

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

UM SISTEMA DE GERAÇÃO AUTOMÁTICA
DE ROTEIROS DE AUDIODESCRIÇÃO

VIRGINIA PINTO CAMPOS

JOÃO PESSOA-PB
Fevereiro-2015

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

**UM SISTEMA DE GERAÇÃO AUTOMÁTICA DE
ROTEIROS DE AUDIODESCRIÇÃO**

VIRGINIA PINTO CAMPOS

**JOÃO PESSOA-PB
Fevereiro-2015**

VIRGINIA PINTO CAMPOS

**UM SISTEMA DE GERAÇÃO AUTOMÁTICA DE
ROTEIROS DE AUDIODESCRIÇÃO**

DISSERTAÇÃO APRESENTADA AO CENTRO DE INFORMÁTICA
DA UNIVERSIDADE FEDERAL DA PARAÍBA, COMO REQUISITO
PARCIAL PARA OBTENÇÃO DO TÍTULO DE MESTRE EM
INFORMÁTICA (SISTEMAS DE COMPUTAÇÃO).

Orientador: Prof. Dr. Tiago Maritan Ugulino de Araújo
Coorientador: Prof. Dr. Guido Lemos de Souza Filho

JOÃO PESSOA-PB
Fevereiro-2015

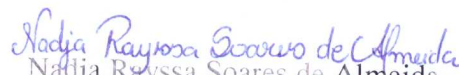
C198u Campos, Virginia Pinto.
Um sistema de geração automática de roteiros de
audiodescrição / Virginia Pinto Campos.-- João Pessoa, 2015.
172f. : il.
Orientador: Tiago Maritan Ugulino de Araújo
Coorientador: Guido Lemos de Souza Filho
Dissertação (Mestrado) – UFPB/CI
1. Informática. 2. Sistema de geração automática.
3. Audiodescrição. 4. Roteiro. 5. Acessibilidade.

UFPB/BC

CDU: 004(043)

Ata da Sessão Pública de Defesa de Dissertação de Mestrado de **Virgínia Pinto Campos**, candidata ao título de Mestre em Informática na Área de Sistemas de Computação, realizada em 27 de fevereiro de 2015.

Aos vinte e sete dias do mês de fevereiro do ano de dois mil e quinze, às dez horas, no laboratório 2 da Escola Superior de Redes - Universidade Federal da Paraíba, reuniram-se os membros da Banca Examinadora constituída para examinar a candidata ao grau de Mestre em Informática, na área de "*Sistemas de Computação*", na linha de pesquisa "*Computação Distribuída*", a **Sra. Virgínia Pinto Campo**. A comissão examinadora foi composta pelos professores doutores: **Tiago Maritan Ugulino de Araújo (PPGI-UFPB)**, orientador e presidente da Banca, **Guido Lemos de Souza Filho (UFPB)**, examinador interno, **Tatiana Aires Tavares (UFPB)**, examinadora interna, **Joana Belarmino de Sousa (UFPB)** externa ao Programa e **Maria da Graça Campos Pimentel (USP)**, externa à Instituição. Dando início aos trabalhos, o professor Tiago Maritan Ugulino de Araújo cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra à candidata para que a mesma fizesse, oralmente, a exposição do trabalho de dissertação intitulado "*Um Sistema de Geração Automática de Roteiros de Autodescrição*". Concluída a exposição, a candidata foi arguida pela Banca Examinadora que emitiu o seguinte parecer: "*aprovada*". Assim sendo, deve a Universidade Federal da Paraíba expedir o respectivo diploma de Mestre em Informática na forma da lei e, para constar, eu, Nadja Rayssa Soares de Almeida, auxiliar em administração, lavrei a presente ata que vai assinada por mim e pelos membros da Banca Examinadora. João Pessoa, 27 de fevereiro de 2015.


Nadja Rayssa Soares de Almeida

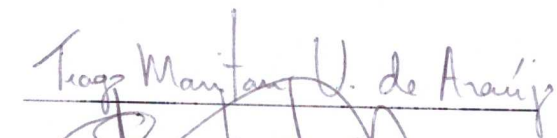
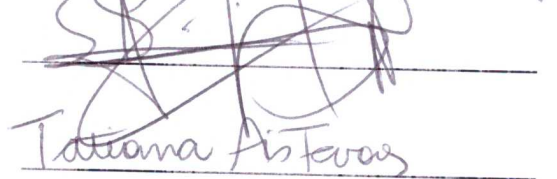
Prof^o Tiago Maritan Ugulino de Araújo
Orientador (PPGI-UFPB)

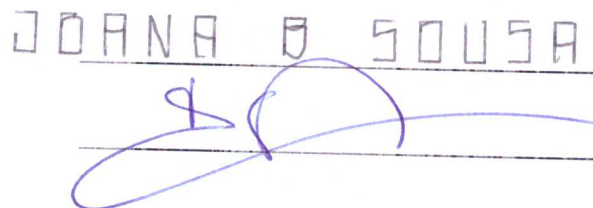
Prof^o Guido Lemos de Souza Filho
Examinador Interno (PPGI-UFPB)

Prof^a Tatiana Aires Tavares
Examinadora (UFPB)

Prof^a Joana Belarmino de Sousa
Examinadora Externa ao Programa (UFPB)

Prof^a Maria da Graça Campos Pimentel
Examinadora Externa (USP)



AGRADECIMENTOS

A Deus, pelas bênçãos recebidas todos os dias, iluminando o meu caminho e dando forças para enfrentar os desafios.

Agradeço a todos os meus familiares pelo apoio incondicional. Em particular, meus pais, Valdemir Campos e Walkíria Freitas, que são uma fonte de inspiração em todos os momentos da minha vida e que tanto me incentivam a batalhar para conquistar os objetivos.

A meu irmão, Bito, pelo carinho, companheirismo e incentivo em todas as minhas jornadas desde sempre.

A Augusto, um agradecimento especial pelo incentivo, paciência, companheirismo e pela transmissão de confiança e força em todos os momentos que passamos juntos.

A meu orientador, Prof. Tiago Maritan, pela disponibilidade, generosidade e orientações ao longo deste trabalho. Um exemplo de profissional e pessoa, sempre disposto a ajudar, cujo apoio foi determinante na elaboração deste trabalho.

A meu coorientador, Prof. Guido, por sua dedicação ao trabalho e pela orientação que foram fundamentais na concepção deste trabalho.

A Leonardo Araújo, Yurika Sato, Felipe Lacet, Manuella Aschoff, Eliza Oliveira, Danilo Assis e tantos outros colegas do LAViD, pelas contribuições que foram fundamentais no desenvolvimento deste trabalho.

Ao Instituto dos Cegos da Paraíba (Valéria Santos), Prof. Joana Belarmino, Liliana Tavares pelo suporte durante a realização da pesquisa.

Por fim, agradeço a todos os amigos que tanto me incentivaram a prosseguir.

RESUMO

A audiodescrição (AD) é um recurso de acessibilidade fundamental para que pessoas cegas ou com baixa visão possam ter acesso ao cinema. Entretanto, este recurso não é utilizado amplamente devido ao alto custo e tempo envolvidos na elaboração da audiodescrição. Este trabalho apresenta uma solução de geração automática de roteiros de audiodescrição de filmes, denominada CineAD. A solução detecta os intervalos entre as falas do filme candidatos a receber a AD, e gera essas descrições a partir da análise do roteiro original e da legenda. De forma alternativa, a solução pode ser incorporada a uma ferramenta de sintetização de voz ou utilizada por um audiodescritor para a criação do áudio com as descrições. Para avaliar a solução, foi desenvolvido um conjunto de experimentos com usuários cegos e audiodescritores. Os resultados mostraram que a solução proposta tem potencial para gerar as descrições dos eventos mais importantes do filme e, desta forma, reduzir as barreiras de acesso ao cinema que as pessoas com deficiência visual enfrentam.

