro ponto importante da avaliação foi a qualidade da síntese de voz utilizada, que obteve 55,5% da aceitação dos participantes do experimento.

Apesar da contribuição do trabalho desenvolvido pelos pesquisadores Encelle, Beldame e Prié (2013), a abordagem investigada possui algumas limitações. Dentre essas limitações, talvez a mais importante seja a restrição do público, ou seja, a solução é direcionada apenas para os usuários com deficiência visual. Dessa forma, a experiência do cinema deixa de ser inclusiva para e passa a ser exclusiva.

Outro ponto que precisa ser melhor investigado na solução de (ENCELLE; BELDAME; PRIÉ, 2013) é a inserção de pausas no vídeo. Uma possível forma de investigar esse aspecto seria oferecer uma plataforma Web, por exemplo, onde os usuários pudessem avaliar a inteligibilidade dos conteúdos e o nível de satisfação com a utilização das pausas. Dessa forma, seria possível validar a solução de modo mais consistente.

3.2.2 Describing Online Videos with Text-to-Speech Narration

Para entender o conteúdo de um vídeo, as pessoas com deficiência visual necessitam de recursos acessíveis, preferencialmente a audiodescrição, para que, dessa forma, as ações e as imagens possam ser descritas, e os textos apresentados durante a exibição do vídeo possam ser lidos. No entanto, esses recursos ainda não estão disponíveis para os vídeos que trafegam na Web.

Nesse contexto, a contribuição de Kobayashi et al. (2010) consiste em criar uma plataforma online para inserir audiodescrição nos vídeos disponíveis na Internet. Sua abordagem utiliza uma tecnologia de *Text-to-Speech* (TTS) para gerar voz sintética e, *scripts* de audiodescrição em formato texto, contendo as descrições das imagens e o momento em que cada uma deve ser apresentada.

A arquitetura da plataforma proposta por Kobayashi et al. (2010) é ilustrada na Figura 3.4. Observando a composição da arquitetura, nota-se existem essencialmente três componentes: (1) editor de script (ou roteiro); (2) repositório de metadados; e (3) player de vídeo.

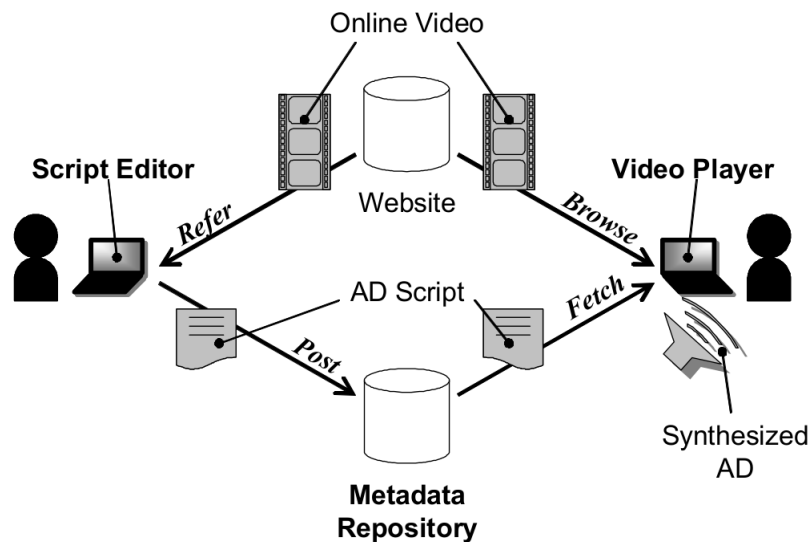


Figura 3.4: Tecnologia formada por um editor de script, um player de vídeo, um repositório de metadados, e um formato baseado em texto para trocar os scripts de Audiodescrição entre os componentes (Fonte: Kobayashi et al. (2010)).

O editor de script oferece uma interface visual para gerar ou editar as descrições de um vídeo, além de determinar o momento em que cada sentença deve ser apresen-

tada, isto é, reproduzida. Dessa forma, uma vez finalizada a edição do script com as descrições, o mesmo é então enviado para o repositório de metadados. O repositório é responsável por armazenar e disponibilizar os scripts de audiodescrição gerados. Por fim, através de uma interface Web acessível (*aiBrowser*), o player de vídeo recupera o script de audiodescrição no repositório de metadados e utiliza um sintetizador de voz para converter as descrições textuais em falas.

Para investigar a receptividade de sua proposta, Kobayashi et al. (2010) realizou uma série de experimentos com pessoas cegas e deficientes visuais. O objetivo foi, dentre outros, estudar como as diferenças na qualidade da voz podem afetar a inteligibilidade dos conteúdos e suas preferências. Nessa etapa da investigação, foram usadas audiodescrições tradicionais fornecidos por narradores profissionais.

Segundo Kobayashi et al. (2010), os resultados obtidos foram promissores. Essa conclusão foi percebida depois que a alta qualidade dos discursos por voz sintética foram semelhantes às narrações humanas, tanto no aspecto da integibilidade quanto na preferência. Os resultados também mostraram que a audiodescrição, ainda que seja de baixa qualidade, podem trazer benefícios significativos para a vida das pessoas com problemas na visão.

A contribuição dos pesquisadores Kobayashi et al. (2010) é pertinente, uma vez que conteúdos acessíveis podem ser gerados de forma colaborativa e distribuída em uma plataforma Web. No entanto, apesar de oferecer Audiodescrição para conteúdos de vídeo, a solução é voltada para um cenário específico, onde a experiência coletiva não é contemplada. Dessa forma, o contexto da solução desenvolvida por (KOBAYASHI et al., 2010) torna-se exclusivo, e não inclusivo.

3.2.3 Whatscine

Desenvolvido na *Universidad Carlos III de Madrid*, o *Whatscine* tem o propósito de oferecer recursos de acessibilidade em salas de cinema. A ideia é tornar as salas de cinema acessíveis, de forma que as pessoas com deficiência visual possam ouvir a audiodescrição do filme sem que haja interferências no áudio principal do filme ou incômodos para o resto do público, e que os deficientes auditivos tenham acesso às legendas ou à tradução em língua de sinais através de óculos especiais ou *smartphones*. A Figura 3.5

apresenta os principais componentes que integram a arquitetura do Whatscine.



Figura 3.5: Arquitetura do Whatscine — fonte: (WHATSCINE, 2015).

De acordo com a Figura 3.5, o *Digital Cinema Server* (ou Servidor de Cinema Digital) é o responsável por enviar as mídias do filme (isto é, áudio e vídeo) para o projetor instalado na sala de projeção, como também transmitir os recursos de acessibilidade (audiodescrição, legendas ou tradução em língua de sinais) para uma rede WiFi local. Além disso, o *Digital Cinema Server* também utiliza o sistema do Whatscine para gerar os conteúdos acessíveis.

No ambiente de exibição — o *Auditorium*, de acordo com a Figura 3.5, os usuários com deficiência podem ter acesso aos conteúdos acessíveis através de *smartphones*, tablets, óculos especiais equipados com microdisplays, dentre outros. Então, dessa forma, esses dispositivos são usados para receber os conteúdos da legenda em formato de texto ou os conteúdos com língua de sinais, no caso dos usuários surdos, e os conteúdos de Audiodescrição, no caso dos usuários cegos.

A proposta do Whatscine (2015) é uma contribuição que desperta interesse tanto no mercado da produção de conteúdos cinematográficos quanto nos deficientes auditivos e nos deficientes visuais. No entanto, fazendo um comparativo qualitativo entre a

proposta do Whatscine e a solução desenvolvida neste trabalho, é possível identificar alguns aspectos importantes que valorizam e diferenciam ambas as propostas.

Embora no contexto do cinema digital os conteúdos acessíveis não precisem ser gerados em tempo real, uma vez que o filme encontra-se pronto, essa é uma característica da solução desenvolvida neste trabalho e que não se encontra presente no Whatscine (2015). Uma das vantagens do suporte à geração de conteúdos acessíveis é o ganho de tempo com a produção desses conteúdos, caso não estejam disponíveis. Além disso, outra vantagem importante é o baixo custo de produção. Para a solução proposta neste trabalho, são necessários apenas os arquivos de legenda e os arquivos de audiodescrição, enquanto que feito com profissionais intérpretes de língua de sinais ou com audiodescritores, existe um custo de produção mais elevado, pois, engloba o cenário, câmeras, iluminação, edição, técnicos de filmagem, dentre outros.

Outro fator que limita a realização de um comparativo mais amplo entre o Whatscine (2015) e a solução proposta é a ausência de produções científicas que descrevam de forma detalhada a estratégia utilizada pelo Whatscine no processo de distribuição e sincronização dos conteúdos acessíveis dentro das salas de cinema. Além disso, também não foram encontrados na literatura científica resultados de testes com usuários que viessem a validar a proposta e, do mesmo modo, resultados de testes computacionais para varificar sua viabilidade.

3.3 Considerações

Foram apresentadas neste Capítulo as principais contribuições científicas relacionadas com o tema de pesquisa abordado neste trabalho (TAMBASCIA et al., 2012), (KATO et al., 2010), (BUTTUSSI; CHITTARO; COPPO, 2007), (SONY, 2012), (WEST et al., 2009), (ENCELLE; BELDAME; PRIÉ, 2013), (KOBAYASHI et al., 2010), (WHATSCINE, 2015). Essas contribuições foram selecionadas na literatura após a realização de buscas feitas em bibliotecas digitais, como *IEEE Xplore*¹, *ACM Digital Library*², *ScienceDirect*³,

¹ieeexplore.ieee.org

²dl.acm.org

³www.sciencedirect.com

*Scopus*⁴ e, além disso, também foram feitas pesquisas aleatórias no *Google*⁵.

Embora os trabalhos investigados e discutidos neste Capítulo utilizem recursos tecnológicos sofisticados, estes apresentam algumas limitações que podem comprometer a inteligibilidade dos conteúdos e consequentemente sua aceitabilidade por parte dos usuários com deficiência auditiva ou deficiência visual.

Dentre as principais limitações encontradas nas propostas direcionadas para as pessoas surdas, as que utilizam as legendas em formato de texto como recurso de acessibilidade são as que mais chamam atenção. O motivo pelo qual essa abordagem é considerada ineficiente é pelo simples fato de que os surdos se comunicam naturalmente através das línguas de sinais, portanto, uma língua baseada em sons, como o português do Brasil, por exemplo, funciona como uma segunda língua para essas pessoas. De acordo com (WAUTERS, 2005), as legendas em formato de texto não endereçam as necessidades da maioria das pessoas surdas, uma vez que esses usuários possuem muita dificuldade para ler e escrever na língua oral do seu país.

Com relação aos trabalhos voltados para resolver os problemas de acesso à informação dos deficientes visuais apresentados neste Capítulo, as contribuições são pertinentes, porém, levando em consideração o escopo desta pesquisa, que direciona o desenvolvimento de uma solução capaz de oferecer conteúdos acessíveis em dispositivos de segunda tela em salas de cinema, as propostas apresentadas possuem algumas limitações. Por exemplo, a inserção de pausas no vídeo para descrever informações excedentes ao tempo disponível pode ser uma estratégia que venha causar bastante desconforto para o usuário cego. Além do mais, com essa abordagem, as salas de cinema deixam de ser inclusivas e passam a ser exclusivas.

Outra limitação encontrada nos trabalhos é a geração de automática dos conteúdos acessíveis, ou seja, das trilhas de vídeo com língua de sinais e da audiodescrição. Uma das vantagens de oferecer suporte à geração desses conteúdos é o baixo custo e a velocidade do tempo de produção. Portanto, como a solução proposta neste trabalho apresenta essas características, isto a torna uma contribuição relevante para a literatura e para os usuários surdos e para os usuários cegos.

⁴www.scopus.com

⁵www.google.com

Tabela 3.2: Classificação das principais características dos trabalhos relacionados.

Ref.	Características				
	Tipo de conteúdo para Surdos	Tipo de conteúdo para Cegos	Geração Automática de conteúdos	Domínio	Licença
Tambascia et al. (2012)	Legendas e Língua de Sinais	-	Sim	Sala de Aula	Acadêmico
Sony (2012)	Legendas	Audiodescrição	Não	Cinema Digital	Comercial
West et al. (2009)	Legendas	Audiodescrição	Não	Geral	Comercial
Encelle, Beldame e Prié (2013)	-	Audiodescrição	Sim	Geral	Acadêmico
Kobayashi et al. (2010)	-	Audiodescrição	Não	Web	Comercial
Whatscine (2015)	Legendas e Língua de Sinais	Audiodescrição	Não	Cinema Digital	Comercial
<i>Solução Proposta</i>	Língua de Sinais	Audiodescrição	Sim	Cinema Digital	Acadêmico

Com o objetivo de fazer um comparativo entre os trabalhos relacionados discutidos neste Capítulo e a solução proposta, a Tabela 3.2 apresenta a classificação das principais características de cada um desses trabalhos, como tipo de conteúdo gerado, suporte à geração automática, domínio da solução e tipo de licença.

De acordo com a Tabela 3.2, observando a coluna referente ao domínio para o qual a solução foi desenvolvida, é possível perceber que foram investigados trabalhos de domínio geral e de domínio específico. No caso dos trabalhos de domínio específico, as contribuições são voltadas para o contexto da educação em salas de aula, para conteúdos disponíveis na Web, e para as salas de cinema digital.

Dentre os trabalhos apresentados na Tabela 3.2, aqueles que oferecem conteúdos em línguas de sinais para os surdos e conteúdos de audiodescrição para os cegos, ao mesmo tempo, são considerados mais relevantes para a linha de pesquisa deste trabalho. Além disso, outro ponto importante que pode ser observado na classificação dos trabalhos é com relação ao suporte à geração automática dos conteúdos e o tipo de licença, pois, apenas os trabalhos com licença acadêmica oferecem esse tipo de recurso.

Por fim, com base nas limitações dos trabalhos relacionados apresentados e discutidos ao longo deste Capítulo, e somando os problemas de comunicação e acesso à informação das pessoas com deficiência auditiva ou com deficiência visual, o Capítulo 4 apresenta em detalhes o desenvolvimento da solução proposta neste trabalho.

Capítulo 4

Solução Proposta

Conforme mencionado no Capítulo 1, o objetivo deste trabalho consiste em desenvolver uma solução tecnológica prática e eficiente, que seja capaz de gerar automaticamente e distribuir, conteúdos acessíveis para usuários surdos e usuários cegos em salas de cinema digital. A ideia é oferecer trilhas alternativas de vídeo com língua de sinais em Libras e Audiodescrição, de modo que os usuários deficientes possam receber esses conteúdos através de dispositivos móveis de segunda tela.

A geração desses conteúdos acessíveis pode ser feita de duas formas: em tempo real ou sob demanda. Para a geração dos conteúdos em tempo real, são enviadas mensagens — especificadas em um protocolo de comunicação — ao player de cinema para obter os arquivos de legenda, roteiro de audiodescrição e informações de controle. A geração em tempo real é uma das características que diferencia esta solução da proposta do Whatscine (2015), principal trabalho correlato investigado. Para a geração dos conteúdos sob demanda, os arquivos de legenda e do roteiro de audiodescrição devem estar disponíveis previamente. Então, após gerar os conteúdos, o sistema proposto utiliza o player de cinema apenas para controlar sua execução.

Os detalhes do desenvolvimento da solução proposta são apresentados neste Capítulo, que está organizado da seguinte forma: a Seção 4.1 apresenta o *Sistema de Acessibilidade*, onde é apresentada a arquitetura geral da solução proposta; a Seção 4.2 mostra a especificação do protocolo de comunicação entre a solução proposta e o player de cinema; e, finalmente, a Seção 4.3 apresenta as considerações finais, onde é feita uma síntese dos principais pontos abordados e discutidos neste Capítulo.

4.1 Sistema de Acessibilidade

O Sistema de Acessibilidade constitui a solução proposta neste trabalho. O objetivo desse sistema é gerar automaticamente trilhas alternativas de vídeo com janelas de Libras e Audiodescrição para usuários surdos e para usuários cegos, respectivamente, e distribuir esses conteúdos em salas de cinema. A organização dos componentes que constituem a arquitetura desse sistema é apresentada na Figura 4.1.

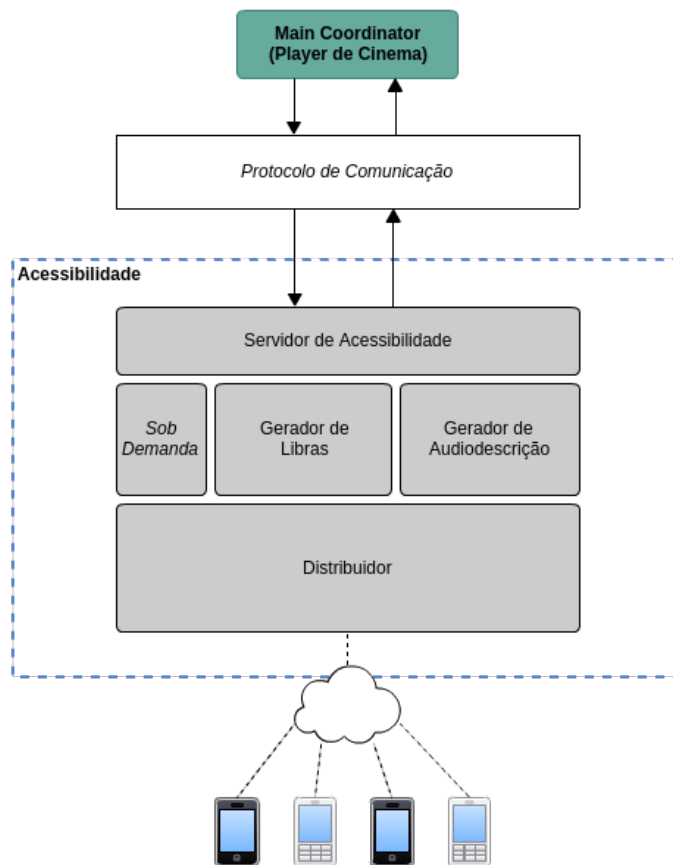


Figura 4.1: Arquitetura do Sistema de Acessibilidade.