Palavras-chave: Audiodescrição; Roteiro; Acessibilidade.

ABSTRACT

Audio description (AD) is a fundamental accessibility feature for blind or low vision people. However, this feature is not used widely due to the high cost and time involved in its development. This paper presents a solution for automatic generation of audio description, called CineAD. The solution detects gaps in speech and generates these descriptions from the analysis of the original script and subtitles. Alternatively, the solution can be incorporated into a speech synthesis tool or used by a professional. To evaluate the solution, we developed a set of experiments with blind users and audio description specialists. The results show that the proposed solution is able to generate descriptions of the most important events of the film, reducing the barriers of blind people to access cinema.

Keywords: Audio description; Screenplay; Accessibility.

LISTA DE ABREVIATURAS E SIGLAS

ABNT	<i>Associação Brasileira de Normas Técnicas</i>
AD	<i>Audiodescrição</i>
TIC	<i>Tecnologias de Informação e Comunicação</i>

LISTA DE FIGURAS

Figura 1: Estrutura do roteiro e identificação dos elementos (Trecho do roteiro do filme <i>A Cartomante</i> , 2004)	23
Figura 2: Arquitetura da Plataforma (KOBAYASHI et al., 2010)	30
Figura 3: Visão esquemática do sistema <i>CineAD</i>	41
Figura 4: Ambiente de edição do CELTX	42
Figura 5: Detecção de intervalos sem diálogos (<i>gaps</i>) a partir da legenda.....	43
Figura 6: Processo de extração das sentenças candidatas.....	45
Figura 7: Fluxo da geração do roteiro AD.....	47
Figura 8: Exemplo da similaridade entre diálogos do roteiro e legenda (Trecho do roteiro do filme <i>A Cartomante</i> , 2004).....	47
Figura 9: Exemplo de ações entre diálogos do roteiro (Trecho do roteiro do filme <i>A Cartomante</i> , 2004)	48
Figura 10: Exemplo de geração de roteiro AD (Trecho do filme <i>3 Minutos</i> , 1999).....	50
Figura 11: Experimento com usuários	55
Figura 12: Grau de escolaridade dos usuários da avaliação	56
Figura 13: Resultados dos testes do nível de compreensão dos conteúdos	56
Figura 14: Gráfico <i>Box Plot</i> (caixa) dos testes do nível de compreensão dos conteúdos.....	57

LISTA DE QUADROS

Quadro 01: Trecho do roteiro de AD do filme <i>Águas de Romanza</i>	21
---	----

LISTA DE TABELAS

Tabela 01: Classificação dos artigos	25
Tabela 02: Classificação de cada trabalho em relação às categorias..	27
Tabela 03: Classificação do CineAD em relação às categorias.....	38
Tabela 04: Peso das Lista de Palavras Frequentes.....	45
Tabela 05: Características dos vídeos de teste.....	52
Tabela 06: Dados da geração do roteiro AD.....	53
Tabela 07: Resultado do teste-t para a Hipótese.....	58
Tabela 08: Resultados da avaliação.	59
Tabela 09: Dados gerais sobre os participantes.	60
Tabela 10: Resultados da avaliação com especialistas (Valores na escala de 1 a 6).....	61

SUMÁRIO

1 INTRODUÇÃO	13
1.1 JUSTIFICATIVA	13
1.2 OBJETIVOS	16
1.2.1 Geral	16
1.2.2 Específicos	16
1.3 METODOLOGIA	16
1.4 ESTRUTURA DA DISSERTAÇÃO	17
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 AUDIODESCRIÇÃO	18
2.2 AUDIODESCRIÇÃO NO CINEMA	19
2.2.3 Roteiro de Audiodescrição	20
2.2.1 Roteiros Cinematográficos	22
2.3 CONSIDERAÇÕES FINAIS	23
3 TRABALHOS RELACIONADOS	24
3.1 REVISÃO SISTEMÁTICA DA LITERATURA	24
3.2 TOWARDS THE USAGE OF PAUSES IN AUDIO-DESCRIBED VIDEOS	28
3.3 DESCRIBING ONLINE VIDEOS WITH TEXT-TO-SPEECH NARRATION	29
3.4 ACCESSIBLE VIDEODESCRIPTION ON-DEMAND	30
3.5 ARE SYNTHESIZED VIDEO DESCRIPTIONS ACCEPTABLE?	32
3.6 TOWARDS COMPUTER-VISION SOFTWARE TOOLS TO INCREASE PRODUCTION AND ACCESSIBILITY OF VIDEO DESCRIPTION FOR PEOPLE WITH VISION LOSS	33
3.7 THE SEMI-AUTOMATIC GENERATION OF AUDIO DESCRIPTION FROM SCREENPLAYS	36
3.8 CONSIDERAÇÕES FINAIS	37
4 SISTEMA PROPOSTO - CINEAD	40
4.1 FUNCIONAMENTO	40
4.2 <i>PARSER</i> DE ROTEIROS	41
4.3 IDENTIFICAÇÃO DE <i>GAPS</i>	42
4.4 SUMARIZAÇÃO	43
4.5 GERAÇÃO DE ROTEIRO DE AUDIODESCRIÇÃO	46
4.5.1 Verificação de Similaridade	47
4.5.2 Seleção de ações entre diálogos	48
4.5.3 Divisão de Gaps	48
4.5.4 Geração do roteiro de audiodescrição	49
4.6 CONSIDERAÇÕES FINAIS	51
5 EXPERIMENTOS	52
5.1 ANÁLISE TÉCNICA DOS ROTEIROS	52
5.2 AVALIAÇÃO DO NÍVEL DE COMPREENSÃO DOS CONTEÚDOS	54
5.3 AVALIAÇÃO COM ESPECIALISTAS	60
5.4 CONSIDERAÇÕES FINAIS	63
6 CONCLUSÃO	64
6.1 TRABALHOS FUTUROS	65
REFERÊNCIAS BIBLIOGRÁFICAS	66

ANEXO A - PROTOCOLO DE REVISÃO SISTEMÁTICA	70
APENDICE A – LISTA DE PALAVRAS FREQUENTES	75
APENDICE B – PLANEJAMENTO DOS EXPERIMENTOS	76
APENDICE C - QUESTIONÁRIO DE AVALIAÇÃO DOS CONTEÚDOS ACESSÍVEIS.....	83
APENDICE D - QUESTIONÁRIO DE AVALIAÇÃO COM AUDIODESCRITORES....	87

1 INTRODUÇÃO

Uma grande carga de informação é veiculada através de meios audiovisuais. Em uma sociedade que apresenta um aumento no uso desses conteúdos como fonte de informação e entretenimento, as pessoas cegas ou com baixa visão podem ser excluídas por enfrentarem barreiras de acesso a esses conteúdos. Para reduzir essas barreiras, um recurso de acessibilidade bastante útil para as pessoas cegas é a audiodescrição (AD).

Este trabalho apresenta uma solução para geração automática de roteiros de audiodescrição com o objetivo de melhorar o acesso de pessoas com deficiência visual aos filmes cinematográficos, e desta forma, reduzir as barreiras de acesso à esse tipo de informação visual.

Este capítulo apresenta a motivação deste trabalho na Seção 1.1, detalha os objetivos gerais e específicos na Seção 1.2, a metodologia que foi utilizada na Seção 1.3 e, por fim, na Seção 1.4 é apresentada a organização deste documento.

1.1 JUSTIFICATIVA

Pessoas que não conseguem processar conteúdos visuais, ou que fazem isto de forma limitada, usualmente tem acesso restrito a alguns tipos de informações como, por exemplo, expressões gestuais, descrição de cenários e personagens, entre outros elementos, que são essenciais em experiências audiovisuais como o cinema, teatro, televisão e óperas. Essas pessoas, em muitas situações, acabam ficando excluídas do contexto sociocultural e intelectual da sociedade, uma vez que não são disponibilizados recursos de acessibilidade que lhes garantam a participação e o consumo desses conteúdos.

Para garantir a acessibilidade a conteúdos, a preocupação não deve estar apenas na criação da informação. É preciso trabalhar a forma de apresentação, visualização e interação com o usuário. As dificuldades e limitações do usuário devem ser estudadas, a fim de propiciar meios de acesso à informação.

As pessoas com deficiência visual, tema central a ser explorado neste trabalho, possuem dificuldades para captação e compreensão da informação, além de limitações nos processos de comunicação e mobilidade. A deficiência visual caracteriza-se pelas debilidades na percepção da presença da luz e na sensação da forma, do tamanho e da cor de um estímulo visual (TORRES; MAZZONI; MELLO, 2007).

Segundo o censo demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2010, existem cerca de 35,7 milhões de pessoas com algum nível de deficiência visual no Brasil, o que representa cerca de 18,8% da população brasileira (IBGE, 2010). Dentre as deficiências investigadas no censo, a deficiência visual apresenta a maior incidência na população e são essas pessoas que, devido a ausência de percepção visual, enfrentam barreiras no acesso às informações visuais, como os filmes. Além do aspecto social, este cenário representa também uma grande área de mercado a ser explorado pela indústria audiovisual.

Diante desta situação, a audiodescrição se apresenta como um meio de fornecer informações adicionais ao público. A audiodescrição é um recurso de acessibilidade desenvolvido para melhorar o acesso de pessoas, cegas ou não, à essas informações visuais, permitindo que elas participem de experiências intrinsecamente visuais, como o cinema, através de descrições de imagens, narração de ações e detalhes visuais como figurinos, indicação de tempo e espaço.

A audiodescrição surgiu nos Estados Unidos em 1975 e hoje é amplamente utilizada em diversos países, como Espanha, França, Inglaterra e Japão. Em alguns deles, existem normas e regulamentações que formalizam a criação e a exibição obrigatória da audiodescrição na televisão (FRANCO; SILVA, 2010).

No Brasil, a primeira utilização pública da audiodescrição foi em 2003 no Festival Internacional de Cinema “Assim Vivemos”, voltado para o público deficiente. Em 2011, a audiodescrição passou a ser obrigatória nas emissoras de televisão brasileiras com transmissão digital, de acordo com a Lei Nº 10.098, de 19 de dezembro de 2000 (BRASIL, 2000). Foi instituído pela Portaria 188/2010 que, atualmente, as emissoras devem incluir 4 horas de audiodescrição na programação semanal (BRASIL, 2010).

O objetivo da audiodescrição de filmes é a tradução dos elementos visuais que é inserida entre os diálogos, através de uma narração, de forma que não haja interferência nos efeitos sonoros do áudio original (BENECKE, 2004). Essa tradução transforma toda a informação visual em palavras, ou seja, audiodescreve tudo o que é visto.

Para audiodescrever os filmes, os roteiros de audiodescrição são geralmente produzidos de forma manual, o que acaba demandando muito tempo (LAKRITZ; SALWAY, 2002). Segundo Lakritz e Salway (2002) é necessário, em média, cerca de 60 horas para um profissional descrever um filme com 2 horas de duração. Ademais, segundo pesquisas com empresas que prestam o serviço de audiodescrição, o custo para a criação de um roteiro de AD é de R\$ 40,00 por minuto de vídeo. No entanto, diversos filmes brasileiros

ainda não contam com este recurso de acessibilidade, e, em consequência disso, as pessoas cegas ou com baixa visão acabam não tendo acesso a esse tipo de informação cultural.

Para reduzir esses problemas, uma solução automatizada para gerar roteiros de AD pode auxiliar nesta tarefa, tornando o processo de descrição mais eficiente, diminuindo o tempo investido e o custo associado. Dessa forma, a geração automática surge como uma importante alternativa, tanto para a criação de audiodescrição de filmes, como para o auxílio aos descritores humanos durante a produção manual de uma descrição.

Alguns trabalhos vem sendo desenvolvido relacionados ao uso computacional dentro do processo de criação e utilização de audiodescrição. Os trabalhos aplicam a audiodescrição em domínios diversos como na descrição de fórmulas e dados matemáticos (MURPHY, BATES FITZPATRICK, 2010; MOSKOVITC, WALKER, 2010), ambientes 3D (OKTAY, FOLMER, 2011; DULYAN, EDMONDS, 2010), dados estatísticos (THOMAN, SRIPADA, NOORDZIJ, 2010) e textos (DEMIR, OLIVER, SCHWARTZ, 2010; WANG, LI, 2012; FERRES, SEPULVEDA, 2011).

No caso da audiodescrição específica para filmes, alguns trabalhos abordam o uso de sintetizadores de voz na narração (O'CONNELL *et al*, 2010; CHAPDELAINE, GAGNON, 2009; KOBAYASHI *et al.*, 2010), a apresentação de diferentes níveis de detalhamento de audiodescrição com versões resumidas e estendidas (ENCELLE, BELDAME, PRIÉ, 2013; CHAPDELAINE, GAGNON, 2009) e o uso de players próprios para facilitar a adição e consumo da audiodescrição (KOBAYASHI *et al.*, 2010; CHAPDELAINE, GAGNON, 2009). Quanto à geração automática das descrições, pesquisas, em estágios iniciais, estão sendo feitas com o uso de técnicas de reconhecimento de elementos do vídeo (GAGNON *et al.*, 2009) e de análise de roteiro cinematográfico sem vínculos com o tempo do vídeo (LAKRITZ; SALWAY, 2002). Porém estas apresentam limitações que não ainda não atingem as necessidades de automatização da AD de forma satisfatória. Diante disso, a geração automática do roteiro de audiodescrição de filmes, especialmente para filmes brasileiros, apresenta-se como uma área pouco explorada.

Dentro deste contexto, este trabalho apresenta uma proposta de solução para a geração da audiodescrição de filmes na língua portuguesa, e como consequência, para facilitar o acesso das pessoas com limitações visuais às produções audiovisuais cinematográficas brasileiras.

1.2 OBJETIVOS

1.2.1 Geral

O objetivo geral deste trabalho é apresentar um sistema de geração automática de roteiros de audiodescrição - CineAD, voltado para filmes brasileiros, em língua portuguesa, tornando-os acessíveis às pessoas cegas ou com baixa visão. Por meio da análise do roteiro e da legenda, as descrições são elaboradas de forma vinculada ao tempo do filme. O sistema gera o roteiro contendo as descrições das ações do filme, o qual, por sua vez, pode ser utilizado por narrador humano ou software de sintetização de voz na criação da audiodescrição.