Conforme observado na Figura 4.1, o Sistema de Acessibilidade é formado pelos módulos *Servidor de Acessibilidade*, *Gerador de Libras*, *Gerador de Audiodescrição*, *Sob Demanda* e *Distribuidor*. O *Servidor de Acessibilidade*, apresentado em detalhes na Subseção 4.1.1, é o módulo do sistema responsável por enviar e receber requisições provindas do player de cinema — componente ilustrado na Figura 4.1 como *Main Coordinator* —, cuja especificação das mensagens é definida pelo Protocolo de Comunicação, discutido detalhadamente na Seção 4.2.

O processo de geração das trilhas de vídeo com tradução em língua de sinais é feita pelo módulo *Gerador de Libras*. O *Gerador de Libras* utiliza os arquivos de legenda fornecidos pelo player de cinema para gerar os vídeos acessíveis e, através das informações de tempo de cada legenda, para determinar o momento em que os sinais em Libras devem ser exibidos, ou seja, estabelecer o sincronismo entre os sinais em Libras e os diálogos do vídeo principal. Os detalhes do desenvolvimento do *Gerador de Libras* é apresentado na Subseção 4.1.2.

O conteúdo de Audiodescrição desta solução é gerado pelo módulo *Gerador de Audiodescrição*. Para a geração desses conteúdos, o *Gerador de Audiodescrição* necessita dos arquivos de descrição, que também são fornecidos pelo player de cinema. Nesse componente, são utilizadas técnicas de *Text-to-Speech* para transformar texto em voz, necessárias para descrever as imagens exibidas no vídeo principal. O módulo *Gerador de Audiodescrição* é apresentado detalhadamente na Subseção 4.1.3.

Além da geração automática dos conteúdos, outro cenário contemplado nesta solução é a possibilidade de receber os conteúdos prontos, ou seja, caso tenham sido produzidos por outra fonte, ou mesmo por esta solução em outro momento, é possível pular a etapa de geração e avançar direto para a etapa de distribuição. Essa alternativa é ilustrada na Figura 4.1 pelo módulo *Sob Demanda*.

Por fim, o módulo *Distribuidor* é responsável pela transmissão das mídias de Libras e de Audiodescrição nas salas de cinema. Além disso, o *Distribuidor* também é responsável pela implementação de algoritmos de sincronização de relógios entre os dispositivos móveis dos usuários e o servidor de conteúdos acessíveis. Esse é um aspecto crucial desta solução, pois, caso os conteúdos não estejam sincronizados dentro de um limiar aceitável, a inteligibilidade dos conteúdos pode ser comprometida. O *Distribuidor* é apresentado em detalhes na Subseção 4.1.4.

4.1.1 Servidor de Acessibilidade

No processo de geração dos vídeos de Libras e da Audiodescrição, alguns conteúdos são imprescindíveis. Por exemplo, os arquivos com as legendas em formato de texto são necessários para gerar os vídeos de Libras, assim como o roteiro de produção do filme — ou arquivos com marcações temporais que contenham todas as descrições das cenas —

é usado para gerar os conteúdos da Audiodescrição. Portanto, é fundamental que esses arquivos estejam disponíveis para que o sistema proposto possa gerar adequadamente os conteúdos acessíveis, caso ainda não tenham sido gerados.

Normalmente, no contexto dos cinemas digitais, esses arquivos estão disponíveis apenas para o sistema responsável pela reprodução dos conteúdos, ou seja, para o player de cinema. Então, em virtude da utilização do Fogo Player (ver Seção 2.4) no desenvolvimento desta solução, e pela viabilidade de integração entre os dois sistemas, foi desenvolvido um mecanismo para auxiliar na comunicação e na recuperação dos arquivos e das informações de controle, isto é, foi concebido o Protocolo de Comunicação, discutido em detalhes na Seção 4.2.

A comunicação entre o sistema da solução proposta e o player de cinema é feita por meio da permutação de mensagens, cuja especificação de cada mensagem encontra-se definida no Protocolo de Comunicação. Dessa forma, seguindo essa especificação, o *Servidor de Acessibilidade* — ilustrado na Figura 4.1 — envia mensagens ao Fogo Player para obter os arquivos de legenda e do roteiro do filme, assim como outras informações importantes para controlar a execução do sistema.

Nesse cenário de comunicação, os papéis são atribuídos a cada sistema conforme o modelo *cliente-servidor*, onde o player de cinema assume o papel de provedor dos conteúdos e serviços, e a solução proposta, nesse caso, assume o papel de requerente desses conteúdos. Do mesmo modo, outros sistemas que necessitem obter arquivos ou informações do Fogo Player devem assumir o papel do cliente.

4.1.2 Arquitetura do Sistema de Libras

A solução apresentada neste trabalho é composta por um conjunto de componentes de software. No módulo *Gerador de Libras*, conforme ilustrado na Figura 4.1, esses componentes são responsáveis por gerar automaticamente, trilhas de vídeo com janelas de Libras. A Figura 4.2 apresenta a arquitetura deste módulo da solução.

De acordo com a Figura 4.2, este módulo é formado por cinco componentes principais: *Controlador de Libras*, *Extrator de Legendas*, *Tradutor*, *Sincronizador* e *Animação*. Além desses componentes, existe também o Dicionário de Libras. Esse dicionário consiste em um repositório de vídeos no formato MPEG2-TS. Cada um desses vídeos

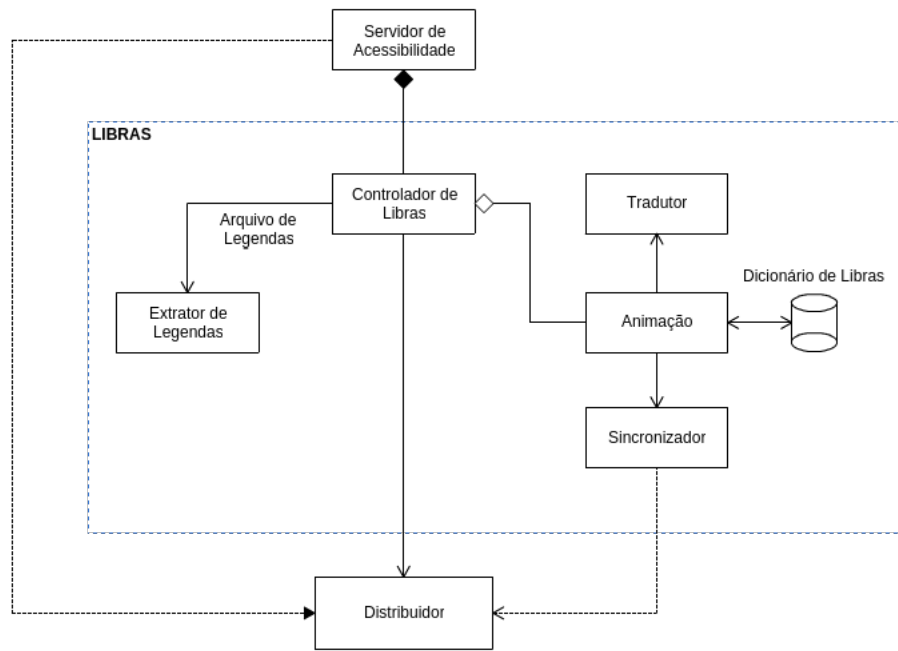


Figura 4.2: Visão arquitetural do sistema de geração de Libras.

corresponde a um sinal em Libras, representados por um agente virtual 3D.

A coordenação do processo de geração e distribuição dos vídeos de Libras é feita pelo componente de controle, isto é, o *Controlador de Libras*. O Controlador é um dos componentes principais dessa arquitetura, pois, sua função consiste em articular a execução dos outros componentes.

Inicialmente, no processo de geração dos conteúdos acessíveis de Libras, o componente *Controlador de Libras* recorre ao *Servidor de Acessibilidade* para que o mesmo inicie uma conexão com o Fogo Player. Então, após o estabelecimento da comunicação, os arquivos de legenda e do roteiro podem ser requisitados, assim também como as informações de controle podem ser recebidas pelo módulo *Gerador de Libras*.

Quando os arquivos de legenda chegam até o componente *Controlador de Libras*, esses são repassados para o componente responsável por extrair as informações importantes, o *Extrator de Legendas*. O Extrator de Legendas tem a função de consumir o texto e as marcações de tempo de cada legenda. Essas informações de tempo são representadas por um tempo em que a legenda deve ser apresentada, denotado como *TimeIn*, e o tempo em que deve ser ocultada, *TimeOut*. Esses arquivos de legenda são especificados pela *Digital Cinema System Specification* (DCSS).

Dessa forma, baseado no relógio lógico, o *Controlador de Libras* monitora o instante em que as legendas devem ser apresentadas. Dessa forma, quando esse tempo é atingido, o Controlador encaminha os dados da legenda para o Sincronizador. O Sincronizador é o componente responsável por receber e converter uma sequência de glosas em uma trilha de vídeo contínua, sincronizada com o vídeo principal.

Para converter o texto que está no domínio da língua oral (neste caso, o português do Brasil) para uma representação intermediária em Libras (isto é, a Glosa), o Sincronizador encaminha o texto da legenda para o Tradutor. O método de tradução automática utilizado no Tradutor foi desenvolvido por Araújo (2012).

O processo de tradução automática combina métodos de compressão estatística para classificar as palavras, com estratégias de simplificação textual para reduzir a complexidade do texto de entrada, e um conjunto de regras morfológicas e sintáticas Araújo (2012). Maiores detalhes sobre o processo de tradução automática podem ser encontrados em Araújo (2012).

Após a tradução de cada legenda, o Tradutor repassa a glosa para o componente de Animação para que os sinais possam ser recuperados. Dessa forma, durante a animação dos sinais, o Dicionário de Libras é consultado para associar cada *token* da glosa ao seu vídeo correspondente. Caso o vídeo do sinal não exista no dicionário, este deverá ser soletrado. O processo de soletração consiste em sinalizar, individualmente, cada letra de uma glosa. Por exemplo, em Libras, a palavra “CINELIBRAS” não possui um sinal, logo, esta irá se tornar “C-I-N-E-L-I-B-R-A-S”. Essa é a estratégia utilizada pelos surdos e intérpretes para descrever termos que não possuem sinais, como, por exemplo, nomes próprios e termos técnicos de modo geral.

Durante a reprodução de um filme, nos intervalos onde não existem diálogos, ou seja, que não há legendas para serem exibidas, o *Sincronizador* adiciona um vídeo com o avatar em pose neutra. A configuração da pose neutra significa que o avatar ficará posicionado com os braços estendidos para baixo e sem qualquer expressão facial. Essa configuração foi definida como a mais recomendada pela comunidade dos surdos juntamente com as sugestões dos intérpretes de Libras.

Após gerar um vídeo para cada sinal em Libras, o *Sincronizador* agrupa esses segmentos formando uma trilha de vídeo contínua. Um detalhe importante desse processo

é que os sinais de Libras que são apresentados no vídeo final devem estar sincronizados com cada legenda correspondente. Esse sincronismo possui uma tolerância máxima de 4 segundos de atraso, de acordo com a norma técnica 15209 da ABNT (2005). Dessa forma, para manter o sincronismo, o *Sincronizador* utiliza as informações de tempo disponíveis nos arquivos de legenda. Então, ao final desse processo, o vídeo acessível com janelas de Libras é encaminhado para o *Distribuidor*.

Por fim, o *Distribuidor* é o componente responsável pela transmissão do vídeo de Libras. Essa transmissão é feita por meio de uma rede local sem fio, previamente configurada dentro da sala de cinema. Dessa forma, os usuários podem se conectar no servidor de aplicação e receber os conteúdos acessíveis através dos dispositivos móveis de segunda tela, como *smartphones*, tablets, dentre outros. Os detalhes de implementação do componente *Distribuidor* são apresentados na Subseção 4.1.4.

4.1.3 Arquitetura do Sistema de Audiodescrição

De acordo com a Figura 4.1, o sistema de acessibilidade também oferece suporte à geração de conteúdos acessíveis para pessoas com deficiência visual, isto é, Audiodescrição. O módulo responsável pela produção desses conteúdos é composto por um conjunto de componentes utilitários de software. A Figura 4.3 apresenta a organização dos componentes que compõem a arquitetura deste módulo.

De acordo com a Figura 4.3, o módulo de audiodescrição é formado por quatro componentes principais: *Controlador de Audiodescrição*, *Leitor de Descrições*, *Sintetizador de Voz (TTS)* e *Compositor de Mídia*. A conversão de texto em voz (técnica denominada *Text-to-Speech*) significa que os sistemas devem interpretar representações linguísticas através de símbolos como transcrições fonéticas, para que, dessa forma, a voz sintética possa ser gerada. Para isso, foi usado um *Modelo Acústico de Voz* (ver Figura 4.3).

Para a geração da audiodescrição, o sistema proposto necessita de um arquivo no formato SRT (*SubRip Text*), cujo conteúdo são as descrições — em formato texto — do filme e o instante de tempo (*Presentation Time Stamp - PTS*) em que cada uma deve ser apresentada, ou seja, reproduzida. Uma vez com o arquivo de descrições, o componente *Controlador de Audiodescrição* encaminha esse arquivo para o *Leitor de*

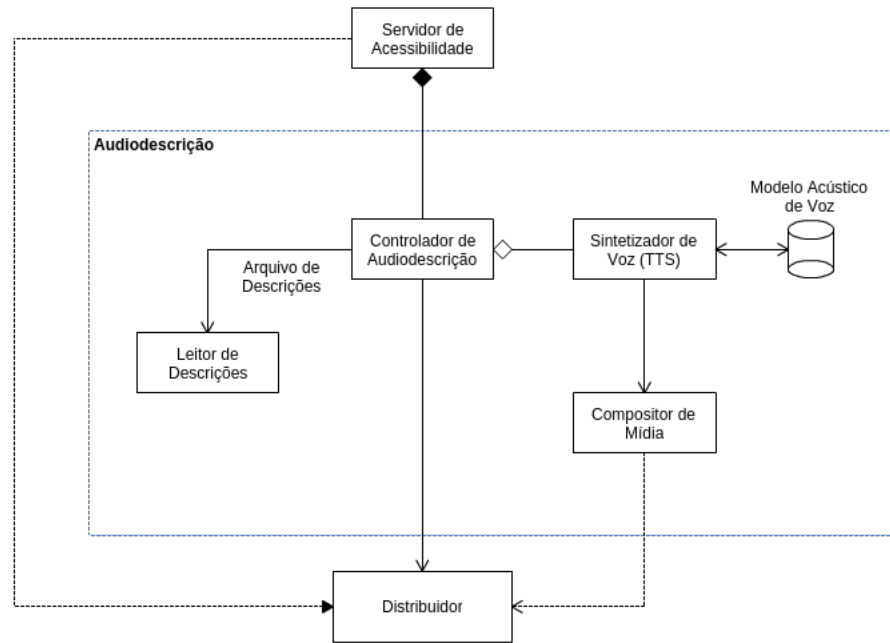


Figura 4.3: Arquitetura do sistema de geração de Audiodescrição.

Descrições, que irá extrair as informações necessárias.

No estudo apresentado neste trabalho, o arquivo de descrições utilizado na geração da Audiodescrição foi fornecido pelo sistema CineAD (CAMPOS; ARAÚJO; FILHO, 2014). De acordo com Campos, Araújo e Filho (2014), para a geração desses arquivos, o CineAD precisa analisar as legendas e o roteiro original do filme, pois, dessa forma, é possível extrair as informações mais importantes do roteiro, como títulos de cena, personagens, ações, dentre outros elementos, além de identificar os intervalos onde não há diálogos (os *gaps*) com base no tempo de cada legenda e, finalmente, gerar um arquivo contendo apenas o roteiro de Audiodescrição. O CineAD, assim como a solução apresentada neste trabalho, encontram-se em fase de desenvolvimento pelo mesmo grupo de pesquisa, o Núcleo de Pesquisa e Extensão/LAViD¹.

Além da utilização do CineAD, é possível que as descrições do filme já tenham sido produzidas, o que é feito normalmente por um profissional audiodescritor. Nesse caso, as marcações de tempo não são monitoradas pela sistema proposto, pois, presume-se que o audiodescritor tenha levado em consideração os *gaps* entre os diálogos para que não haja sobreposições entre as vozes dos personagens e a do audiodescritor, o que

¹<http://www.lavid.ufpb.br>

poderia gerar certo desconforto para os usuários cegos.

Durante o processo de geração das mídias de Audiodescrição, o controlador é notificado com cada descrição textual que precisa ser processada para então gerar o conteúdo de áudio correspondente. Dessa forma, ao receber essas notificações, o controlador repassa esses metadados para o componente *Sintetizador de Voz*, que por sua vez utiliza um modelo acústico de voz humana para transformar texto em voz. Essa técnica é normalmente conhecida como *Text-to-Speech* (TTS). Na implementação do componente de síntese de voz apresentado neste trabalho, foi utilizada a biblioteca de software *Ivona*².

Após a geração do conteúdo de Audiodescrição, uma trilha de áudio no formato MP3 (MPEG-1/2 Audio Layer 3) deve ser fornecida por este módulo do sistema. Então, esse conteúdo é encaminhado para o componente responsável pela transmissão, nesse caso, o *Distribuidor*. Os detalhes do desenvolvimento do componente Distribuidor são apresentados e discutidos na Subseção 4.1.4.

4.1.4 Distribuidor

O Distribuidor é o componente responsável por fornecer os conteúdos de vídeo com tradução em Libras e de Audiodescrição para os usuários surdos e cegos, respectivamente, dentro das salas de cinema. A Figura 4.4 ilustra a organização deste componente.

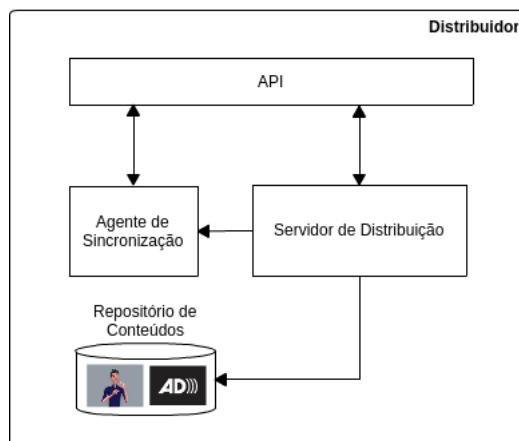


Figura 4.4: Organização do componente de distribuição.