1.2.2 Específicos

Para alcançar o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

1. Fazer um levantamento bibliográfico detalhado sobre as principais soluções relacionadas a geração automática/semiautomática de audiodescrição de filmes;
2. Investigar alternativas para audiodescrever filmes;
3. Projetar uma estratégia de geração automática de roteiro audiodescrição de filmes a partir da análise do roteiro e legenda originais;
4. Implementação da solução proposta e a aplicação em filmes brasileiros;
5. Realização de um experimento prático quantitativo com usuários cegos brasileiros para analisar e testar a solução.
6. Realização de uma pesquisa qualitativa com audiodescriptores para analisar a qualidade do roteiro gerado pelo sistema.

1.3 METODOLOGIA

O desenvolvimento deste trabalho iniciou-se com o levantamento das informações na literatura sobre a audiodescrição de filmes e o uso computacional no processo de criação e apresentação das descrições. Essas informações apontaram uma lacuna na geração dos

roteiros de audiodescrição e serviram como ponto de partida na elaboração de um sistema que realiza a geração de forma automática.

Em seguida, foi desenvolvido o sistema CineAD com uma abordagem de geração automática de roteiros de audiodescrição de filmes brasileiros, em língua portuguesa, através da análise da legenda e do roteiro cinematográfico.

A avaliação do sistema consistiu de experimentos de caráter quantitativo e qualitativo. A primeira parte dos experimentos avaliou o entendimento dos conteúdos visuais com a audiodescrição gerada pelo sistema, diretamente com pessoas cegas ou deficientes visuais, realizada no Instituto dos Cegos da Paraíba. Foram aplicados questionários para quantificar a melhoria no entendimento do filme com o uso da audiodescrição do CineAD em comparação com o filme sem a audiodescrição.

Na segunda parte dos experimentos, foi realizada uma pesquisa de caráter qualitativo com o objetivo de coletar dados e estudar a qualidade das descrições geradas pelo sistema. Para isso, foi utilizado o método de entrevistas, feito com especialistas em audiodescrição.

1.4 ESTRUTURA DA DISSERTAÇÃO

O restante deste documento está organizado em seis capítulos. A seguir serão apresentados um sumário de cada um deles.

Capítulo 2: Fundamentação Teórica. Contextualiza a audiodescrição, apresenta definições e características sobre a audiodescrição no cinema e apresenta detalhes sobre roteiros de filmes e roteiros de AD.

Capítulo 3: Trabalhos Relacionados. Apresenta trabalhos relacionados a audiodescrição e ao uso computacional no processo de geração e apresentação da audiodescrição.

Capítulo 4: Sistema Proposto - CineAD. Apresenta uma descrição detalhada da solução proposta neste trabalho, detalha seus principais componentes e as estratégias utilizadas no seu desenvolvimento.

Capítulo 5: Experimentos. Descreve detalhes sobre os experimentos de avaliação do sistema CineAD realizados com usuários cegos e especialistas, e apresenta uma análise dos resultados obtidos nos testes.

Capítulo 6: Conclusão. Apresenta conclusões sobre o trabalho e uma discussão sobre os possíveis trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma visão geral da audiodescrição e áreas correlatas que englobam aspectos tratados na criação e apresentação da audiodescrição de filmes.

Na Seção 2.1, o conceito e as principais características da audiodescrição são apresentados. Por sua vez, a Seção 2.2 descreve a audiodescrição específica para o cinema, além de detalhar aspectos do roteiro que são relevantes para a criação da AD.

2.1 AUDIODESCRIÇÃO

Em uma sociedade justa e democrática, as pessoas devem ter as mesmas oportunidades de acesso à informação. Segundo a ABNT – Associação Brasileira de Normas Técnicas (ABNT, 1994), o termo acessibilidade refere-se à possibilidade de qualquer indivíduo desfrutar dos benefícios da vida em sociedade, incluindo-se, dentre estes, a Internet e o acesso a conteúdos digitais, independente de suas capacidades físico-motoras, perceptivas ou sociais.

As atividades como cinema, teatro, televisão, espetáculos de dança e ópera representam situações potencialmente positivas à integração dos indivíduos. Partindo destes conceitos, surge a necessidade de, levando-se em conta as mais diversas características e limitações que as pessoas podem possuir, criar e desenvolver meios que garantam a acessibilidade à informação que sejam, ao mesmo tempo, atrativos e fáceis de utilizar.

Pessoas que apresentam limitações no sentido da visão enfrentam, consequentemente, barreiras para participar de eventos ou experiências que exijam tal sentido. O indivíduo possui deficiência visual quando, de forma parcial ou total, não detém a capacidade funcional da visão. Segundo Flor (2009), o termo está referido a um problema permanente, congênito¹ ou adquirido, que não pode ser reparado por tratamentos clínicos ou cirúrgicos. As pessoas que estão neste grupo de usuários apresentam os seguintes problemas: cegueira total, quando objetos ou luminosidade são imperceptíveis; baixa visão, quando a acuidade visual é menor que a de pessoas com visão normal; e daltonismo, quando não é possível distinguir combinações ou pares de cores (DIAS, 2007) (QUEVEDO & ULBRICHT, 2011).

¹ Congênito: (adj.) inato; que nasce com o indivíduo; natural, nato.

Desta forma, a audiodescrição (AD) apresenta-se como recurso de acessibilidade que permite às pessoas que possuem algum tipo de deficiência visual ter acesso e usufruir de produções audiovisuais. Segundo Motta (2014), a audiodescrição é um recurso para pessoas com deficiência visual que facilita o entendimento de obras audiovisuais, transformando tudo que é visto no que é ouvido. É uma tradução que converte toda a informação visual em palavras, ou seja, audiodescreve tudo o que é apresentado visualmente. Diferentemente de outras tecnologias que promovem a acessibilidade, a AD não é um recurso a ser utilizado isoladamente, mas sim, vinculado ao produto visual (NUNES *et al.*, 2011).

A AD surgiu na década de 1970. Desde então, alguns países estabeleceram normas que determinam o modelo e o foco da audiodescrição, cada um com critérios específicos para atender a necessidade das pessoas com deficiência visual. Nos Estados Unidos, o modelo enfatiza a ambientação visual resultando em uma descrição mais detalhada. Já na Espanha, o foco da AD está nas ações, com vocabulário mais simples e frases curtas (ALVES *et al.*, 2011).

De forma geral, a audiodescrição deve descrever os elementos fundamentais para o entendimento da obra audiovisual. De acordo com CASADO (*apud* BRAGA, 2011), os elementos a serem descritos se dividem em: visuais não verbais e visuais verbais. Os visuais não verbais são os personagens, figurinos, características físicas, expressões gestuais, ambientação, localização espacial dos personagens, elementos temporais como hora do dia, ações, etc. Os visuais verbais são títulos, legendas e elementos textuais que fazem parte da história como, por exemplo, os nome de ruas.

2.2 AUDIODESCRIÇÃO NO CINEMA

A audiodescrição pode ser aplicada às várias linguagens artísticas como forma de tornar um evento acessível para pessoas com limitações visuais. Contudo, este trabalho tem como escopo a audiodescrição para cinema.

De acordo com Benecke (2004), a AD de filmes é a tradução das imagens, do enredo, do cenário e da ação, que é inserida entre os diálogos dos filmes, através de uma narração, de forma que não haja interferência nos efeitos sonoros do áudio original.