De acordo com a Figura 4.4, o Distribuidor é composto por uma API (*Application Programming Interface*), um *Servidor de Distribuição*, um *Agente de Sincronização* e

²<http://www.ivona.com/en>

um *Repositório de Conteúdos*. A API é uma documentação cujo propósito consiste em especificar os tipos de serviços oferecidos e quais estão disponíveis para serem consumidos — ou requisitados — pelos clientes registrados na aplicação. De modo simplificado, os serviços disponíveis nesta versão da API, são:

- `/api/entry/:user`: o cliente faz uma requisição ao servidor solicitando seu registro para que possa receber os conteúdos acessíveis a serem transmitidos;
- `/api/start`: o servidor utiliza essa especificação para notificar os clientes sobre o momento em que o filme é iniciado;
- `/api/stop`: da mesma forma como acontece no início do filme, o servidor usa essa especificação para notificar os clientes sobre o fim do filme;
- `/api/time`: durante a reprodução dos conteúdos, os clientes podem, em qualquer momento, consultar o tempo atual de reprodução do filme;
- `/api/time/sync`: o servidor usa essa especificação para enviar informações de tempo do filme para que os clientes possam manter o sincronismo.

Para que os clientes possam se registrar na aplicação é preciso que especifiquem o tipo de conteúdo que desejam receber. Isso é feito enviando juntamente com a requisição, o tipo de conteúdo como parâmetro para a API. Por exemplo, para um surdo receber os conteúdos fornecidos pelo servidor, o mesmo deve enviar o identificador “LIBRAS” como parâmetro. Dessa forma, o servidor associa o endereço do cliente ao tipo de conteúdo que será enviado. O mesmo acontece para os usuários cegos, que devem enviar o identificador “AD” para receber o conteúdo da audiodescrição.

De acordo com as especificações da API, todas as requisições devem ser precedidas do endereço de rede do servidor, assim como o protocolo de comunicação entre as aplicações. Por exemplo, no cenário em que um cliente surdo faça uma requisição para se registrar na aplicação, e que o servidor encontre-se registrado no endereço 192.168.0.100, porta 8000, o cliente deve enviar a requisição da seguinte forma: `http://192.168.0.100:8000/api/entry/libras`.

O protocolo de comunicação utilizado na distribuição desses conteúdos multimídia é o *Hypertext Transfer Protocol* (HTTP). O HTTP é um protocolo de nível de aplicação para sistemas de informação hipermídia distribuídos e colaborativos (FIELDING, 1999). O objetivo de usar o HTTP para fazer streaming dos vídeos de Libras e da Audiodescrição é que há a possibilidade de quebrar o fluxo em pequenos segmentos de dados. Então, dessa forma, os clientes podem gerenciar seu *buffer* e requisitar em qualquer momento novos segmentos de dados conforme necessidade.

A API especifica os tipos de serviços oferecidos, porém, é o Servidor de Distribuição (ver Figura 4.4) que processa as requisições. A função do Servidor de Distribuição é implementar a lógica do servidor, ou seja, dentre outras funções, é responsável por manter a lista dos clientes conectados na aplicação, pelo envio dos conteúdos certos para cada tipo de cliente, e pelo controle do envio de mensagens de sincronismo.

Para implementar o Servidor de Distribuição, foi utilizada a tecnologia conhecida como *Node.js*. Node.js é uma plataforma que permite construir aplicações de rede escaláveis e de alto desempenho. A plataforma do Node.js utiliza um modelo orientado a operações de E/S não bloqueantes, que o torna leve e eficiente, adequado para aplicações distribuídas de tráfego intenso e de tempo real (NODEJS, 2015).

Outro componente observado na Figura 4.4 é o Agente de Sincronização. Esse componente é responsável por enviar mensagens periodicamente para os clientes registrados na aplicação. O objetivo dessas mensagens é informar aos clientes o tempo atual de reprodução do filme. Dessa forma, é possível que os clientes observem seu tempo de reprodução local e verifiquem se há algum atraso ou adiantamento com relação ao tempo recebido do servidor. A Figura 4.5 apresenta um diagrama que ilustra o envio de mensagens de sincronismo e o método de ajuste de relógio.

Conforme observado na Figura 4.5, inicialmente o servidor envia uma mensagem (*Start*) notificando o início da sessão. O tempo que essa mensagem leva para chegar até o cliente é apresentado como “Caminho crítico”. Nesse caso, o caminho crítico significa um intervalo de tempo desconhecido, que depende do adaptador de rede e da velocidade de transmissão. Porém, para ilustrar o funcionamento do método de ajuste de relógio, foi considerada uma unidade de tempo.

Ao receber a mensagem notificando o início do sessão, o cliente então carrega uma

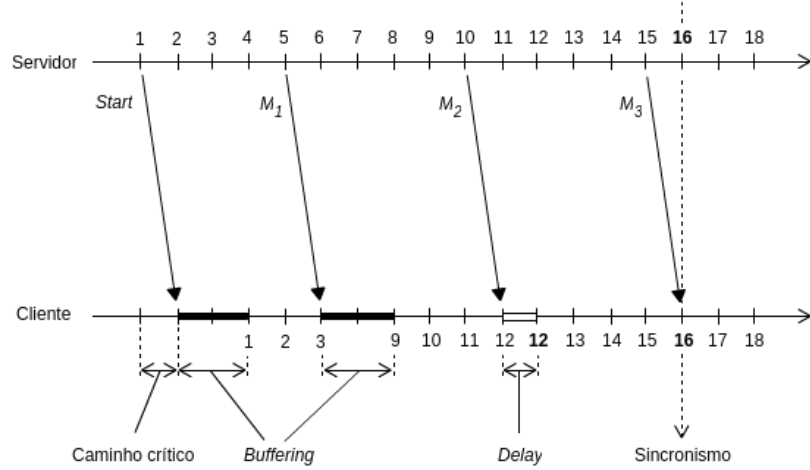


Figura 4.5: Diagrama de sincronização entre cliente e servidor.

quantidade de dados suficientes para reproduzir o conteúdo, ou seja, realiza o *Buffering*. Em seguida, após obter esses dados e estar pronto para reproduzir, observando o diagrama na Figura 4.5, nota-se que os relógios estão diferentes, isto é, não estão sincronizados. Então, para ajustar a reprodução dos conteúdos, o cliente deve efetuar o cálculo de deslocamento dado através da função:

$$\theta = S' + 2 \cdot (S' - C'),$$

onde S' representa o instante de tempo do servidor, e C' o tempo do cliente.

De acordo com o diagrama apresentado, as mensagens de sincronismo são denotadas pelo conjunto $T = \{M_1, M_2, M_3, M_n\}$. Cada mensagem é responsável por carregar o tempo lógico atual do servidor. Dessa forma, como exemplo do cálculo de deslocamento, podemos considerar inicialmente o envio da mensagem M_1 , que parte do servidor para o cliente. Quando a mensagem M_1 é enviada, o servidor está tempo igual a 5. Ao chegar no cliente — já considerando o caminho crítico, seu tempo lógico é igual a 3. Essa diferença pode variar bastante de dispositivo para dispositivo, pois depende de recursos de processamento, memória, hardware, dentre outros.

Assim que as informações de tempo são obtidas, o cliente então utiliza a função apresentada anteriormente para calcular o deslocamento. Ao substituir os valores na função, será obtido o valor 9. Esse valor indica que o cliente deve requisitar ao servidor os quadros do conteúdo referentes ao tempo a partir do instante 9. Dessa forma, após

carregar quadros suficientes, o cliente retoma a reprodução a partir do instante 9.

Após essa operação, é possível observar que o cliente agora está adiantado com relação ao tempo do servidor. Essa diferença é percebida quando o servidor envia a mensagem seguinte, M_2 . Ao receber a mensagem M_2 , o cliente nota que está no tempo 12, enquanto o servidor em 10. Nesse caso, o cálculo de ajuste é feito de maneira diferente, pois, como o custo de requisitar novos dados ao servidor é uma tarefa custosa, torna-se mais viável inserir atrasos até conseguir o sincronismo. O cálculo de ajuste de tempo nesse caso é dado através da função a seguir:

$$\sigma = 0.5 \cdot (|S'' - C''|),$$

onde S'' indica o tempo atual do servidor, e C'' o tempo do cliente.

Substituindo os valores na função, será encontrado o valor 1. Isso significa que o cliente deve introduzir um atraso de uma unidade. Esse procedimento pode ser visto na Figura 4.5 pelo indicador *Delay*. É possível que haja a necessidade de inserir novos atrasos em mensagens subsequentes, resultantes do ajuste contínuo até chegar o sincronismo. No entanto, é possível observar que o sincronismo foi alcançado após receber a mensagem M_3 . De maneira empírica, baseado apenas em testes analíticos, foi definido um limiar de diferença máxima de 300 ms (trezentos milissegundos), que para um vídeo codificado em 30 fps³, a diferença não pode ser superior a 9 quadros.

O último componente observado na Figura 4.4 é o *Repositório de Conteúdos*. O Repositório de Conteúdos, como o próprio nome sugere, é um repositório responsável por armazenar as mídias acessíveis (isto é, vídeos de Libras e Audiodescrição) geradas pela solução apresentada neste trabalho. De acordo com a Figura 4.1, esses conteúdos são podem ser gerados pelos módulos “LIBRAS” e “Audiodescrição”, e só depois de gerados são encaminhados para o componente de distribuição.

Antes da etapa de geração dos conteúdos acessíveis, é necessário obter algumas informações sobre o filme a ser exibido. Dentre essas informações, estão os arquivos de legenda e de audiodescrição, além do relógio e eventos de início e fim. Dessa forma, para obter essas informações, foi desenvolvido um protocolo de comunicação com o

³*Frames per Second* — Quadros por Segundo.

Fogo Player (Seção 2.4). Esse protocolo é discutido em detalhes na próxima Seção.

4.2 Protocolo de Comunicação

Segundo Araújo, Pedroza e Mesquita (2003), um Protocolo de Comunicação deve ser definido a partir de um conjunto de regras que governam o formato e o significado de quadros, mensagens ou pacotes que são trocados através de entidades pares. Baseado nisso, a especificação do Protocolo de Comunicação entre a solução proposta e o Fogo Player foi elaborada com base nas propriedades de comunicação definidas por Araújo, Pedroza e Mesquita (2003).

A comunicação entre o sistema proposto neste trabalho (isto é, o Servidor de Acessibilidade) e o Fogo Player é feita através da troca de mensagens. Por meio dessa comunicação, os sistemas podem enviar arquivos, dados e informações entre si. Para o sistema de acessibilidade, a comunicação é imprescindível para obter os arquivos de legenda e as informações referentes ao relógio lógico do Fogo Player, essenciais para manter o sincronismo entre os conteúdos acessíveis gerados pela solução proposta e as mídias do filme. Além disso, esse canal de comunicação também é usado pelo Fogo Player para enviar mensagens de controle. Pelo fato de serem estruturadas de maneira particular, a Figura 4.6 apresenta sua composição e os elementos que a compõe.

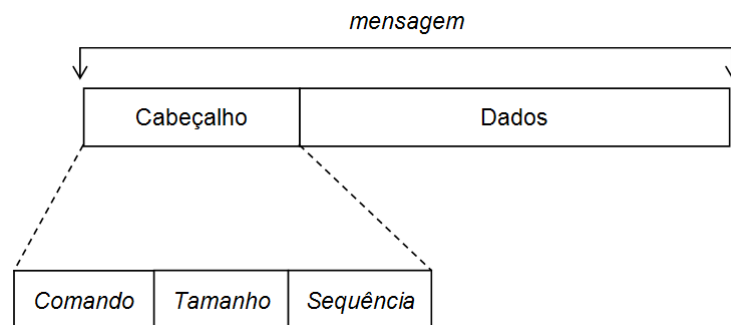


Figura 4.6: Estrutura das mensagens usadas na comunicação entre o sistema da solução proposta e o Fogo Player.

De acordo com a Figura 4.6, as mensagens são divididas em duas partes: *Cabeçalho* e *Dados*. O *Cabeçalho* é responsável por carregar informações a respeito das características da mensagem, por exemplo, o identificador do serviço (*Comando*) que

deve ser interpretado e executado pelo sistema receptor, o comprimento (*Tamanho*) da mensagem em bytes, e a continuidade (*Sequência*) de envio, caso uma mensagem precise ser fragmentada e transmitida em mais de uma seção. A Tabela 4.1 apresenta a especificação das mensagens utilizadas no Protocolo de Comunicação entre a solução proposta e o Fogo Player.

Tabela 4.1: Especificação dos tipos de mensagens utilizadas na comunicação entre a solução proposta e o Fogo Player.

REQUEST		RESPONSE	
IDENTITY	Solicita autenticação no player por meio do <i>Unique Identification</i> (UID) da aplicação	IDENTITY	Se a autenticação for realizada com sucesso, retorna OK, caso contrário, responde ERRO
CLOCK	Solicita o instante atual do player para o ajuste do relógio	PCR	Envia o Program Clock Reference (PCR) do player de cinema
SUB	Solicita os arquivos de legenda	SUB	Envia os arquivos de legenda
AD	Solicita os arquivos de Audiodescrição	AD	Envia os arquivos do roteiro de Audiodescrição
		INITIALIZE	Indica o início do filme
		FINALIZE	Indica que o filme foi encerrado

De acordo com a Tabela 4.1, é possível observar que as mensagens são agrupadas em duas categorias: *REQUEST* e *RESPONSE*. Toda mensagem do tipo REQUEST está relacionada a uma mensagem do tipo RESPONSE, exceto as mensagens cujo identificador são: *INITIALIZE* e *FINALIZE*, utilizadas pelo Fogo Player para informar

o início e o fim da reprodução do filme, respectivamente.

A outra parte da mensagem, isto é, *Dados*, é usada para encapsular os metadados e as informações de contexto. Por exemplo, os *tokens* de autenticação, as informações sobre o valor do PCR (*Program Clock Reference*) do Fogo Player, e os metadados dos arquivos de legenda são inseridos neste segmento da mensagem. A Figura 4.7 ilustra um diagrama de sequência que representa um possível cenário de comunicação entre o sistema da solução proposta e o Fogo Player.

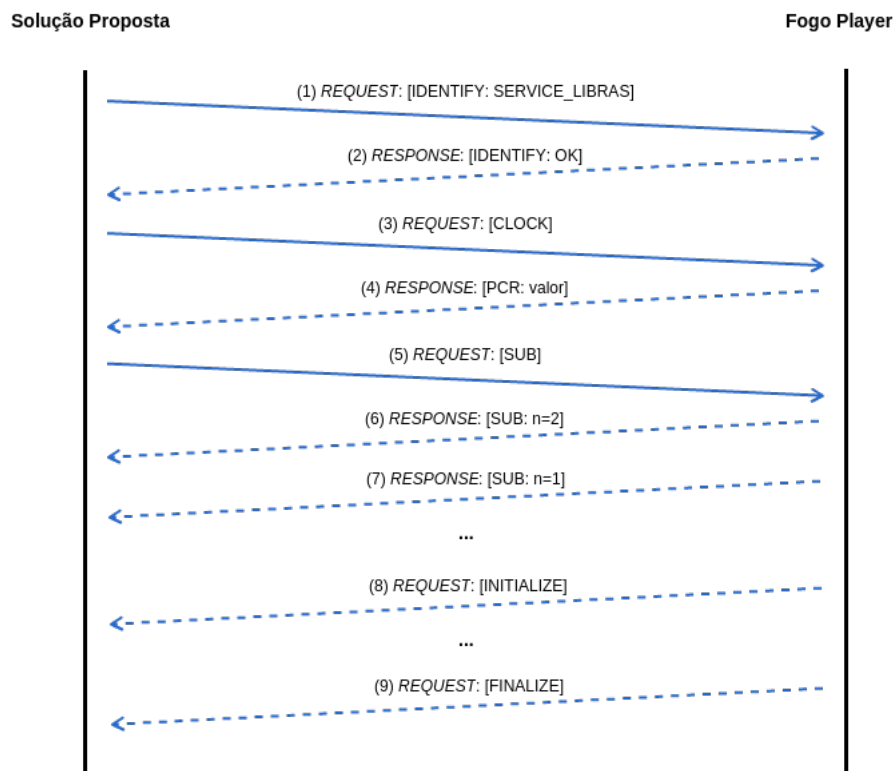


Figura 4.7: Cenário de comunicação entre a solução proposta e o Fogo Player.

Conforme ilustrado na Figura 4.7, inicialmente, para estabelecer uma comunicação, o servidor de acessibilidade (ver Figura 4.1) solicita sua autenticação no Fogo Player. Para isso, é enviada uma mensagem do tipo “IDENTIFY”, associada ao identificador único (UID). Neste cenário, os identificadores disponíveis para autenticação são: “SERVICE_LIBRAS”, que reconhece o sistema de geração das janelas de Libras, e “SERVICE_AD”, que autentica o módulo de geração de Audiodescrição.

Quando o Fogo Player recebe uma mensagem de autenticação, o mesmo verifica se o identificador recebido é válido, ou seja, se o token da mensagem consta em sua

base de dados. Por isso, é importante que os identificadores dos sistemas (Libras e Audiodescrição) sejam inseridos previamente na base do Fogo Player. Caso o UID seja válido, o Fogo Player responde enviando uma mensagem com o identificador “OK” ou, caso contrário, se não foi reconhecido, envia o identificador “ERRO”. Uma vez autenticado pelo Fogo Player, o sistema da solução proposta está habilitado para fazer requisições seguindo a especificação definida neste protocolo.

O Protocolo de Comunicação é fundamental para sincronizar os componentes de apresentação e geração de conteúdos acessíveis, isto é, geração das trilhas de vídeo com as janelas e Libras e da Audiodescrição.

4.3 Considerações

Este Capítulo apresentou uma proposta para tentar solucionar o problema de acesso à informação nos Cinemas Digitais enfrentados por usuários surdos ou cegos. Então, a ideia foi desenvolver um sistema computacional capaz de gerar de forma automática, trilhas alternativas de vídeo com janelas de Libras e Audiodescrição, e distribuir esses conteúdos em salas de cinema, onde os usuários podem utilizar dispositivos móveis de segunda tela para receber esses conteúdos. A arquitetura desta solução é ilustrada conforme a Figura 4.1.

Além dos componentes de software da solução proposta, também foi desenvolvido um protocolo de comunicação, apresentado na Seção 4.2. Este protocolo tem o função de especificar o tipo das mensagens utilizadas na comunicação entre o sistema desenvolvido e o player de cinema. Dessa forma, é possível garantir a coerência das requisições e dos conteúdos oferecidos, como, por exemplo, os arquivos de legenda, roteiro, informações de relógio e controle, dentre outras.