O processo de produção da audiodescrição para filmes engloba um conjunto de etapas em comum e é aplicável em longas, médias, curtas metragem e seriados. Santana (2010) destaca as seguintes etapas:

- **Decupagem para AD:** O filme é assistido na íntegra, os personagens são elencados (nomes e características), são detectadas as informações sobre quando e onde acontece a história e, por fim, ocorre o mapeamento da informação que é indispensável na audiodescrição. No Brasil, ainda não há modelo ou normas que determinem quais os elementos que devem ser descritos (NUNES *et al.*, 2011). Atualmente, os eventos descritos são bastante diversificados, dependendo da história, do tempo disponível e da necessidade do filme. Os movimentos, ações e expressões dos personagens são elementos que comumente fazem parte das descrições.
- **Marcação de Cena:** Os espaços entre os diálogos, silêncios e pontos importantes para inserção de AD são detectados. Aqui podem ser utilizadas as “deixas”, que são as últimas falas antes de iniciar a audiodescrição.
- **Roteiro de Audiodescrição:** O texto da audiodescrição é criado cena por cena, a partir das informações coletadas nas etapas anteriores.
- **Gravação da AD:** Seleciona-se a voz que reproduzirá a narração. Essa voz deve ser clara, agradável e não pode ser confundida com outras vozes apresentadas no filme. Após a escolha da voz, as falas da audiodescrição são gravadas, geralmente em estúdio, com a presença de um diretor.
- **Tratamento, Mixagem e Finalização:** O áudio é tratado para eliminar qualquer ruído que possa ter sido captado durante a gravação, e é misturada (mixada) com o som original do filme.

O presente trabalho trata, especificamente, das três primeiras etapas supracitadas (decupagem para AD, marcação de cena e roteiro de audiodescrição). O sistema descrito no Capítulo 4 apresenta uma solução para a geração automática de roteiros de audiodescrição para filmes.

2.2.1 Roteiro de Audiodescrição

Segundo Nunes *et al.* (2011), para se fazer a audiodescrição é preciso elaborar um roteiro adaptado ao tipo de conteúdo visual que se deseja audiodescrever. No caso de filmes, faz-se necessária a criação prévia de um roteiro de audiodescrição que, em seguida, será narrado. O audiodescritor roteirista é o responsável por criar um roteiro dotado de precisão,

com as devidas inserções de tempo e orientações para leitura, apresentando todas as informações relevantes para a compreensão do filme, ao mesmo tempo em que procura eliminar descrições desnecessárias. Importante ressaltar que, no processo de descrição, um número mínimo de palavras deve ser utilizado, evitando-se, desta forma, o excesso de informações no áudio.

Em seguida, a partir das orientações do roteiro, um audiodescritor realiza a gravação da voz. Um exemplo de roteiro AD é apresentado em Araújo (2010). O trabalho expõe um trecho do roteiro de audiodescrição para o filme *Águas de Romanza* (ver Quadro 1). O roteiro contém os tempos iniciais e finais (*time code*) das inserções da AD, as descrições textuais, as deixas (últimas falas antes de iniciar a audiodescrição) e as rubricas, que são orientações para a narração.

TIME-CODE	AUDIODESCRIÇÃO
00:04:22:03 - -> 00:04:25:20	– Mas Vó, como é que é feita a chuva? A avó enxuga as lágrimas no vestido. → DEIXA
00:04:34:23 - -> 00:04:36:20	– Você não lembra? Romanza balança a cabeça negativamente. → DEIXA
00:05:35:22 - -> 00:05:37:11	–... só esperando o sol, pra fazer eles brilhar. Os olhos verdes de Romanza sorriem. → DEIXA [Rápido] [Falar assim que aparece o rosto de Romanza] → RUBRICA
00:06:11:22 - -> 00:06:13:21	– Você vai ver só. A avó olha para São José. Lá fora... → Deixa

Quadro 1: Trecho do roteiro de AD do filme *Águas de Romanza* (ARAÚJO, 2010).

No presente trabalho, a solução apresentada utiliza uma abordagem na qual a geração automática de roteiros de AD tem o roteiro cinematográfico como base para a elaboração das descrições. Através da análise do roteiro original do filme, as informações sobre os eventos principais da história são selecionadas para formar o roteiro de audiodescrição. Desta forma, é preciso entender os elementos e o tipo de informação que está contida dentro desses roteiros cinematográficos, como também a sua relevância dentro do processo de geração.

2.2.2 Roteiros Cinematográficos

As descrições da AD devem possuir uma linguagem adequada e clara, utilizando-se palavras capazes de determinar de forma concreta os objetos, personagens e cenários, o que, por sua vez, evidencia a importância da utilização de um vocabulário que traduza as características, gestos e ações do filme de forma objetiva (BRAGA, 2011). Diante disso, o roteiro original do filme tem função essencial na etapa de criação da audiodescrição, pois nele estão contidas as palavras e termos apropriados a linguagem do filme. Segundo Payá (*apud* BRAGA, 2011), o audiodescritor deve escolher as melhores palavras para descrever os eventos e, para isso, tem o roteiro do filme como uma importante ferramenta de tradução.

De acordo com Field (2001), um roteiro cinematográfico é uma história contada com imagens. O roteiro é formado por elementos básicos dentro de sua estrutura como, por exemplo, pontos de viradas, cenas, diálogos, personagens, ações, dentre outros. Todas as informações sobre o filme, desde pontos de vista até diálogos de personagens, são apresentadas no roteiro, o qual serve como base para as filmagens. No que diz respeito à audiodescrição, o roteiro cinematográfico contém muitas informações com a descrição mais completa do filme, porém como a AD está limitada aos espaços sem diálogos e de silêncio, ela não comporta a totalidade destas descrições. Por esta razão, é necessário que se realize a seleção das informações mais relevantes para a audiodescrição.

Em ALVES *et al.* (2011), foi feita uma pesquisa com participantes com deficiência visual sobre os modelos de AD aplicados em outros países. Foi identificado que há uma maior eficácia quando se utiliza um modelo de audiodescrição baseado na ação, como na norma espanhola, em detrimento da versão mais extensa com descrição dos objetos e detalhes do cenário, pois a AD muito detalhada pode causar confusões nos usuários. Diante disso, neste trabalho, o roteiro de audiodescrição gerado pela solução utiliza unicamente as ações.

Uma ação é uma atividade ou fala gerada por algum elemento de estória, como personagens, incidentes, situações em que se encontram, objetos, entre outros (CAMPOS, 2009). Para exemplificar, seguem algumas ações extraídas do roteiro cinematográfico do filme *Cidade de Deus*:

- Cabeção e Touro caminham lado a lado, rumo ao que parece ser um encontro inesperado e inevitável com Alicate.

- Cabeleira tira a camiseta enquanto olha apaixonado para Berenice.
- A mulher do Ceará vai até o carrinho comprar peixe.

Por sua vez, *softwares* editores de roteiros permitem a identificação de cada elemento do roteiro. Com isso, cada componente escrito possui um rótulo específico de acordo com sua função dentro do roteiro. Os rótulos mais utilizados são: título de cena, ação, personagem, diálogo, planos, transição entre cenas e etc. A Figura 2 apresenta um exemplo da estrutura do roteiro e a identificação dos rótulos de cada elemento. É importante ressaltar que os elementos contidos no roteiro original do filme com o rótulo “ação” constituem os pontos centrais do roteiro de AD gerado pela solução descrita neste trabalho.