A seguir, o Capítulo 5 apresenta e discute os resultados de alguns testes de compreensão realizados com um protótipo do sistema. Esses testes são importantes para avaliar a qualidade dos conteúdos gerados pela solução e para investigar a satisfação dos usuários em utilizar um dispositivo de segunda tela para ter acesso à tradução em língua de sinais, no caso dos surdos, ou acompanhar os recursos da Audiodescrição através de fones de ouvido conectados ao dispositivo, no caso dos cegos.

Além dos testes de inteligibilidade dos conteúdos gerados, o Capítulo 5 também apresenta alguns resultados obtidos após a execução de testes de desempenho do sistema. Então, para que os testes fossem realizados, foi desenvolvido um protótipo da solução proposta. O principal objetivo desses testes foi investigar a média de atraso entre os conteúdos apresentados nos dispositivos de segunda tela e o conteúdo apresentado na primeira tela, ou seja, avaliar o nível de sincronismo entre os conteúdos acessíveis recebidos pelos usuários e o conteúdo da tela principal, isto é, o filme.

Capítulo 5

Resultados e Discussões

O Capítulo 1 contextualiza o problema de pesquisa investigado neste trabalho, que em síntese, trata dos problemas de comunicação e acesso à informação das pessoas com deficiência auditiva e visual nos cinemas digitais. Então, após apresentar alguns fundamentos e pesquisas equivalentes nos capítulos seguintes, o Capítulo 4 mostra em detalhes a solução desenvolvida. Embora a solução do problema tenha sido modelada com base em evidências apresentadas pela literatura científica, o fato de ser um sistema voltado para usuários com necessidades especiais levou à necessidade de realizar alguns testes com um protótipo da solução. Dessa forma, os testes foram planejados e divididos em duas etapas: testes computacionais e testes de inteligibilidade. Os testes computacionais mostram alguns resultados sobre o comportamento e o desempenho do sistema. Os testes de inteligibilidade observam o nível de compreensão dos conteúdos gerados e o grau de satisfação dos usuários com a solução proposta. Os resultados obtidos em cada etapa dos testes é apresentado e discutido neste capítulo.

5.1 Testes Computacionais

A execução do teste permitiu observar o comportamento do sistema com relação do nível de sincronismo dos conteúdos, e uma análise de desempenho de tráfego de rede na distribuição dos conteúdos para múltiplos dispositivos.

As Subseções 5.1.1, 5.1.2 e 5.1.3 apresentam as principais etapas da realização dos testes computacionais. A Subseção 5.1.1 mostra como o testes foi organizado, o qual

descreve o cenário utilizado e os materiais necessários para sua realização. Em seguida, na Subseção 5.1.2, é feita uma análise dos resultados obtidos. A Subseção 5.1.3 mostra algumas informações do tráfego de rede, como, por exemplo, taxa de transmissão, quantidade de pacotes perdidos, dentre outras informações. Por fim, a Subseção 5.1.4 apresenta uma discussão sobre os testes e os resultados observados.

5.1.1 Ambiente de Testes

Foram utilizados 10 dispositivos móveis para a realização do teste, onde, dentre esses, foram 4 *tablets* e 6 *smartphones*. Uma característica importante dos dispositivos foi que, para garantir a compatibilidade com o aplicativo desenvolvido, eles deveriam estar com a versão 4.4 (*Kitkat*), ou superior, do sistema operacional Android¹ instalada. Esse requisito é essencial em virtude da versão da API do sistema utilizada no desenvolvimento, versão 19. As configurações de hardware dos dispositivos móveis não foram observadas, dado que o objetivo do teste é simular um cenário próximo da realidade, ou seja, manter a dinâmica de participação do usuário independente de hardware utilizado.

Além dos dispositivos móveis, também foram utilizados um computador pessoal convencional — um notebook — e um roteador sem fio. O computador usado possui um processador Intel® Core™ i5-3210M com 2.5GHz, memória principal (RAM) de 6GB DDR3, e interface de rede sem fio *Qualcomm Atheros*, modelo AR9485, e clock de 33MHz. Neste computador, foram instalados todos os componentes do sistema apresentados na Figura 4.1.

O roteador, do fabricante D-Link², foi usado como ponto de acesso para a distribuição dos conteúdos de áudio e vídeo acessível, e o mesmo oferece suporte aos protocolos IEEE 802.11b, IEEE 802.11g e IEEE 802.11n de rede Wireless, além de uma taxa máxima de transferência de dados de 300 Mbps, e frequência de banda de 2.4GHz.

Como a dinâmica do teste foi restrita à transmissão de conteúdos de vídeo, até mesmo para facilitar a coleta dos dados, foi gerado um vídeo contendo a apresentação de um cronômetro durante toda a sua exibição. Para a geração do vídeo, foi usado o padrão

¹<https://www.android.com>

²<http://www.dlink.com.br>

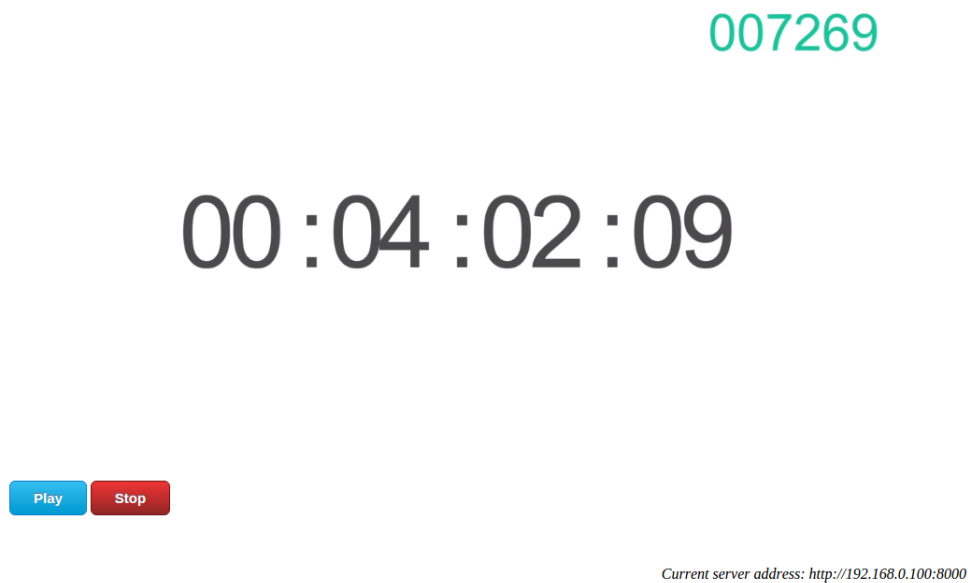


Figura 5.1: Vídeo com cronômetro usado no teste.

de compressão H.264 baseado no MPEG-4. O vídeo possui duração de 30 minutos, com resolução de 1280x720, e codificado a 30 quadros por segundo (ou simplesmente 30 *fps*³). Dessa forma, a ideia é que todos os clientes (isto é, dispositivos móveis) registrados na aplicação recebam a mesma fonte de dados. A Figura 5.1 mostra uma captura da tela em um determinado momento durante a exibição do vídeo.

Observando a Figura 5.1, é possível notar que o cronômetro é composto por quatro contadores. Esses contadores seguem o padrão HH:mm:ss:fps, que indica, nesta mesma ordem, o tempo em horas, minutos, segundos, e *frames* por segundo. O fps faz a contagem de cada quadro do vídeo até completar um segundo, e então é zerado novamente. Por exemplo, para um vídeo codificado a 30 fps, esse contador vai de 0 (zero) a 29 (vinte e nove) a cada segundo de vídeo. Além desses contadores, outro contador pode ser visto no canto superior direito da Figura 5.1. Esse contador indica a quantidade de quadros exibidos até o momento, ou seja, é um contador acumulativo.

Na execução do teste, foi utilizada a tela de um monitor para exibir o vídeo com o cronômetro ao mesmo tempo em que os dispositivos recebiam e apresentavam o mesmo conteúdo. O vídeo apresentado na tela do monitor foi controlado pelo Fogo Player (ver Seção 2.4). Dessa forma, os botões de *Play* e *Stop* observados na Figura 5.1 são

³*Frames Per Second*.

responsáveis por gerar os eventos de início e fim do vídeo, respectivamente.

Para facilitar a coleta dos dados, foi utilizada uma câmera para gravar todo o teste. Essa gravação foi feita de tal forma que fosse possível visualizar, simultaneamente, a tela do monitor e a dos dispositivos. O tempo total de gravação foi de 30 minutos, e com esse intervalo de tempo, optou-se por coletar 50 amostras selecionadas aleatoriamente. Então, para isso, foram gerados 50 números não repetidos de 0 a 54.000 — em um vídeo com duração de 30 minutos e codificado a 30 fps, no total, são 54 mil quadros. Dessa forma, os números gerados representam o quadro em que o vídeo do servidor deve ser pausado, e então, contabilizados os quadros de cada dispositivo. A Figura 5.2 mostra alguns quadros capturados.

De acordo com a Figura 5.2, é possível visualizar o instante em que os quadros 1459 (Figura 5.2a), 14799 (Figura 5.2b), 19813 (Figura 5.2c) e 42231 (Figura 5.2d)

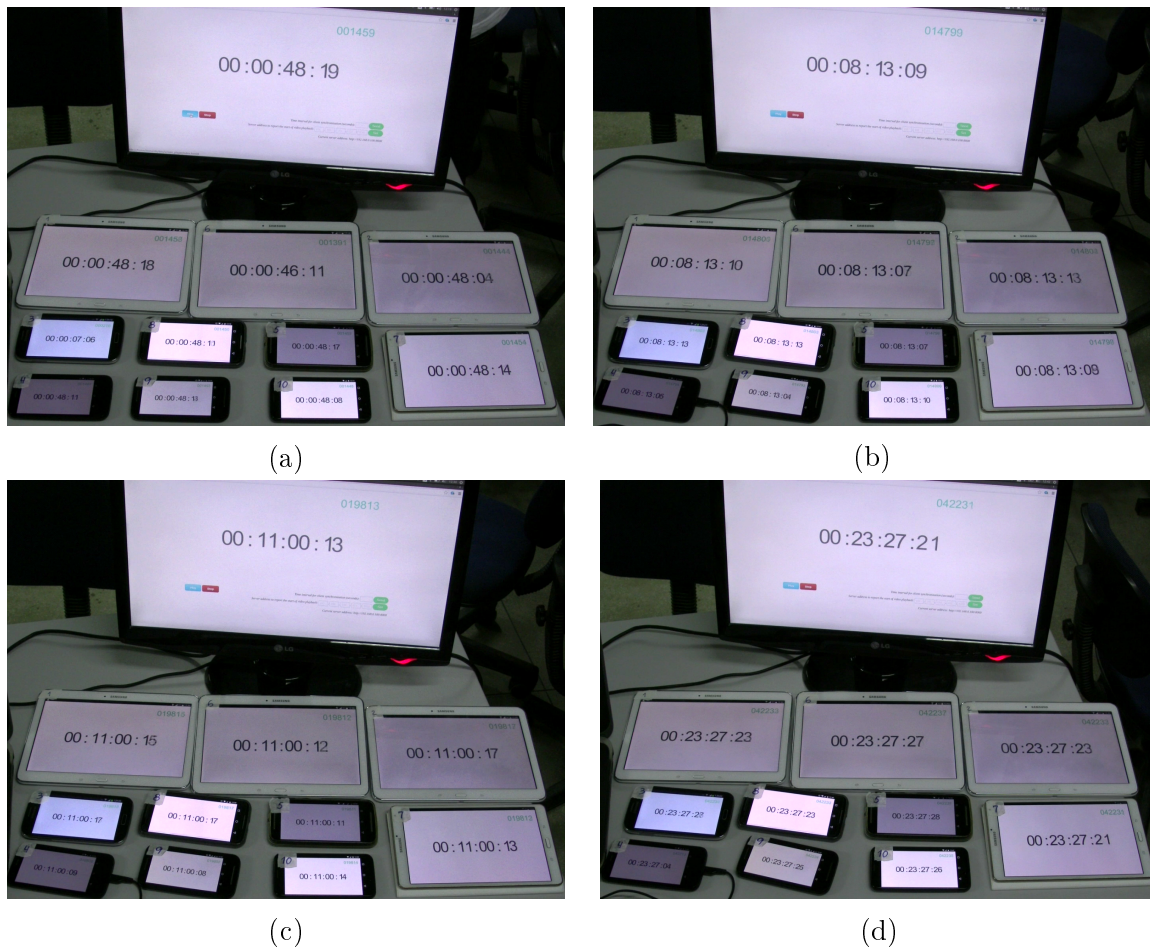


Figura 5.2: Momento da captura dos quadros: (a) 1459; (b) 14799; (c) 19813; (d) 42231.

foram capturados. Então, dessa forma, são observados e contabilizados os quadros do dispositivo em cada uma das marcações de tempo geradas.

A Subseção 5.1.2 apresenta o conjunto de amostras e os resultados obtidos após uma análise estatística.

5.1.2 Análise dos Resultados

A Tabela 5.1 apresenta a referência de cada quadro dos dispositivos no momento em que o vídeo principal foi pausado. Os valores expostos na coluna *Quadro* correspondem ao quadro do vídeo principal usado como base. Esses valores foram gerados aleatoriamente dentro de um intervalo de 30 minutos, ou seja, valores entre 0 e 54.000.

Observando a Tabela 5.1 é possível notar que alguns quadros referentes aos dispositivos não foram contabilizados. Esse comportamento aconteceu especificamente com os dispositivos 3, 4, 5 e 6. Com os dispositivos 3 e 6, houve um certo atraso para iniciar a reprodução do conteúdo. Já com os dispositivos 4 e 5, essa irregularidade aconteceu em tempo de reprodução, que pode ser percebido nas amostras de 31 a 38 referente ao dispositivo 4, e de 42 a 46 no dispositivo 5.

Uma possível explicação para o atraso inicial sofrido pelos dispositivos 3 e 6 é que, por limitações de hardware, por exemplo, não foi possível carregar dados suficientes até a chegada de uma nova mensagem de sincronismo. Dessa forma, ao observar o novo tempo calculado, e que ainda não havia sincronismo, os dados carregados até então foram descartados e então, uma nova requisição feita ao servidor em busca de dados mais atualizados.

O comportamento imprevisível dos dispositivos 4 e 5 pode ser explicado de duas maneiras. A primeira, que pode justificar a falha com o dispositivo 4 é que durante os testes, uma ligação foi recebida. Então, por alguma razão interna, é possível que o evento da ligação tenha comprometido a execução normal da aplicação. Em relação a falha com o dispositivo 5, a explicação provável é que, enquanto os testes eram executados, uma mensagem foi recebida alertando que a bateria estava fraca.

A falha com os dispositivos 3 e 6 foram corrigidas antes de iniciar a reprodução do conteúdo. Porém, o problema com os dispositivos 4 e 5 foi reparado em tempo de execução. Para isso, no momento em que o problema foi detectado — não imediatamente,

Tabela 5.1: Captura dos quadros referente a cada amostra do teste.

Amostra	Quadro	Dispositivos									
		1	2	3	4	5	6	7	8	9	10
1	1459	1458	1444	-	1451	1457	-	1454	1451	1453	1448
2	1844	1844	1823	-	1836	1843	-	1839	1836	1839	1833
3	3024	3021	3024	-	3036	3023	3018	3019	3015	3019	3013
4	4571	4573	4575	4574	4567	4574	4569	4570	4553	4570	4572
5	6864	6866	6868	6868	6861	6867	6863	6864	6868	6863	6865
6	8174	8176	8179	8178	8171	8178	8173	8174	8167	8174	8175
7	8702	8704	8706	8706	8699	8705	8700	8702	8694	8701	8703
8	9713	9714	9717	9717	9719	9716	9712	9713	9705	9712	9714
9	12572	12574	12577	12576	12569	12571	12571	12572	12576	12571	12574
10	12880	12882	12885	12884	12877	12878	12879	12880	12884	12879	12881
11	12895	12897	12900	12899	12892	12894	12894	12895	12899	12894	12896
12	13034	13035	13038	13038	13030	13032	13035	13033	13034	13033	13035
13	13782	13784	13786	13786	13778	13780	13781	13782	13786	13781	13783
14	13939	13941	13943	13943	13935	13937	13938	13939	13943	13938	13940
15	14799	14800	14803	14803	14796	14797	14797	14803	14803	14794	14800
16	17479	17480	17483	17483	17476	17477	17477	17478	17483	17474	17480
17	18391	18392	18395	18394	18387	18389	18389	18390	18395	18386	18391
18	19101	19102	19105	19105	19097	19099	19100	19101	19104	19096	19102
19	19813	19815	19817	19817	19809	19811	19812	19813	19817	19808	19814
20	20369	20371	20373	20373	20365	20367	20367	20369	20373	20364	20370
21	20563	20564	20567	20567	20559	20558	20562	20562	20566	20558	20564
22	20574	20576	20578	20578	20570	20569	20572	20574	20577	20569	20575
23	21022	21024	21026	21026	21018	21017	21021	21022	21017	21017	21023
24	22259	22260	22263	22263	22255	22254	22257	22258	22256	22263	22260
25	25145	25146	25149	25146	25142	25140	25143	25145	25142	25149	25145
26	30477	30477	30478	30478	30474	30481	30475	30477	30482	30480	30476
27	30536	30537	30537	30537	30533	30540	30534	30536	30541	30539	30536
28	31276	31277	31277	31277	31273	31280	31274	31275	31280	31279	31276
29	31396	31397	31397	31397	31393	31400	31394	31395	31400	31399	31395
30	31458	31459	31459	31459	31456	31463	31457	31458	31463	31462	31458
31	32917	32918	32918	32918	-	32919	32915	32916	32921	32920	32916
32	34010	34011	34011	34011	-	34014	34005	34009	34026	34013	34010
33	34125	34127	34127	34126	-	34129	34101	34124	34134	34128	34125
34	36521	36522	36522	36522	-	36520	36527	36520	36523	36526	36521
35	38165	38166	38166	38166	-	38165	38171	38165	38167	38169	38165
36	38440	38441	38441	38441	-	38440	38446	38440	38441	38443	38439
37	39263	39264	39264	39264	-	39262	39269	39262	39264	39266	39262
38	41805	41807	41807	41812	-	41811	41812	41805	41807	41809	41810
39	42231	42233	42233	42237	42214	42238	42237	42231	42233	42235	42236
40	43313	43315	43315	43320	43321	43316	43320	43314	43316	43318	43319
41	43674	43676	43676	43677	43681	43681	43681	43674	43676	43674	43679
42	45968	45969	45969	45966	45972	-	45975	45968	45970	45963	45967
43	46145	46147	46147	46143	46150	-	46152	46145	46147	46141	46144
44	46415	46416	46416	46413	46420	-	46422	46415	46417	46411	46414
45	47535	47536	47527	47533	47513	-	47542	47522	47537	47531	47534
46	48512	48513	48507	48510	48521	-	48519	48525	48514	48508	48507
47	49461	49462	49456	49459	49470	49471	49468	49464	49463	49457	49456
48	49580	49581	49575	49578	49589	49588	49587	49584	49582	49576	49575
49	52183	52185	52191	52190	52180	52187	52187	52187	52185	52179	52195
50	53745	53747	53755	53752	53740	53749	53750	53749	53747	53742	53752

Tabela 5.2: Atraso dos dispositivos em relação a cada amostra de quadro do servidor

Amostra	Quadro	Dispositivos									
		1	2	3	4	5	6	7	8	9	10
1	1459	1	15	-	8	2	-	5	8	6	11
2	1844	0	21	-	8	1	-	5	8	5	11
3	3024	3	0	-	12	1	6	5	9	5	11
4	4571	2	4	3	4	3	2	1	18	1	1
5	6864	2	4	4	3	3	1	0	4	1	1
6	8174	2	5	4	3	4	1	0	7	0	1
7	8702	2	4	4	3	3	2	0	8	1	1
8	9713	1	4	4	6	3	1	0	8	1	1
9	12572	2	5	4	3	1	1	0	4	1	2
10	12880	2	5	4	3	2	1	0	4	1	1
11	12895	2	5	4	3	1	1	0	4	1	1
12	13034	1	4	4	4	2	1	1	0	1	1
13	13782	2	4	4	4	2	1	0	4	1	1
14	13939	2	4	4	4	2	1	0	4	1	1
15	14799	1	4	4	3	2	2	4	4	5	1
16	17479	1	4	4	3	2	2	1	4	5	1
17	18391	1	4	3	4	2	2	1	4	5	0
18	19101	1	4	4	4	2	1	0	3	5	1
19	19813	2	4	4	4	2	1	0	4	5	1
20	20369	2	4	4	4	2	2	0	4	5	1
21	20563	1	4	4	4	5	1	1	3	5	1
22	20574	2	4	4	4	5	2	0	3	5	1
23	21022	2	4	4	4	5	1	0	5	5	1
24	22259	1	4	4	4	5	2	1	3	4	1
25	25145	1	4	1	3	5	2	0	3	4	0
26	30477	0	1	1	3	4	2	0	5	3	1
27	30536	1	1	1	3	4	2	0	5	3	0
28	31276	1	1	1	3	4	2	1	4	3	0
29	31396	1	1	1	3	4	2	1	4	3	1
30	31458	1	1	1	2	5	1	0	5	4	0
31	32917	1	1	1	-	2	2	1	4	3	1
32	34010	1	1	1	-	4	5	1	16	3	0
33	34125	2	2	1	-	4	24	1	9	3	0
34	36521	1	1	1	-	1	6	1	2	5	0
35	38165	1	1	1	-	0	6	0	2	4	0
36	38440	1	1	1	-	0	6	0	1	3	1
37	39263	1	1	1	-	1	6	1	1	3	1
38	41805	2	2	7	-	6	7	0	2	4	5
39	42231	2	2	6	17	7	6	0	2	4	5
40	43313	2	2	7	8	3	7	1	3	5	6
41	43674	2	2	3	7	7	7	0	2	0	5
42	45968	1	1	2	4	-	7	0	2	5	1
43	46145	2	2	2	5	-	7	0	2	4	1
44	46415	1	1	2	5	-	7	0	2	4	1
45	47535	1	8	2	22	-	7	13	2	4	1
46	48512	1	5	2	9	-	7	13	2	4	5
47	49461	1	5	2	9	10	7	3	2	4	5
48	49580	1	5	2	9	8	7	4	2	4	5
49	52183	2	8	7	3	4	4	4	2	4	12
50	53745	2	10	7	5	4	5	4	2	3	7

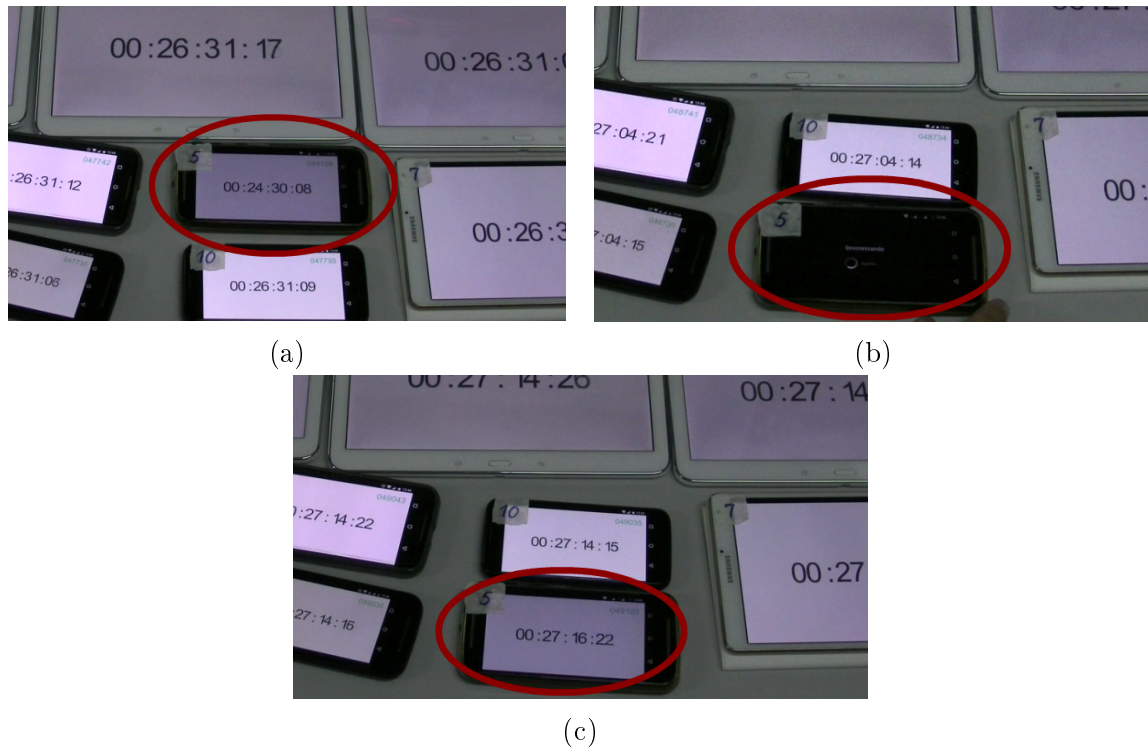


Figura 5.3: (a) Momento em que o dispositivo 5 interrompe a exibição do conteúdo; (b) depois de solicitar novamente seu registro na aplicação, aguarda o processo de sincronismo; (c) reinicia a exibição do vídeo.

o dispositivo foi retirado, e então, uma nova conexão foi estabelecida com o servidor. A Figura 5.3 mostra esse procedimento com o dispositivo 5.

De acordo com a Figura 5.3, ao perceber que o conteúdo exibido no dispositivo 5 parou após receber uma mensagem alertando que a bateria estava fraca (Figura 5.3a), o mesmo então foi desconectado da aplicação e, em seguida, uma nova requisição feita ao servidor (Figura 5.3b). Dessa forma, após a aplicação ajustar o relógio e carregar uma quantidade suficiente de dados, o dispositivo voltou a reproduzir o conteúdo como pode ser visto na Figura 5.3c.

Apesar de existir indícios de que as falhas observadas na Tabela 5.1 foram ocasionadas pelos eventos imprevisíveis ocorridos durante o teste, não é possível garantir que essa correlação, de fato, existe. Dessa forma, para assegurar que as falhas foram decorrentes desses eventos, é necessário realizar novos testes com a aplicação, e reproduzir os mesmos eventos para tentar provar sua efetividade.

Observando os dados apresentados na Tabela 5.1, é possível extrair informações sobre o nível de sincronização dos conteúdos. Então, com esse propósito, foi calculada

a diferença entre os quadros de cada dispositivo em relação ao quadro do servidor, e o resultado é apresentado na Tabela 5.2. Os valores expostos na Tabela 5.2 indicam a diferença entre os quadros independente da posição que se encontre, ou seja, indica a quantidade de quadros atrasados ou adiantados.

Analisando valores apresentados na Tabela 5.2, percebe-se que os dispositivos conseguiram manter um nível de sincronismo razoável. No entanto, houve uma diferença relativamente alta em alguns casos, como, por exemplo, quando a diferença alcançou os valores 24, 22, 21, dentre outros inferiores a estes. Então, para ilustrar essa variação de sincronismo, a Figura 5.4 apresenta um gráfico com informações de atraso mínimo, médio e máximo para cada amostra.

Conforme ilustrado na Figura 5.4, existem alguns pontos no gráfico que demonstram que os dispositivos atingiram um nível de sincronismo satisfatório. Porém, é possível notar também que houve uma certa variação em alguns pontos. Por exemplo, analisando a linha que representa os maiores valores, percebe-se que até a quarta amostra, houve bastante variação de atraso. Uma das possíveis explicações para essa variação é que nos primeiros momentos do teste, pelo fato da maior parte dos dispositivos não estarem sincronizados, algumas requisições acabaram atrasando mais do que outras.

Além dos atrasos observados nos primeiros momentos do teste, outros pontos que

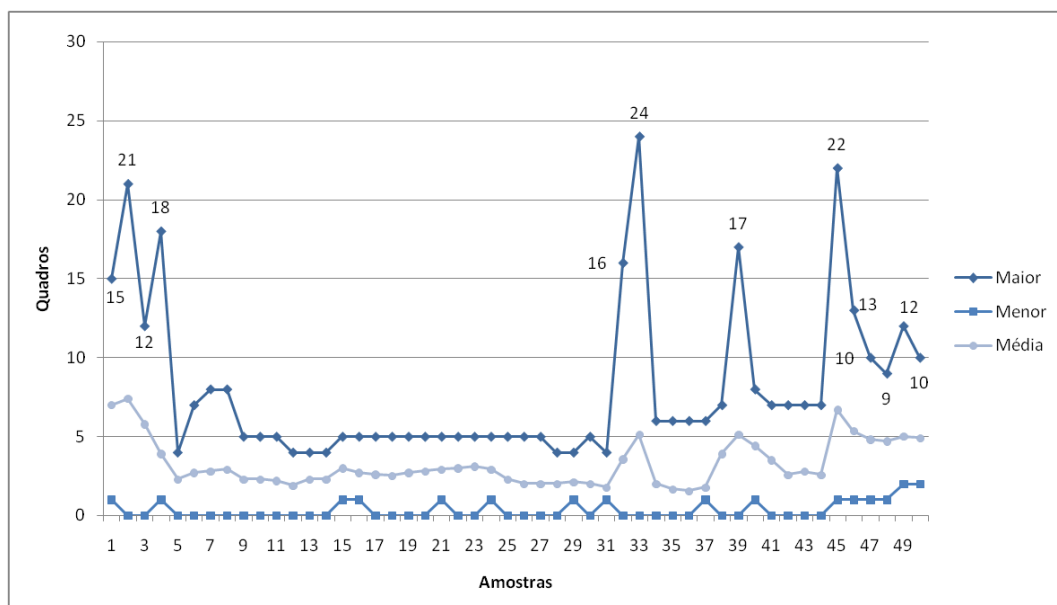


Figura 5.4: Variação de atraso mínimo, médio e máximo de cada dispositivo.

sofreram bastante variação podem ser vistos nas amostras 32, 33, 39, e nas últimas cinco, ou seja, nas amostras de 45 a 50. Essa variação mais acentuada nesses pontos indica que alguns dispositivos tentavam sincronizar seus conteúdos naquele momento. Ao consultar a Tabela 5.2, é possível descobrir quais dispositivos apresentam esse comportamento.

Outra maneira de analisar o nível de sincronismo dos dispositivos é através de um gráfico de dispersão. Então, como esse propósito, foi gerado um gráfico *box plot*, conforme ilustrado na Figura 5.5. De acordo com o gráfico, houve uma baixa dispersão dos valores observados. O maior valor referente ao limite superior calculado foi igual a 9 quadros, que pertenceu ao dispositivo 4. O dispositivo que apresentou maior dispersão dos valores foi o número 6, que teve uma variação inter-quartil de 1 a 6 quadros. Por outro lado, o dispositivo 7 foi o que mostrou a maior concentração dos valores, ou seja, foi o dispositivo que se manteve mais tempo sincronizado.

Ainda de acordo com o gráfico apresentado na Figura 5.5, é possível perceber também a presença de alguns *outliers*. Um *outlier* é uma observação que se encontra a uma distância anormal de outros valores em uma amostra aleatória de uma população (NIST-SEMATECH, 2012). Dessa forma, considerando o conjunto de amostras, exceto as 18 que não foram contabilizadas (ver Tabela 5.2), 21 foram considerados *outlier*.

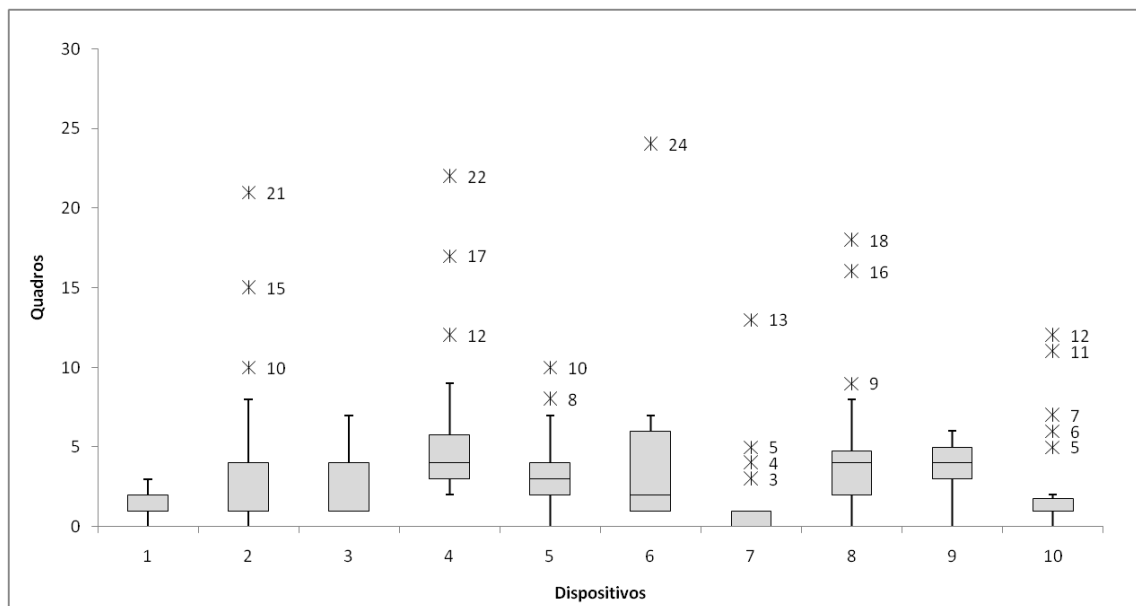


Figura 5.5: Dispersão do nível de sincronismo de cada dispositivo.

Tabela 5.3: Cálculos das medidas de dispersão.

	Dispositivos									
	1	2	3	4	5	6	7	8	9	10
Maior	3	21	7	22	10	24	13	18	6	12
Menor	0	0	1	2	0	1	0	0	0	0
Média	1,42	3,88	3,11	5,38	3,31	3,88	1,48	4,30	3,36	2,36
Desvio Padrão	0,61	3,62	1,81	3,93	2,12	3,85	2,82	3,38	1,63	3,18
Mediana	1	4	4	4	3	2	0	4	4	1

Isso representa 4,36% das 482 amostras observadas.

A Tabela 5.3 apresenta o cálculo das medidas de dispersão para cada dispositivo usado no teste. Dentre essas medidas, foram calculados os valores da média, mediana e desvio padrão, além de apresentar o maior e o menor valor encontrado. Uma observação importante na Tabela 5.3 é que o cálculo do desvio padrão foi relativamente alto para os dispositivos que apresentaram pelo menos um *outlier*. Em contrapartida, aqueles que não tiveram nenhum *outlier* apresentaram um desvio padrão considerado baixo. Esse foi o caso dos dispositivos 1, 3 e 9.

O gráfico apresentado na Figura 5.5 mostra a dispersão do nível de sincronismo em cada dispositivo individualmente. No entanto, com o objetivo de visualizar a dispersão de sincronismo de modo geral — isto é, contabilizando todas as amostras analisadas, foi gerado um gráfico *box plot* com todos os valores apresentados na Tabela 5.2. Esse gráfico é mostrado na Figura 5.6.

Examinando o gráfico ilustrado na Figura 5.6, nota-se que a maior concentração dos valores ficou entre o primeiro e o terceiro quartil, isto é, 1 e 4, respectivamente. Isso significa que o nível de sincronismo entre os dispositivos pode ser considerado satisfatório. Para esse gráfico, o valor da média foi igual a 3,21, e desvio padrão igual a 3,08. Novamente, o motivo do desvio padrão ter sido alto em relação a média, foi pelo fato da presença dos *outliers*, conforme pode ser visto na Figura 5.6.