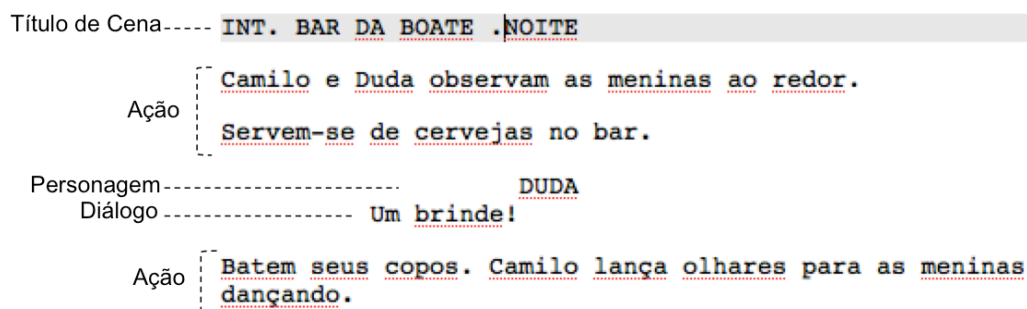


Figura 1: Estrutura do roteiro e identificação dos elementos (Trecho do roteiro do filme *A Cartomante*, 2004)

2.3 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentada uma visão geral da audiodescrição e algumas características específicas para a sua aplicação no cinema. A audiodescrição é um recurso fundamental para tornar produtos audiovisuais acessíveis às pessoas com qualquer deficiência visual. Especificamente em filmes, o audiodescritor precisa conhecer bem a obra para escolher, da melhor maneira possível, os eventos relevantes ao entendimento do enredo e que serão descritos. Por esta razão, o roteiro torna-se uma ferramenta fundamental para o audiodescritor buscar palavras e termos que farão parte da AD e que melhor descrevam as ações.

3 TRABALHOS RELACIONADOS

Atualmente, o processo de criação de roteiros AD é uma tarefa que demanda muito tempo, principalmente pelo fato de ser, predominantemente, feito de forma manual (LAKRITZ;SALWAY, 2002). O uso de tecnologia pode auxiliar significativamente na inserção e elaboração da audiodescrição em filmes. Em razão disso, alguns estudos vem sendo desenvolvidos para incorporar técnicas computacionais nos processos envolvidos. Neste capítulo é apresentada uma revisão de literatura sobre o desenvolvimento de técnicas de geração automática e semiautomática de audiodescrição.

3.1 REVISÃO SISTEMÁTICA DA LITERATURA

Uma revisão sistemática da literatura utilizada para a geração automática e semiautomática de audiodescrição, o tema de pesquisa foco deste trabalho, é apresentada em Muniz (2014). O processo de revisão em Muniz (2014) foi conduzido por três pesquisadores: Raul Correa Muniz (autor do trabalho Muniz (2014)), a autora deste trabalho e Dr. Tiago Maritan Ugulino de Araújo (professor orientador dos trabalhos). A participação de mais de um pesquisador permitiu que as fases do processo que necessitam de subjetividade, como a seleção, classificação e avaliação dos artigos fossem conduzidas de forma mais imparcial, o que torna os resultados mais confiáveis. O protocolo da revisão sistemática é apresentado no Anexo A.

Na revisão foram selecionados 14 artigos que apresentavam relevância em relação ao tema. Eles passaram por um processo de classificação que consistiu da leitura e da classificação de acordo com critérios definidos no protocolo de revisão. Este processo foi realizado por dois revisores e, em caso de divergência na classificação, o terceiro atuaria como árbitro, desempatando a decisão sobre a classificação correta do trabalho. O resultado da classificação dos trabalhos é apresentado na Tabela 1.

Tabela 1: Classificação dos artigos

Categoria	Sub categoria	Número de artigos	Porcentagem
Tipo de entrada	Texto	3	21,42%
	Ambiente 3D	2	14,28%
	Dados estatísticos	1	7,14%
	Vídeo	2	14,28%
	Roteiro	2	14,28%
	Dados matemáticos	2	14,28%
	Não classificado	2	14,28%
Tipo de saída	Texto	5	35,71%
	Áudio	8	57,14%
	Impressão tátil	1	7,14%
	Gráfico	2	14,28%
	Vídeo	1	7,14%
	Roteiro	2	14,28%
	Não classificado	1	7,14%
Tipo de geração	Automática	8	57,14%
	Semiautomática	2	14,28%
	Manual	1	7,14%
	Não classificado	3	21,42%
Canal de geração de áudio	Canal original	5	35,71%
	Canal em separado	1	7,14%
	Não utiliza	5	35,71%
	Não classificado	3	21,42%
Domínio da áudio descrição	Domínio geral	5	35,71%
	Domínio específico	8	57,14%
	Não classificado	1	7,14%
Método de avaliação	Automática	1	7,14%
	Com usuários	10	71,42%
	Híbrida	0	0,00%
	Especializada	1	7,14%
	Não classificada	2	14,28%

Sobre o tipo de entrada, a subcategoria Texto apresentou a maior quantidade de trabalhos (3 trabalhos – 21,42%). As subcategorias Ambiente 3D, Vídeo, Roteiro e Dados Matemáticos tiveram 2 trabalhos cada uma, correspondendo a 14,28% em cada subcategoria, enquanto que Dados Estatísticos teve apenas um trabalho (7,14%). Por fim, os trabalhos que não especificaram o tipo de entrada foram considerados na subcategoria “Não Classificado”, com um total de 2 trabalhos (14,28%).

Em relação ao tipo de saída, cinco trabalhos, correspondente a 35,71% do total, apresentaram mais de um tipo de saída, portanto a contagem ultrapassa o número total de artigos selecionados. O tipo de saída mais utilizado é o áudio, com 8 trabalhos (57,14%). Em seguida estão as saídas em texto com 5 trabalhos (35,71%). O tipo de saída gráfico e roteiro foi utilizado em 2 trabalhos (14,28%) cada um. A impressão tátil e o vídeo foram utilizados

como tipo de saída em um trabalho cada um. Apenas um trabalho não especificou o tipo de saída e foi considerado como “não classificado”.

Sobre o tipo da geração, a estratégia automática está presente na maior parte dos trabalhos, correspondendo a 57,14% dos trabalhos pesquisados (8 trabalhos, especificamente). A geração semiautomática foi utilizada em 2 trabalhos (14,28%). Apenas um trabalho apresenta a geração manual da audiodescrição (7,69%) realizada por audiodescritores ou profissionais de audiodescrição. Por fim, 3 trabalhos (21,42%) não especificaram a estratégia de geração da audiodescrição.

Em relação ao canal de geração de áudio, 5 trabalhos (35,71%) mixam a audiodescrição no áudio original do conteúdo, e apenas um trabalho utiliza um canal de áudio separado (7,14%). Adicionalmente, 5 trabalhos (35,71%) não utilizam áudio como saída e outros 3 (21,42%) não explicitaram o tipo de canal de áudio e foram considerados como “não classificado”.

Sobre o domínio da audiodescrição, a maior parte dos trabalhos (8 trabalhos – 57,14%) utiliza a audiodescrição em algum domínio específico, enquanto que 5 trabalhos (35,71%) apresentam uma abordagem de geração de audiodescrição para domínio geral. Apenas um trabalho (7,69%) não especifica um sistema de geração de audiodescrição por se tratar de um trabalho teórico (isto é, sem uma prototipação ou implementação prática) e, desta forma, foi considerado como “não classificado”.