De acordo com a Associação Brasileira de Normas Técnicas (ABNT), em sistemas de transmissão de *closed caption* em tempo real, o atraso máximo da informação não pode exceder 4 segundos (ABNT, 2005). Dessa forma, pode-se assumir que os surdos toleram até 4 segundos para ter acesso à informação do conteúdo. Nesse caso, como

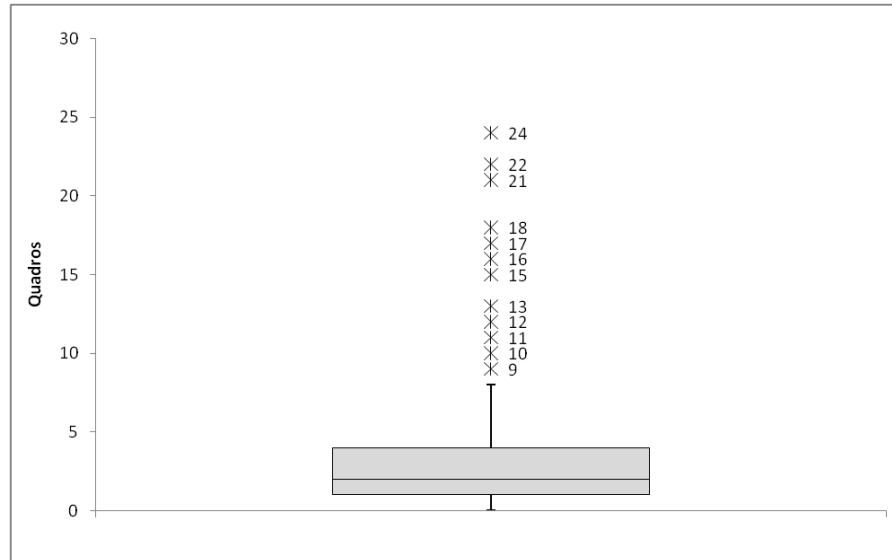


Figura 5.6: Dispersão geral do nível de sincronismo entre os dispositivos.

o atraso médio observado (ver Figura 5.6) foi igual a 3,21 quadros — com um desvio padrão de 3,08, a solução apresentada neste trabalho se mostrou eficiente, e também pode ser considerada viável para a geração dos conteúdos de Libras em salas de cinema digital, uma vez que, em média, o atraso é inferior a 210 milissegundos⁴.

5.1.3 Análise do Tráfego de Rede

Além da análise de sincronismo dos dispositivos, outro aspecto investigado durante o teste foi o tráfego da rede. Para isso, foi utilizada a ferramenta *ntopng*. O *ntopng* permite usar um navegador web para obter diversas métricas sobre o tráfego de rede, como, por exemplo, taxa de transmissão, quantidade de pacotes perdidos, dentre outras informações (NTOPNG, 2015). A Figura 5.7 ilustra um gráfico que demonstra o comportamento da rede durante o teste.

De acordo com o gráfico ilustrado na Figura 5.7, é possível observar o tráfego da rede durante o experimento. Esse tráfego representa o consumo de dados pelos clientes. Em alguns pontos no gráfico, nota-se que houve uma variação mais intensa na taxa de transmissão. Possivelmente, esse aumento na taxa ocorreu devido o processo de sincronização de alguns dispositivos, que naturalmente enviam várias requisições

⁴O cálculo foi baseado em um vídeo codificado a 30 quadros por segundo (fps), onde cada quadro tem aproximadamente 33,4 milissegundos.

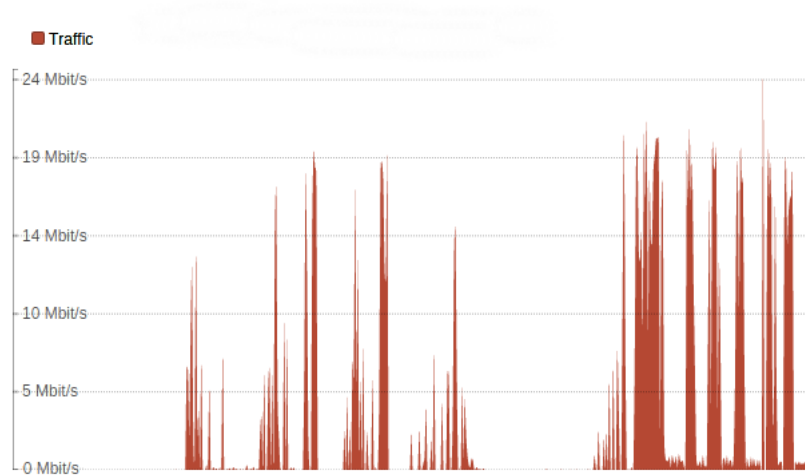


Figura 5.7: Análise do tráfego de rede durante o teste.

ao servidor solicitando novos segmentos de mídia do conteúdo. A Figura 5.8 mostra algumas informações estatísticas sobre o tráfego ilustrado na Figura 5.7.

Conforme as informações expostas na Figura 5.8, no momento em que esses dados foram analisados, a taxa máxima de transmissão atingiu 24,26 Mbit/s. Entretanto, a taxa média de transmissão registrada foi de 2,62 Mbit/s, e a última observada de 113,75 Kbit/s. Novamente, essa variação pode ter sido ocasionada por algumas amostras consideradas *outlier*, como é possível observar na Figura 5.6.

	Time	Value
Min	08/17/15 12:13:55	0 bps
Max	08/17/15 12:28:43	24.26 Mbit/s
Last	08/17/15 12:32:32	113.75 Kbit/s
Average	2.62 Mbit/s	

Figura 5.8: Informações estatísticas sobre o tráfego de rede após 13 minutos.

Por fim, a Figura 5.9 trás outras informações estatísticas da rede após 27 minutos de teste. Nesse momento, a análise do tráfego feita pela ferramenta *ntopng* registrou uma taxa máxima de transmissão de 2,11 Mbit/s, bem inferior a taxa média vista na Figura 5.8. Além dessa informação, é possível observar também que a taxa mínima foi de 71,9 Kbit/s e a média igual a 812,2 Kbit/s. É provável que essa redução na taxa de transmissão tenha ocorrido em virtude do grau de sincronismo dos dispositivos. Essa

	Time	Value
Min	08/17/15 12:42:48	71.9 Kbit/s
Max	08/17/15 12:45:49	2.11 Mbit/s
Last	08/17/15 12:46:49	48.12 Kbit/s
Average	812.2 Kbit/s	

Figura 5.9: Informações estatísticas sobre o tráfego de rede após 27 minutos.

análise de sincronismo entre os dispositivos é apresentada em detalhes na Subseção 5.1.2.

5.1.4 Discussão

Nas Subseções 5.1.1, 5.1.2 e 5.1.3 foram apresentados alguns testes computacionais realizados com um protótipo da solução. O objetivo do teste foi investigar o nível de sincronismo e o seu desempenho da distribuição dos conteúdos. Em relação ao nível de sincronismo, os resultados foram considerados satisfatórios (ver Figura 5.6). A média da disparidade entre os quadros dos dispositivos foi de 3,21 quadros — com desvio padrão de 3,08. Isso significa que a diferença entre os conteúdos, em média, não ultrapassou 210 milissegundos. Portanto, considerando o atraso máximo de 4 segundos definido pela NBR-15290 (ABNT, 2005), a solução é capaz de gerar conteúdos acessíveis em Libras em salas de cinema digital.

Em relação ao desempenho da distribuição dos conteúdos, durante a realização do teste, a ferramenta *ntopng* (NTOPNG, 2015) foi utilizada para analisar o tráfego da rede. Os resultados dessa análise mostraram que a média de tráfego dos conteúdos aos 13 minutos foi igual a 2,62 Mbit/s (Figura 5.8), e após 27 minutos de teste, a média caiu para 812,2 Kbit/s (Figura 5.9). Apesar da pouca quantidade de dispositivos, é possível perceber que o servidor conseguiu atender as requisições e também distribuir os conteúdos com certa folga.

Por fim, mesmo tendo alcançado um grau de sincronismo considerável e uma taxa de transmissão sob controle, é necessário utilizar uma quantidade maior de dispositivos para realizar avaliações estatísticas mais aprofundadas e, dessa forma, garantir que o desempenho da solução será satisfatório, ou então, definir até que ponto pode

ser considerada viável. A avaliação do sincronismo e desempenho de distribuição dos conteúdos de Audiodescrição gerados pela solução fica como uma proposta de trabalho futuro.

5.2 Avaliação dos Conteúdos de Libras

Com o objetivo de investigar o nível de compreensão dos conteúdos gerados e o grau de satisfação dos surdos, foram feitos alguns testes usando um protótipo da solução proposta. Os testes foram realizados em duas etapas. Na primeira, os surdos tiveram que utilizar as legendas no formato texto. Na segunda, os surdos utilizaram os conteúdos acessíveis gerados pela solução proposta.

Os testes foram realizados com vinte usuários surdos brasileiros. O grupo era heterogêneo e foi composto por sete homens e treze mulheres, com idade entre 13 e 56 anos e média de 30,4 anos. O grau de escolaridade dos surdos também foi observado. Dentre eles, haviam desde surdos com ensino fundamental incompleto até surdos com formação superior incompleta.

A realização dos testes aconteceu em duas localidades, sendo uma para cada etapa da avaliação. A primeira aconteceu na Fundação Centro Integrado de Apoio a Pessoa com Deficiência (FUNAD), que fica localizada na cidade de João Pessoa-PB. Nesta, participaram dez usuários surdos, que tiveram o recurso das legendas como principal meio de acesso à informação.

A outra localidade em que foram realizados os testes foi no Instituto Nacional de Educação de Surdos (INES), que fica situado na cidade do Rio de Janeiro-RJ. Do mesmo modo, foram envolvidos nesta etapa do teste, dez usuários surdos. Cada surdo desse grupo utilizou um dispositivo móvel (referente ao modelo iPad 2 da marca *Apple Inc.*) para acompanhar a tradução do filme gerada pela solução proposta. O software usado no dispositivo para receber o conteúdo acessível foi o VLC⁵.

O processo de avaliação foi o mesmo nas duas etapas. Dessa forma, em um primeiro momento, foi apresentado para os surdos o objetivo principal dos testes. Em seguida, os surdos receberam um questionário, dividido em três partes, no qual tiveram que

⁵<http://www.videolan.org/vlc>

preencher com algumas informações pessoais, como, idade, grau de escolaridade, sexo, dentre outros. Concluído o preenchimento da primeira parte do questionário, os surdos foram convidados para assistir o filme Sintel ⁶. Sintel é um filme de gênero animado, produzido de forma colaborativa pela *Blender Foundation*⁷, disponível em 4K (isto é, UHD), e com tempo de duração de aproximadamente quinze minutos. Depois da exibição do filme, os usuários foram convidados a preencher o restante do questionário.

A segunda parte do questionário apresentou quatro questões relacionadas ao conteúdo do filme. Essas questões foram elaboradas de modo que era necessário compreender o diálogo (através da legenda ou pela tradução em Libras) para responder corretamente. Para esse conjunto de perguntas, os surdos tinham que escolher dentre as alternativas A, B, C ou D, onde a alternativa D representou a opção “Não sei”, que foi incluída para prevenir que fossem escolhidas respostas aleatórias quando os surdos não soubessem a resposta correta.

A terceira parte do questionário foi aplicada apenas para o grupo de surdos que acompanhou a tradução do filme no dispositivo móvel. O objetivo desse conjunto de perguntas foi investigar o nível de satisfação dos surdos em utilizar um dispositivo móvel para acompanhar a tradução do filme. Nessa parte, foram feitas duas perguntas: a primeira sobre o grau de dificuldade em usar um dispositivo móvel; e a segunda, sobre o nível de complexidade em acompanhar, simultaneamente, o conteúdo do filme e a tradução em Libras na tela do dispositivo móvel. Do mesmo modo, para representar sua resposta, os surdos escolheram um número em uma escala de 1 a 6, sendo 1 muito difícil e 6 muito fácil. Durante a realização dos testes, intérpretes de Libras da FUNAD e do INES auxiliaram na comunicação com os surdos.

A Figura 5.10 apresenta o percentual do resultado obtido nos testes de inteligibilidade, referente a segunda parte do questionário.

De acordo com a Figura 5.10, é possível perceber que os surdos têm muitas dificuldades em acompanhar o conteúdo de um filme através das legendas, pois, conforme ilustrado no gráfico (Figura 5.10), com essa abordagem, os surdos responderam, em média, apenas 35% das respostas corretas. Em contrapartida, com a solução proposta

⁶<http://www.sintel.org>

⁷<http://www.blender.org/foundation/>

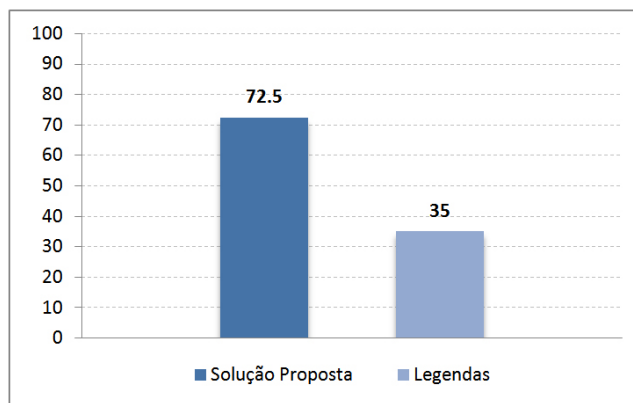


Figura 5.10: Percentual de acertos no teste de inteligibilidade.

neste trabalho, os surdos conseguiram responder até 72,5% das questões corretamente.

Na terceira parte do questionário, foi investigado o nível de compreensão, o entendimento da gramática e a naturalidade dos sinais gerados pelo Avatar 3D. Nessa parte, os surdos tiveram que escolher um número em uma escala de 1 a 6, onde 1 significa muito ruim (ou pouco) e 6, excelente (muito).

Para os conteúdos com legendas, algumas inconsistências nos resultados foram encontradas. Por exemplo, com respeito a gramática, alguns surdos indicaram que os conteúdos foram compatíveis com a gramática de Libras (com média de 4,3), onde, na realidade, as legendas foram apresentadas de acordo com a gramática do português brasileiro. Além disso, o nível de entendimento não foi compatível com a performance desses usuários nos testes de compreensão, que obteve um valor médio de 3,7.

Para os conteúdos acessíveis gerados pela solução proposta, os aspectos investigados tiveram resultados considerados modestos, além de uma média de desvio padrão (DP) relativamente alta, chegando a 1,89 (31,5%). Isso significa que houve uma divergência considerável entre as opiniões dos surdos. De acordo com San-Segundo e AL (2008), uma possível explicação para isso é a flexibilidade da estrutura de algumas sentenças em língua de sinais. Essa flexibilidade pode, em alguns momentos, comprometer o entendimento das pessoas surdas. Outra possibilidade é o regionalismo, que dependendo da localidade, a configuração de alguns sinais pode ser diferente.

Observando a dispersão dos resultados obtidos no teste de compreensão (apresentada no gráfico da Figura 5.11), nota-se que os surdos que avaliaram a solução proposta tiveram pouca variação em seus resultados. Esse comportamento pode ser percebido

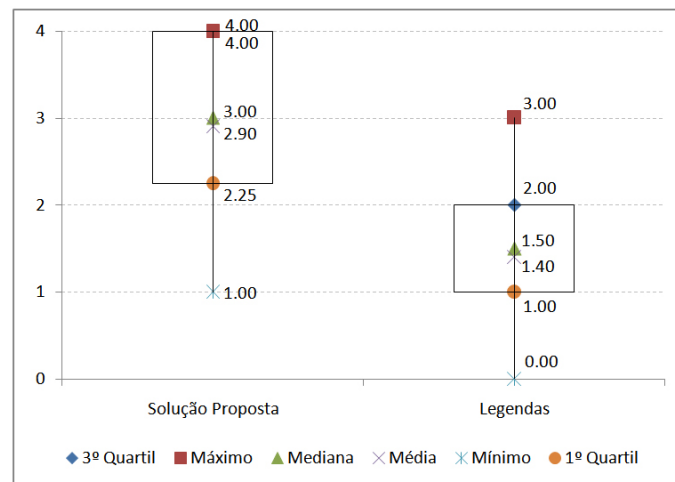


Figura 5.11: Análise de dispersão dos resultados entre as abordagens investigadas.

através do cálculo da mediana, primeiro e terceiro quartil da distribuição ilustrada, que foram 3.0, 2.25 e 4.0, respectivamente.

A dispersão dos resultados obtidos com a utilização das legendas mostra que os surdos tiveram um baixo índice de acertos. Os valores da mediana, primeiro e terceiro quartil (1.5, 1.0 e 2.0, respectivamente) foram inferiores com relação a solução proposta. Outro comportamento importante que foi observado foi que um surdo acertou 75% das questões, e que outros dois não obtiveram nenhum acerto, ou seja, 0% de aproveitamento. Esse último resultado ilustra a dificuldade de alguns surdos em compreender informações veiculadas em textos em língua oral (por exemplo, legenda).

A última parte do questionário investigou aspectos relacionados ao grau de dificuldade para usar um dispositivo móvel e a complexidade de acompanhar a tradução do filme em um dispositivo de segunda tela. Para essas questões, os surdos escolheram um número em uma escala de 1 a 6, onde 1 indica que foi muito fácil, e 6 que foi bastante difícil. Os resultados dessa última parte do questionário são apresentados nos gráficos das Figuras 5.12 e 5.13.

De acordo com a Figura 5.12, observa-se que maior parte dos surdos que participaram dos testes responderam que não houve grandes dificuldades em usar a tecnologia de um dispositivo móvel. Para essa pergunta, a média foi de 3.8 (com desvio padrão de 1.55).

Com relação ao nível de complexidade para acompanhar, simultaneamente, a tradução em Libras gerada pela solução proposta e o filme (isto é, visualizar tanto a tela do

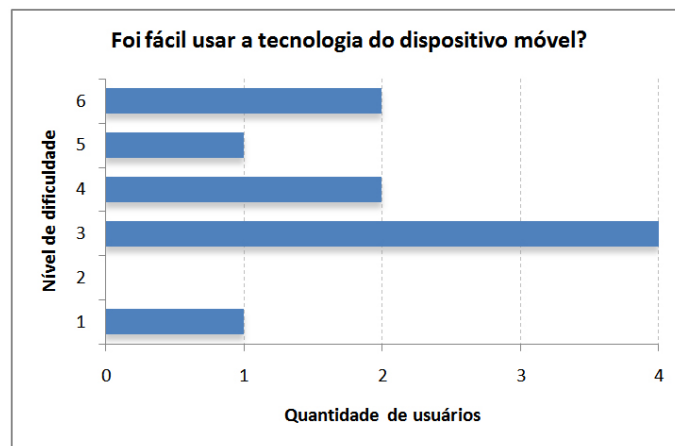


Figura 5.12: Resultado da avaliação do grau de dificuldade em usar um dispositivo móvel.

filme como a do dispositivo móvel), o indicativo é de que existem algumas dificuldades com essa abordagem, conforme o gráfico ilustrado na Figura 5.13. De acordo com as respostas fornecidas pelos surdos, a média de dificuldade foi de 3.1.