Por fim, o método de avaliação mais utilizado nos trabalhos foi a avaliação com usuários, aplicado em 10 trabalhos (71,42%). A avaliação automática e a especializada foi utilizada como método de avaliação em apenas 1 trabalho (7,14%) para cada método. Os demais trabalhos (2 trabalhos – 14,28%) não especificaram como a avaliação da solução/sistema foi feita e foram considerados como “não classificado”.

A Tabela 2 apresenta a classificação específica de cada trabalho selecionado.

Tabela 2: Classificação de cada trabalho em relação às categorias.

Ref	Tipo de geração	Tipo de entrada	Tipo de saída	Canal de geração de áudio	Domínio da audiodescrição	Método de avaliação
DEMIR; OLIVER; SCHWARTZ, 2010	Automática	Texto (Webpage)	Texto	Canal original	Domínio geral	Com usuários
WANG; LI; LI, 2012	Automática	Texto	Impressão tátil, áudio e gráfico	Canal original	Domínio específico	Com usuários
FERRES; SEPÚLVEDA, 2011	Automática	Texto (Código HTML)	Texto	Não utiliza	Domínio específico	Automática
OKTAY; FOLMER, 2011	Automática	Ambiente 3D e texto	Texto	Não utiliza	Domínio geral	Não classificado
THOMAN; SRIPADA; NOORDZIJ, 2010	Automático	Dados estatísticos	Texto	Não utiliza	Domínio específico	Com usuários
ENCELLE; BELDAME; PRIÉ, 2013	Manual	Vídeo	Texto e áudio	Não classificado	Domínio geral	Com usuários
KOBAYASHI et al., 2010	Semi-automática	Roteiro de AD	Áudio	Não classificado	Domínio geral	Com usuários
CHAPDELAIN E; GAGNON, 2009	Não classificado	Não classificado	Áudio	Não classificado	Domínio geral	Com usuários
MOSKOVITC H; WALKER, 2010	Automática	Dados matemáticos	Áudio e gráfico	Canal original	Domínio específico	Com usuários
MURPHY; BATES; FITZPATRICK, 2010	Automática	Fórmulas matemáticas	Áudio	Canal original	Domínio específico	Com usuários
DULYAN; EDMONDS, 2010	Não classificado	Ambiente 3D	Áudio	Canal original	Domínio específico	Com usuários
O'CONNELL et. al., 2010	Não classificado	Não classificado	Não classificado	Não utiliza	Não classificado	Com usuários
GAGNON et al., 2009	Automática	Vídeo	Vídeo, áudio e roteiro	Canal separado	Domínio específico	Não especificado
LAKRITZ; SALWAY, 2002	Semi-automática	Roteiro Cinematográfico	Roteiro	Não utiliza	Domínio específico	Especializada

Segundo Nunes *et al.* (2011), a audiodescrição deve ser adequada ao tipo de produto visual que será descrito, portanto vamos considerar, da revisão sistemática da literatura, apenas os trabalhos relacionados diretamente à audiodescrição para filmes, ou seja, os trabalhos que utilizam vídeo e roteiro em suas estratégias e aplicações. Estes trabalhos serão descritos a seguir.

3.2 TOWARDS THE USAGE OF PAUSES IN AUDIO-DESCRIBED VIDEOS

Uma das dificuldades da audiodescrição é ter tempo disponível entre os diálogos para narrar as descrições. O trabalho de ENCELLE, BELDAME e PRIÉ (2013) se concentra no uso de pausas artificiais em vídeos audiodescritos. A proposta é aumentar o tempo disponível para a narração e, conseqüentemente, transmitir mais informações na audiodescrição.

A abordagem baseia-se em anotações de vídeo (*video annotations*) para enriquecer o vídeo durante a reprodução. Vídeos enriquecidos podem ter vários elementos extras associados a ele como legendas, imagens, links, áudio e etc. Este recurso se apresenta como uma forma de prover a acessibilidade à vídeos, uma vez que esses elementos extras podem facilitar a compreensão e tradução de partes dos vídeos.

Em casos onde a lacuna entre os diálogos é curta para narração da audiodescrição, as pausas artificiais são inseridas para aumentar o tempo disponível. Desta forma, para uma dada lacuna, a audiodescrição associada é transmitida no decorrer do espaço e prolonga-se durante a duração da pausa.

O objetivo do trabalho está relacionado com a utilidade dessas pausas como forma de enriquecer a experiência de audiodescrição das pessoas cegas. Para isso, as seguintes hipóteses foram levantadas: a) quanto maior for a duração da pausa mais desconfortável ela é; b) quanto maior for o número ranking da pausa menos desconfortável ela é (as pausas recebem o número ranking de acordo com sua posição de início); e c) as pausas são mais desconfortáveis durante a primeira exibição de um vídeo do que na segunda exibição.

Para o experimento, foi selecionada a animação “Tong”, com 8 minutos e 59 segundos de duração. Primeiramente, foram identificados os principais elementos que necessitavam de descrição e, como resultado, foram geradas 41 descrições textuais. Em seguida, o vídeo foi enriquecido com anotações do tipo “descrição”, que corresponde a uma descrição textual e uma lacuna associada. Neste experimento, a audiodescrição foi criada

por um sintetizador de voz, que realiza a narração das anotações do vídeo. Dentre as anotações, 26 descrições não se encaixavam dentro da lacuna, portanto necessitavam de mais tempo para serem narradas. Diante disso, 26 anotações do tipo “pausas artificiais” foram inseridas e posicionadas logo após o final das lacunas correspondentes. A duração das pausas artificiais foi definida da seguinte forma: D1) a duração da pausa é igual a 0,5 vezes a duração da lacuna; D2) a duração da pausa é igual à duração da lacuna; e D3) a duração da pausa é igual a 1,5 vezes a duração da lacuna. Diferentes versões do vídeo foram criadas utilizando as durações propostas com o intuito de testar os tipos de pausa.

O grupo de teste consistiu de 18 voluntários cegos, divididos em 3 grupos, cada um correspondendo à uma versão diferente do vídeo. Durante os testes, dados quantitativos e qualitativos foram coletados através de um questionário. Também foi solicitado que os participantes indicassem os pontos de desconforto durante duas exibições do vídeo.

Os resultados mostraram que as pausas artificiais oferecem um bom nível de aceitabilidade e apresentam um nível de desconforto baixo. Foi observado também que as pausas do tipo D3 parecem causar mais desconforto do que as do tipo D1 e D2. De acordo com a análise quantitativa, a localização da pausa tende a influenciar o nível de desconforto, uma vez que este diminuía lentamente ao longo do tempo. Este ponto também foi identificado em relação ao número de vezes em que o vídeo enriquecido foi assistido.

3.3 DESCRIBING ONLINE VIDEOS WITH TEXT-TO-SPEECH NARRATION

Kobayashi *et al.* (2010) apresentam uma plataforma que usa síntese de voz para adicionar audiodescrição à vídeos *on-line* de qualquer site. A tecnologia texto-para-fala (*text-to-speech* – TTS) permite que o roteiro de audiodescrição seja utilizado no formato de texto, sem a necessidade de um audiodescritor-narrador.

A arquitetura da plataforma consiste de um editor de roteiro, um player de vídeo, um repositório de metadados e um formato baseado em texto utilizado para a troca de roteiros entre os componentes. A Figura 2 apresenta esta arquitetura.