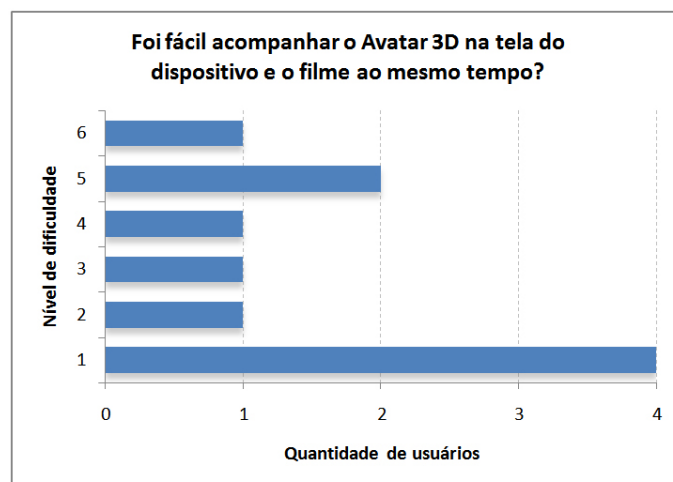


Figura 5.13: Resultado do nível de dificuldade para acompanhar as duas telas ao mesmo tempo.

Por fim, com base nos resultados obtidos nos testes de inteligibilidade, é possível observar que, utilizando a solução proposta, os surdos tiveram um percentual de acerto maior do que os surdos que utilizaram o recurso das legendas. Em virtude disso, é provável que a solução apresentada neste trabalho seja útil para auxiliar usuários surdos (que tenham a Libras como sua língua natural de comunicação) em salas de cinema.

5.3 Avaliação dos Conteúdos de Audiodescrição

Para avaliar geração de Audiodescrição proposta neste trabalho, também foram feitos testes envolvendo deficientes visuais (cegueira total ou parcial). Os testes foram realizados no Instituto dos Cegos da Paraíba, e aconteceram em duas etapas. A primeira etapa foi realizada no dia 31 de Outubro de 2014, e a segunda no dia 3 de Novembro de 2014. No total, 12 voluntários deficientes participaram dos testes, onde, dentre esses, haviam 10 homens e 2 mulheres, com idades entre 18 e 40 anos.

Com relação ao grau de visão dos participantes, 4 eram completamente cegos e 8 tinham cegueira parcial, ou seja, sua capacidade de enxergar era limitada. O nível de escolaridade dos voluntários também foi observado e, de acordo com a classificação, 16,67% ainda não tinham o ensino fundamental completo, 8,33% haviam concluído o ensino fundamental, 25% estavam com o ensino médio em andamento, 8,33% já tinham o ensino médio, 16,67% encontravam-se terminando o ensino superior, e 25% tinham o nível superior de ensino.

Divididos em dois grupos, os usuários foram convidados para assistir dois filmes curta-metragem (um do gênero comédia, com tempo de 5 minutos e 29 segundos, e outro do gênero drama, com 5 minutos e 43 segundos) e um trecho de um longa-metragem (do gênero drama e com 1 minuto e 43 segundos de duração). A ideia foi que um grupo tivesse apenas o áudio original do filme e o outro com o recurso da Audiodescrição gerada pela solução proposta.

A medida em que cada filme era exibido, no final, os usuários tiveram que responder um conjunto de perguntas relacionadas ao contexto do filme apresentado. Com isso, foi possível observar o nível de compreensão dos deficientes visuais com relação aos conteúdos exibidos. O gráfico apresentado na Figura 5.14 mostra os resultados obtidos com esse experimento.

De acordo com a Figura 5.14, os usuários que tiveram acesso ao filme apenas pelo áudio original, isto é, sem o recurso da Audiodescrição, conseguiram 20%, 19,44% e 6,67% de acertos para os vídeos 1, 2 e 3, respectivamente, média de 15,37%. Em contrapartida, o percentual de acerto dos usuários que acompanharam o filme com o auxílio da Audiodescrição gerada pela solução proposta foi superior. De acordo com os

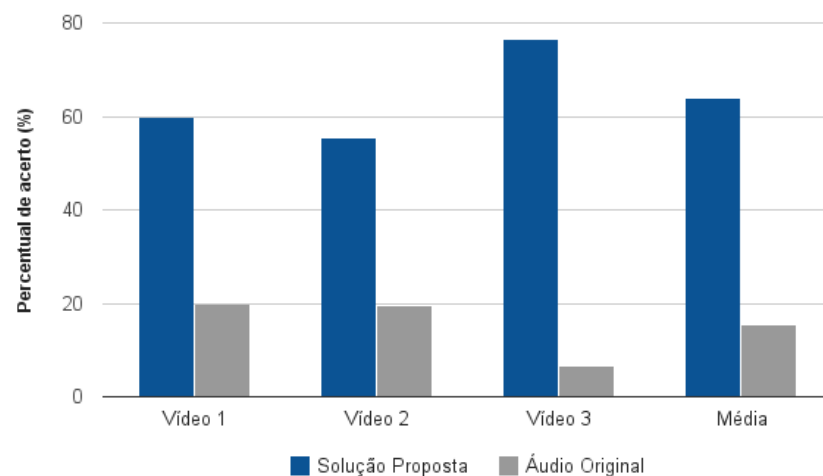


Figura 5.14: Resultado do nível de compreensão dos conteúdos apresentados.

resultados, no primeiro filme foram 60% das perguntas respondidas corretamente, no segundo, foram 55,56% de acertos e, por fim, o terceiro com 76,67% de aproveitamento, média de 64,07%.

Outra questão importante investigada neste experimento foi o nível de dificuldade que os usuários tiveram para entender os filmes com a Audiodescrição gerada pela solução proposta e apenas com o áudio original do filme. O método utilizado para avaliar esse aspecto foi baseado em uma escala numérica de 1 a 6, onde 1 significa que houve muita dificuldade e 6 indica que foi muito fácil. A Figura 5.15 apresenta o resultado obtido nessa investigação.

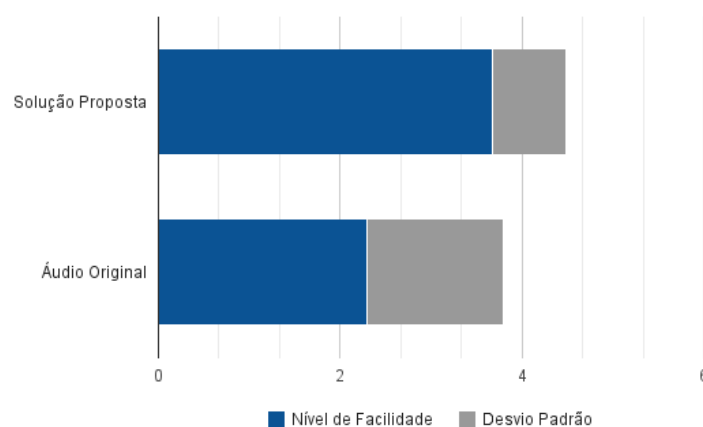


Figura 5.15: Nível de facilidade para entender os filmes utilizados no experimento.

Conforme o gráfico ilustrado na Figura 5.15, os usuários indicaram que, utilizando a Audiodescrição gerada pela solução, foi mais fácil compreender o conteúdo dos filmes exibidos, média de 3,67 (61,17%) e desvio padrão de 0,81. Para o grupo que acompanhou os filmes utilizando apenas o áudio original dos filmes, a média do nível de facilidade foi de 2,3 (38,33%), com desvio padrão de 1,5.

O último ponto abordado nos testes foi com relação aos benefícios que a Audiodescrição gerada por esta solução podem trazer para os deficientes visuais. Então, utilizando a mesma escala numérica de 1 a 6, os usuários indicaram o quanto esses recursos podem ser favoráveis para que pessoas com problemas na visão possam ter acesso às informações disponíveis em ambientes como os cinemas. A Figura 5.16 mostra o gráfico com o resultado dessa análise.

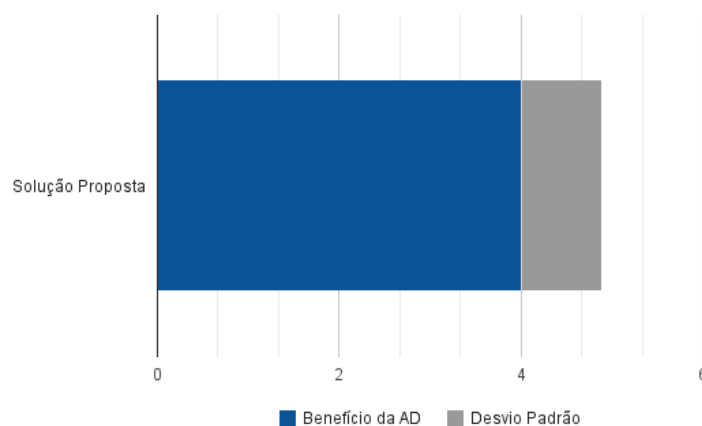


Figura 5.16: Benefícios da Audiodescrição gerada pela solução na compreensão dos filmes.

De acordo com o gráfico ilustrado na Figura 5.16, é possível perceber que os deficientes visuais que participaram desse experimento responderam que a Audiodescrição gerada pela solução proposta é capaz de auxiliar consideravelmente no entendimento dos conteúdos dos filmes. Isso pode ser observado através da média obtida nessa questão, que foi igual a 4 (66,67%) e desvio padrão de 0,89.

Apesar da representatividade dos resultados, é preciso realizar uma análise estatística mais aprofundada. Dessa forma, replicando esses testes com uma quantidade maior de usuários, será possível comprovar a viabilidade em utilizar esta solução como recurso de acessibilidade nas salas de cinema. Além disso, novos testes ajudarão a

identificar possíveis problemas na proposta. A realização de novos testes e análises estatísticas mais aprofundadas é uma das propostas de trabalhos futuros.

Capítulo 6

Considerações Finais

Este trabalho apresentou uma solução tecnológica para viabilizar a participação de pessoas com deficiência auditiva e visual em salas de cinema digital. O objetivo é que trilhas de vídeos acessíveis com janelas de Libras e recursos de Audiodescrição sejam gerados automaticamente a partir dos arquivos de legendas e do roteiro do filme, respectivamente. Então, dessa forma, é possível distribuir esses conteúdos de forma que os usuários possam recebê-los através de dispositivos de segunda tela (e.g., dispositivos móveis).

O desenvolvimento de um protótipo da solução possibilitou a realização de testes computacionais e de testes de inteligibilidade com usuários surdos e usuários cegos. Os testes computacionais, apresentado na Seção 5.1, mostraram o comportamento da ferramenta com relação nível de sincronismo dos conteúdos e o desempenho do tráfego de rede durante a distribuição dos conteúdos. Em relação ao nível sincronismo, os resultados foram considerados satisfatórios, uma vez que a média de atraso foi inferior a 210 milissegundos — média de 3,21 quadros e desvio padrão de 3,08. Dessa forma, considerando o atraso máximo de 4 segundos definido pela NBR-15290 (ABNT, 2005), a solução é consegue gerar conteúdos acessíveis em Libras de maneira eficiente.

Durante a realização do teste computacional, foi utilizada a ferramenta *ntopng* (NTOPNG, 2015) foi utilizada para analisar o tráfego da rede. Os resultados dessa análise mostraram que a taxa média de transmissão não ultrapassou 2,62 Mbit/s nos momentos observados. Mesmo com uma quantidade relativamente pequena de dispositivos, foi possível perceber através dos resultados que o protótipo da solução conseguiu

atender as requisições e distribuir os conteúdos sem grandes problemas.

Foram realizados também alguns testes com usuários surdos e usuários cegos. Os resultados dos testes com os surdos mostraram que, usando a solução apresentada neste trabalho, o nível de compreensão foi superior ao grupo que utilizou apenas o recurso das legendas, isto é, 72,5% de acerto com a solução e 35% de acerto com as legendas. Nos testes com os usuários cegos, os resultados apontaram que a Audiodescrição gerada pelo sistema é capaz de auxiliar na compreensão dos filmes, isto é, 64,07% de acerto usando a solução e 15,37% usando os recursos disponíveis atualmente, ou seja, o áudio original do filme. Esses resultados evidenciam que a solução apresentada neste trabalho é viável para reduzir as barreiras de acesso à informação nos cinemas dos usuários surdos e cegos.

Por fim, mesmo tendo atingido um nível de sincronismo considerável e uma taxa de transmissão de conteúdos razoável, é necessário utilizar uma quantidade maior de dispositivos para realizar avaliações estatísticas mais aprofundadas e, dessa forma, garantir que o desempenho da solução continuará sendo satisfatório, ou então, definir até que ponto pode ser considerada viável. Entretanto, mesmo com os resultados obtidos até o momento, é possível perceber a existência de uma contribuição científica, tecnológica e de grande impacto social.

6.1 Perspectivas de Trabalhos Futuros

Embora este trabalho tenha apresentado uma solução eficaz na geração dos conteúdos acessíveis e eficiente na distribuição desses conteúdos — de acordo com os resultados discutidos no Capítulo 5, existem diversos desafios de pesquisa e propostas de trabalhos que podem ser identificados. Por exemplo, a solução atual utilizou *tablets* e *smartphones* para exibir a tradução em Libras dos conteúdos, porém, presume-se que se esses conteúdos forem apresentados em um *microdisplay* acoplado em óculos especiais, os surdos podem usufruir de um maior conforto.

Outro desafio de pesquisa que pode agregar melhorias à esta solução é a utilização de uma plataforma para dispositivos móveis que ofereça recursos 3D para exibir os sinais em Libras em tempo de execução. Um exemplo de plataforma com essas carac-

terísticas é o Unity¹. Então, dessa forma, ao invés de transmitir conteúdos de vídeo, que naturalmente ocupa boa parte do canal de transmissão, seriam enviados apenas metadados contendo a identificação dos sinais e o tempo de apresentação de cada um deles.

Além das sugestões mencionadas anteriormente, outras propostas de trabalhos que podem ser desenvolvidas posteriormente são:

1. *Aprimorar a estratégia de distribuição dos conteúdos.* Os resultados obtidos nos testes computacionais revelou que a abordagem implementada nesta versão da solução atende ao requisito de atraso máximo permitido de 4 segundos definido pela NBR-15290 (ABNT, 2005). No entanto, é possível melhorar ainda mais o sincronismo e o desempenho da distribuição. Um exemplo que pode ser citado para obter essa melhoria é a utilização do *Dynamic Adaptive Streaming over HTTP* (DASH), também conhecido como MPEG-DASH. O propósito do MPEG-DASH é diminuir o atraso e o *buffering* durante a reprodução de um vídeo; usar *Content Delivery Network* (CDN) baseados em HTTP de custo efetivo, *proxies* e *caches*; lógica de streaming baseada no cliente permitindo maior escalabilidade e flexibilidade; dentre outros.
2. *Realizar novos experimentos com uma quantidade maior de usuários.* Embora o experimento realizado neste trabalho tenha fornecido alguns resultados satisfatórios, é importante ressaltar que se tratando de um público com necessidades especiais, isto é, pessoas surdas e pessoas cegas, é importante que sejam realizados testes exaustivamente com esses usuários, pois, dessa forma, será possível validar uma solução para ser usada em um cenário real, ou seja, nas salas de cinema;
3. *Realizar testes de sincronismo e desempenho com os conteúdos de Audiodescrição.* Em virtude da limitação de tempo, não houve possibilidade de realizar testes de sincronismo e desempenho de distribuição com esta versão da solução. Portanto, como sugestão de trabalho futuro, é recomendado que seja desenvolvida uma

¹<https://unity3d.com>

estratégia que viabilize testar o nível de sincronismo dos conteúdos de áudio gerados pela solução apresentada neste trabalho.

Bibliografia

Aquino Júnior, L. L. et al. A software-based solution for distributing and displaying 3D UHD films. *IEEE Computer Society*, p. 60–68, 2013.

2003. *Uma Metodologia de Projeto de Protocolos de Comunicação Baseada em Técnicas Evolutivas*. [S.l.]: XX Simpósio Brasileiro de Telecomunicações, 2003.

ARAÚJO, T. M. U. *Uma solução para geração automática de trilhas em Língua Brasileira de Sinais em conteúdos multimídia*. Tese (Doutorado) — Universidade Federal do Rio Grande do Norte - UFRN, Rio Grande do Norte, 2012.

ARAÚJO, T. M. U. et al. An approach to generate and embed sign language video tracks into multimedia contents. *Information Sciences - INS 10819*, p. 20, 2014.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 15290: Acessibilidade em comunicação na televisão*. Rio de Janeiro, nov. 2005. 10 p.

BIELER, R. B. *Desenvolvimento Inclusivo: Uma abordagem universal da Deficiência*. [S.l.]: Banco Mundial, out 2005.

BRASIL. *Lei n. 10.098/2000*. 2000. http://www.planalto.gov.br/ccivil_03/leis/110098.htm. Acessado em Dezembro de 2014.

BRASIL. *Lei n. 10.436/2000*. apr 2002. <http://bit.ly/1lFsFq5>.

BRASIL. Ministério de Estado da Educação. *Portaria n 2.678, de 24 de Setembro de 2002*. Brasília, DF, 2002.

BRASIL. Ministério de Estado das Comunicações. *Portaria nº 188, de 24 de março de 2010*. Brasília, DF, 2010.

BRASIL. Presidência da República. *Decreto nº 5.296 de 02 de Dezembro de 2004*. Brasília, DF, 2004.

BRASILMEDIA. *Deficientes Auditivos*. 2012. http://www.brasilmedia.com/tipos-de-deficiencia-auditiva.html#.Ubs0Y_nVBrt. Acessado em Maio de 2013.

BUTTUSSI, F.; CHITTARO, L.; COPPO, M. Using web3d technologies for visualization and search of signs in an international sign language dictionary. *12th International Conference on 3D Web Technology*, p. 61–70, 2007.

2014, João Pessoa, Paraíba. *CineAD: Um Sistema de Geração Automática de Roteiros de Audiodescrição*. [S.l.]: WebMedia2014. 20th Brazilian Symposium on Multimedia and the Web, 2014.

CARVALHO, M. L. B. *Visão subnormal: Orientações ao Professor do Ensino Regular*. 1994.

CINTAS, J. D. Audiovisual translation today: a question of accessibility for all. *Translating Today*, London, n. 4, p. 3–5, 2005.

CISCO. *Cisco Visual Networking Index: Forecast and Methodology, 2013–2018*. USA, 2014.

CONSELHO FEDERAL E REGIONAL DE FONOAUDIOLOGIA. Orientações dos Conselhos de Fonoaudiologia para o Laudo Audiológico. *Audiometria Tonal, Logaudiometria e Medidas de Imitância Acústica*. Brasília, DF, 2009.

2013. *Towards the usage of pauses in audiodescribed videos*. Rio de Janeiro, RJ: 22nd International World Wide Web Conference, 2013.

FEBRABAN. Coleção FEBRABAN de Inclusão Social. *Manual de Acessibilidade para Agências Bancárias*. São Paulo, SP, 2010.

FELIPE, T. A.; MONTEIRO, M. S. *Libras em Contexto: curso básico*. Brasília, DF, 2005. 26 p.

FERREIRA, F. L. S. et al. Providing support for sign languages in middlewares compliant with itu j.202. *Proceddings of the 2011 IEEE International Symposium of Multimedia - ISM*, p. 149–156, 2011.

FIELDING, R. *RFC 2616: Hypertext Transfer Protocol*. [S.l.], jun. 1999. Disponível em: <<http://tools.ietf.org/html/rfc2616>>.

FUSCO, E. *X-libras: Um Ambiente Virtual para a Língua Brasileira de Sinais*. Dissertação (Mestrado) — Centro Universitário Eurípedes de Marília, UNIVEM, Marília, SP, 2004.

GALLAHUE, D. L.; OZMUN, J. C. *Compreendendo o desenvolvimento motor: bebês, crianças, adolescentes e adultos*. 3.ed.. ed. São Paulo - SP: Phorte, 2005.

INSTRUCTIONAL TECHNOLOGY AND EDUCATION OF THE DEAF, 2010, Dublin, Ireland. *Utilizing Lifelike, 3D Animated SigningAvatar Characters for the Instruction of K-12 Deaf Learners*. New York: International Tech Symposium. National Technical Institute of the Deaf - NTID, 2008.

IBGE. *Censo Demográfico 2000*. nov 2000. <http://bit.ly/UY3tkn>. Acessado em: 19 jan. 2014.

KATO, N. et al. Visual communication with dual video transmissions for remote sign language interpretation services. *Dept. of Electron. & Inf. Sci., Tsukuba Coll. of Technol.*, 2010.

2010. *Towards the usage of pauses in audiodescribed videos*. Raleigh, NC: W4A2010, 2010.
- LAKRITZ, J.; SALWAY, A. *The semi-automatic generation of audio description*. [S.l.], 2002.
- LIMA, M. d. S. C. Signing science. *International Society for Technology in Education - ISTE*, U.S. & Canada, v. 32, n. 8, p. 30–33, 2005.
- LIMA, M. d. S. C. Algumas considerações sobre o ensino de português para surdos na escola inclusiva. *Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura Letra Magna*, p. 20–34, 2006.
- MENEZES, D. C.; CAVALCANTE, T. C. F. Compreensão de textos escritos: um estudo com adolescentes surdos. *Distúrb Comun*, p. 29–38, apr 2008.
- MOTTA, L. M. V. M. *Audiodescrição: entrevista com Livia Motta*. 2008. <http://globo.tv.globo.com/rede-globo/programa-do-jo/v/atriz-graziela-pozzobon-faz-audiodescricao/2812698/>. Acessado em Julho de 2014.
- NICOLAIEWSKY, C. A.; CORREA, J. Habilidades cognitivo-linguísticas e segmentação lexical em braille. *Paidéia*, Rio de Janeiro, RJ, v. 19, n. 44, p. 314–348, 2009.
- NIST-SEMATECH. *Engineering Statistics*. 2012. <http://www.itl.nist.gov/div898/handbook/index.htm>. Acessado em: 2 set. 2015.
- NODEJS. *Node.JS on the Road*. 2015. <https://nodejs.org>. Acessado em: 28 ago. 2015.
- NTOPNG. *High-Speed Web-based Traffic Analysis and Flow Collection*. 2015. <http://www.ntop.org/products/traffic-analysis/ntop>. Acessado em: 2 set. 2015.
- ORGANIZATION, W. H. *Visual impairment and blindness*. 2014. <http://www.who.int/mediacentre/factsheets/fs282/en/>. Acessado em Janeiro de 2015.
- PASSADORI, I. *A Importância da Comunicação no Processo de Educação*. 2012. <http://www.passadori.com.br/cursos/palestras/artigos/a-importancia-da-comunicacao-no-processo-de-educacao>. Acessado em Maio de 2013.
- POZZOBON, G. *Entrevista no programa do Jô Soares*. 2010. <http://globo.tv.globo.com/rede-globo/programa-do-jo/v/atriz-graziela-pozzobon-faz-audiodescricao/2812698/>. Acessado em Janeiro de 2015.
- SAN-SEGUNDO, R.; AL, E. Speech to sign language translation system for spanish. *Speech Communication*, p. 1009–1020, 2008.

SILVA, D. A. N. d. S. *Uma Linguagem de Definição Expansível para Língua Brasileira de Sinais*. Dissertação (Mestrado) — PPGI, Programa de Pós-Graduação em Informática, UFPB, 2012.

IRISH HUMAN COMPUTER INTERACTION CONFERENCE'10, 2010, Dublin, Ireland. *HCI for the Deaf community: Developing human-like avatars for sign language synthesis*. Dublin, Ireland: Dublin City University - DCU, 2010.

SONY. *Sony Digital Cinema*. 2012.

Souza Filho, G. L.; ARAÚJO, T. M. U. *GTAAaaS – Acessibilidade como um Serviço (Accessibility as a Service)*. João Pessoa, PB: [s.n.], 2011.

TAMBASCIA, C. d. A. et al. Solução para comunicação e interação com deficientes auditivos em sala de aula. *II ENINED - Encontro Nacional de Informática e Educação ISSN:2175-5876*, p. 115–122, 2012.

UNESCO. Relatório para a UNESCO da Comissão Internacional sobre Educação para o século XXI. *Educação: Um Tesouro a Descobrir*. Brasília, DF, 2010.

WAUTERS, L. N. *Reading comprehension in deaf children: The impact of the mode of acquisition of word meanings*. Tese (Doutorado) — Radboud University, Nijmegen, Holanda, 2005.

WEST, L. L. et al. Display design trade-offs for a wireless mobile captioning system. *Georgia Tech Research Institute*, 2009.

WHATSCINE. *Empowering Theatres*. 2015. <http://www.whatscine.es/accesibilidad.html>. Acessado em Janeiro de 2015.

Apêndice A

Questionário aplicado no teste de inteligibilidade com os surdos

INFORMAÇÃO USUÁRIO:

SEXO: () MASCULINO () FEMININO

VOCÊ É? () SURDO () INTÉRPRETE () INSTRUTOR

IDADE: _____

COMO VOCÊ SURDO? ESCOLHER:

() SURDO NASCENÇA

() TER ALGUÉM FAMÍLIA TAMBÉM SURDO

() SUA FAMÍLIA USAR LIBRAS SUA CASA

() VOCE PARTICIPAR COMUNIDADE SURDO?

() TRABALHA ALGUMA ESCOLA PARA SURDO?

() ESTUDAR ESCOLA PARA SURDOS?

() ESTUDAR ESCOLA JUNTO OUVINTE?

VOCÊ SABER LER? () SIM () NÃO

QUAL SEU GRAU ESCOLARIDADE?

☐ NUNCA ESTUDAR

☐ ENSINO FUNDAMENTAL INCOMPLETO (AINDA NÃO TERMINAR 5º AO 9º ANOS).

☐ ENSINO FUNDAMENTAL COMPLETO (TERMINAR 5º AO 9º ANOS).

☐ ENSINO MÉDIO INCOMPLETO (AINDA NÃO TERMINAR 1º AO 3º ANOS).

☐ ENSINO MÉDIO COMPLETO (TERMINAR 1º AO 3º ANOS).

☐ ENSINO SUPERIOR INCOMPLETO (NÃO TERMINAR UNIVERSIDADE).

☐ ENSINO SUPERIOR COMPLETO (CONCLUIR CURSO UNIVERSIDADE).

☐ OUTRO: _____

ESCOLHER NÍVEL VOCE TER CONHECIMENTO PORTUGUÊS (6 - AVANÇADO, 1 - NÃO SEI NADA)?

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

LIBRAS ESCOLHER NÍVEL VOCE TER CONHECIMENTO LIBRAS (6 - AVANÇADO, 1 - NÃO SEI NADA)?

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

AVALIAR VÍDEO (SINTEL):

1) FÁCIL ENTENDER? (6 - CLARO, 1 - CONFUSO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

2) ACOMPANHA GRAMÁTICA LIBRAS? (6 - PERFEITO, 1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

3) *SINALIZAÇÃO PARECER NATURAL/NORMAL? (6 - PARECER PESSOA, 1 - PARECER ROBÔ)*

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

4) *MULHER TER ANIMAL ESTIMAÇÃO, QUAL?*

☐ GALINHA

☐ GATO

☐ DRAGÃO

☐ SEI-NÃO

5) *O-QUE GUARDIÃO FALAR MULHER?*

☐ FALAR ELA PERTO ENCONTRAR PRÓPRIO ANIMAL.

☐ FALAR ELA PROCURAR IR CAMINHO ERRADO E ESTAR LONGE.

☐ FALAR ELA SER FILHA DELE.

☐ SEI-NÃO

6) *QUAL SER HISTÓRIA LANÇA PERSONAGEM USAR?*

☐ LANÇA TER PASSADO TRISTE. MATAR MUITO INOCENTE.

☐ LANÇA TER GLÓRIA PASSADO. MATAR MUITO BANDIDO.

☐ LANÇA PASSADO PERTENCER GRANDE GUERREIRO HÉRCULES.

☐ SEI-NÃO

7) *O-QUE SIGNIFICAR SINTEL?*

☐ SIGNIFICAR TERRA PERIGOSA.

☐ SIGNIFICAR TERRA FRIO

☐ SIGNIFICAR TERRA DRAGÃO

☐ SEI-NÃO

AVALIAR LIBRAS DISPOSITIVO:

*1) FÁCIL USAR TECNOLOGIA CELULAR OU TABLET? (6 – MUITO FÁCIL);
(1- MUITO DIFÍCIL)*

() 6 () 5 () 4 () 3 () 2 () 1

*2) FÁCIL ACOMPANHAR AVATAR-INTÉRPRETE AO MESMO TEMPO TELA
FILME E CELULAR OU TABLET? (6 – MUITO FÁCIL, 1 - MUITO DIFÍCIL)*

() 6 () 5 () 4 () 3 () 2 () 1

Apêndice B

Artigo Completo Publicado no Webmedia'14

CineLibras: A Proposal for Automatic Generation and Distribution of Windows of Libras on the Cinema Rooms

Leonardo A. Domingues
LAVID/Ci/UFPB
João Pessoa, Brasil
leonardo.araujo@lavid.ufpb.br

Felipe L. S. Ferreira
LAVID/Ci/UFPB
João Pessoa, Brasil
lacet@lavid.ufpb.br

Tiago M. U. de Araújo
LAVID/Ci/UFPB
João Pessoa, Brasil
tiagomaritan@lavid.ufpb.br

Manoel S. Neto
LAVID/Ci/UFPB
João Pessoa, Brasil
manoelgs@lavid.ufpb.br

Lucenildo A. Júnior
LAVID/Ci/UFPB
João Pessoa, Brasil
lucenildo@lavid.ufpb.br

Guido L. de S. Filho
LAVID/Ci/UFPB
João Pessoa, Brasil
guido@lavid.ufpb.br

ABSTRACT

Deaf people face many problems to execute their daily activities. The main reasons to explain this include barriers for both access information as well as communicating with people without disabilities. In this context, the main goal of this paper is to identify the main problems faced by deaf people to access information in movie theaters and to propose a solution to better address their requirements. In this context, it was developed a computational system that is able to automatically generate and distribute accessible video tracks in Brazilian Sign Language (Língua Brasileira de Sinais - LIBRAS) in cinema rooms. This solution uses mobile devices as secondary screens, in a way that deaf people can have access to the content presented in their natural way of communication. Finally, experiments were performed with groups of Brazilian deaf in order to ensure the viability of the proposed solution and the data collected are analyzed and discussed.

Categories and Subject Descriptors

C.2.4 [Distributed Systems]: [Client/server, Distributed applications]; D.2.2 [Design Tools and Techniques]: [User interfaces]; H.5.1 [Multimedia Information Systems]: [Animations, Evaluation/methodology, Video]; K.4.2 [Social Issues]: [Assistive technologies for persons with disabilities]

General Terms

Management, Performance, Languages

Keywords

Cinema Digital, Acessibilidade, Deficiência Auditiva, Aplicações Multimídia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebMedia'14, November 18 - 21 2014, Paraíba, Brazil
Copyright 2014 ACM 978-1-4503-3230-9/14/11 ...\$15.00
<http://dx.doi.org/10.1145/2664551.2664557>.

1. INTRODUÇÃO

Avanços recentes das tecnologias de vídeo digital, assim como o desenvolvimento de câmeras de alta resolução, projetores digitais, modernos padrões de compressão de imagens digitais, redes de computadores de alta velocidade, dispositivos com alta capacidade de processamento e armazenamento, tem direcionado o desenvolvimento da era do cinema digital. Atualmente, é possível capturar, editar, e exibir vídeos digitais de alta definição com ou sem estereoscopia, devido as vantagens inerentes as mídias digitais sobre as analógicas sem perda de qualidade de imagem.

Além de proporcionar conteúdos audiovisuais de qualidade, a digitalização do cinema também tem tornado possível o desenvolvimento e a inclusão de novos serviços e aplicações nos ambientes cinematográficos. Essas aplicações e serviços podem proporcionar novos serviços e benefícios para os espectadores, assim também como expandir as funções do sistema, permitir a participação dos usuários em ambientes inovadores, acesso às novas informações, interação com o conteúdo audiovisual, dentre outros.

Um exemplo dessa nova tendência é o 13th *Street Interactive Movie Experience* [3]. Nesta experiência, os espectadores participam da história e interagem com o protagonista do filme através do seu smartphone. Durante a reprodução do filme, o protagonista inicia automaticamente uma ligação para o espectador na sala do cinema e pede para que ele o ajude a tomar algumas decisões, por exemplo, "Para que direção eu devo ir: direita ou esquerda?". Então, dependendo da decisão do espectador, o filme toma uma trajetória diferente. Dessa forma, a cada nova sessão, o filme pode seguir um roteiro diferente. O processo de interação é feito por meio de um software que efetua chamadas telefônicas para o espectador e permite um diálogo entre ele e o protagonista através de técnicas de reconhecimento de voz.

Outro exemplo importante são os serviços e as tecnologias desenvolvidas para oferecer condições de acesso aos deficientes visuais e auditivos. Por meio desses serviços e tecnologias, é possível gerar conteúdos acessíveis para usuários surdos, por exemplo, e distribuí-los para um display secundário (tablet, smartphone, óculos com microdisplay, dentre outros). Dessa forma, esses usuários tem a possibilidade de acompanhar a tradução do filme em sua língua natural de comunicação, isto é, a língua de sinais.

Na literatura científica, existem alguns trabalhos que fo-

Apêndice C

Prêmio de Melhor Artigo na Trilha Principal do Webmedia'14



Apêndice D

Artigo Completo Publicado no *Journal of Information and Data Management* - JIDM

CineLibras: Automatic Generation and Distribution of Libras Tracks for Digital Cinema Platforms

Leonardo A. Domingues¹, Tiago M. U. de Araújo², Danilo A. N. S. Silva³, Felipe L. S. Ferreira⁴, Guido L. de S. Filho⁵, Fernando T. Paulino⁶

Federal University of Paraíba, Paraíba, Brazil

{leonardo.araujo, tiagomaritan, danilo, lacet, guido}@lavid.ufpb.br and fernandotpaulino@gmail.com

Abstract. Deaf people face many problems to execute their daily activities. These difficulties include barriers for both access information as well as communicating with people without disabilities. In this context, the main goal of this article is to identify the main problems faced by deaf people to access information in movie theaters and to propose a solution to better address their requirements. In this context, a computational system was developed to automatically generate and distribute accessible video tracks in Brazilian Sign Language (Língua Brasileira de Sinais - Libras) in cinema rooms. This solution uses mobile devices as secondary screens, in a way that deaf people can have access to the content presented in their natural way of communication. Finally, experiments were performed with groups of Brazilian deaf in order to ensure the viability of the proposed solution and the data collected are analyzed and discussed.

Categories and Subject Descriptors: C.2.4 [Distributed Systems]: Client/server, Distributed applications; D.2.2 [Design Tools and Techniques]: User interfaces; H.5.1 [Multimedia Information Systems]: Animations, Evaluation/methodology, Video; K.4.2 [Social Issues]: Assistive technologies for persons with disabilities

Keywords: digital cinema, accessibility, hearing impairment, Brazilian Sign Language, multimedia applications

1. INTRODUCTION

Information and Communication Technologies (ICTs) are rarely developed taking into account the specific requirements and needs of deaf people [Haddon and Paul 2001]. They communicate naturally using sign languages, but the support for these languages is rarely addressed in the design of TV, Web and Cinema technologies. When supported, they are generally limited to a window with a sign language human interpreter (wipe) for few contents. In consequence, deaf people have serious difficulties to communicate and access information.

However, the digitization of TV and Cinema, besides the improvement of audiovisual content, is also making possible the development and introduction of new services and applications on these platforms. These applications and services can expand the system functions, allowing users to participate in innovative environments, to access new information, to interact with audiovisual content, among others.

In Cinema, an example is the "13th Street Interactive Movie Experience" [Elkins 2013]. In this experiment, viewers interact with the movie character using their smartphone, taking part in history. More specifically, during the film, a character calls a viewer in the room, and asks him for help to make some decisions, such as "Which way should I follow: left or right?". Then, based on the viewer's

We would like to acknowledge for the support of Rede Nacional de Ensino e Pesquisa (RNP) and Ministério do Planejamento, Orçamento e Gestão (MPOG) of Federal Government of Brazil in partnership with the Núcleo de Pesquisa e Extensão do Laboratório de Aplicações em Vídeo Digital (NPE/LAViD) that helped greatly in developing this work. Copyright©2014 Permission to copy without fee all or part of the material printed in JIDM is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.