O editor de roteiro é uma ferramenta de apoio à criação de roteiros de AD, que fornece uma interface visual para editar as frases e especificar, na linha do tempo, os instantes em que cada descrição deve ser lida. A ferramenta permite que esta edição também possa ser feita por um descritor não especializado para melhorar a AD enquanto visualizam a narração resultante. Neste editor, é possível modificar os parâmetros da voz, tais como o

gênero do narrador e a velocidade. Para facilitar a identificação da localização das audiodescrições inseridas, a interface também exibe a forma de onda da faixa de áudio original.

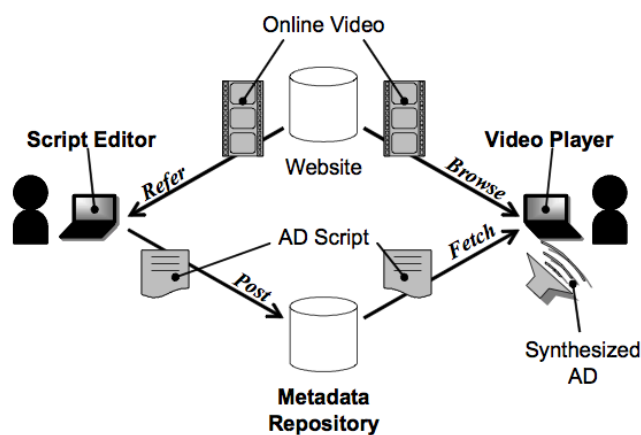


Figura 2: Arquitetura da Plataforma (KOBAYASHI et al., 2010)

Ao inserir uma audiodescrição, a duração esperada da narração é mostrada na linha do tempo. Neste caso, o usuário poderá ouvir a voz sintetizada e realizar ajustes para refinar a audiodescrição. Por fim, os roteiros de AD são armazenados e compartilhados em um repositório de metadados. Uma vez no repositório, o player de vídeo busca a audiodescrição do vídeo acessado e realiza as narrações. A plataforma utilizada permite sintetizar as narrações no servidor ou no cliente.

Experimentos com pessoas com deficiências visuais foram realizados para analisar a aceitabilidade dos sintetizadores de audiodescrição, como também o impacto da qualidade da voz na inteligibilidade dos vídeos. Para isso, foram exibidos vídeos de diversas categorias com audiodescrições em vários níveis de qualidade da voz sintetizada e também com a versão gravada por um narrador profissional humano para servir de comparação na análise dos resultados. Os resultados mostraram que a fala sintetizada era aceitável e poderia melhorar significativamente a experiência do usuário em relação a vídeos que não apresentam AD. O trabalho, contudo, não apresenta experimentos feitos com a ferramenta de edição de roteiro.

3.4 ACCESSIBLE VIDEODESCRIPTION ON-DEMAND

O trabalho apresenta a análise de acessibilidade de um *Website* que disponibiliza

vídeos com audiodescrição, através de um player de vídeo adaptado, chamado *VDPlayer* (CHAPDELAINE; GAGNON, 2009). O site acessível foi desenvolvido para apresentar 5 filmes de curta-metragem, e as audiodescrições são fornecidas em dois níveis de apresentação: 1) modo padrão, cujas descrições são inseridas nas lacunas entre os diálogos; e 2) modo estendida, que utiliza toda a audiodescrição produzida e que excede a duração das lacunas. Os autores mencionam que a audiodescrição dos vídeos é narrada através de sintetizador de voz, entretanto, não especifica como as descrições foram elaboradas.

O *VDPlayer* possui controles de vídeo padrão como *play*, *pause*, entre outros. Por ser um *player* adaptado à acessibilidade de vídeos, ele oferece também um conjunto de controles específicos para a audiodescrição, tais como, seleção do nível de audiodescrição (modo padrão ou estendido) e configurações gerais que permitem o avanço ou retorno à alguma narração de descrição.

No experimento foram selecionados 10 participantes para testar a solução. Estes usuários assistiram aos vídeos e, em seguida, responderam questionários sobre a audiodescrição e a interação com o *player*, com o objetivo de determinar a sua relevância e a facilidade de uso dos controles relacionados à audiodescrição. A avaliação da audiodescrição consistiu de uma série de nove declarações pré-estabelecidas que assumiam valores positivos ou negativos, onde os usuários deveriam selecionar o grau de concordância com cada uma delas.

Os resultados mostraram que 92% dos participantes concordaram fortemente com as declarações positivas, enquanto que 15% concordaram com as negativas. A avaliação da voz sintetizada obteve as pontuações mais baixas, sendo considerada moderadamente inaceitável. Além disso, os participantes relataram a impressão de que a audiodescrição encobria a percepção das informações importantes do áudio original do vídeo. De forma geral, a audiodescrição foi considerada de boa qualidade, mas necessita de algumas melhorias.

A avaliação do *VDPlayer* mostrou que 90% dos participantes consideraram que a seleção da opção entre vários tipos de audiodescrição é uma funcionalidade muito relevante. Os participantes relataram que acessaram os dois modos disponíveis, tendo a versão estendida como a preferência. Sobre a facilidade do uso dos controles, 90% dos participantes julgaram o *player* de fácil utilização.

3.5 ARE SYNTHESIZED VIDEO DESCRIPTIONS ACCEPTABLE?

Este trabalho faz um estudo sobre o uso de sintetizadores de voz para gerar a narração das descrições (O'CONNELL et. al., 2010). Para isso, duas principais questões foram exploradas: a) As narrações sintetizadas são aceitáveis na audiodescrição?; b) Como prover melhores experiências ao usuário nas descrições sintetizadas de vários tipos de filmes? Para investigar estas questões, foram conduzidos experimentos com pessoas deficientes visuais do Japão e dos Estados Unidos.

O experimento no Japão envolveu a exibição de amostras de 2 vídeos seguidas por entrevistas para coletar informações sobre a qualidade das descrições sintetizadas. Essas amostras apresentavam diferentes versões de 3 tipos de voz: a humana, *standard* TTS (*ProTalker*, IBM) e *prototype* TTS (IBM). Quanto às descrições, foram testados dois níveis de detalhamento: a rápida (2x a velocidade normal da fala) e a estendida (com pausas para as descrições que extrapolam o tempo das lacunas entre diálogos). Todas as audiodescrições foram produzidas e narradas por profissionais.

Um grupo de 115 adultos com deficiências visuais assistiram as amostras e os resultados mostraram que a voz humana foi preferível pela maioria, seguida pela voz do *prototype* TTS. Entretanto, um quarto dos participantes preferiram a narração sintetizada. Segundo a análise dos comentários qualitativos, os usuários preferiam a voz sintetizada devido ao fato de ser mais fácil de distinguir a audiodescrição das faixas de áudio original do vídeo. Um participante indicou que, para vídeos de entretenimento, as narrações humanas são mais recomendadas por permitir uma experiência mais natural e agradável.

Na fase das entrevistas, 3 participantes foram selecionados para assistir as amostras com duas apresentações diferentes: sem audiodescrição e com audiodescrição. Cada participante avaliou a experiência em uma escala de pontuação de 0 a 100 pontos, de acordo com os critérios de satisfação e inteligibilidade. Os resultados mostraram que a apresentação do vídeo com as descrições melhorou a pontuação em 30 a 48 pontos em relação ao vídeo sem audiodescrição. Quanto ao nível de detalhamento, todos os participantes preferiram a versão estendida.

Nos Estados Unidos, o experimento avaliou a qualidade da audiodescrição com a voz humana e com o sintetizador *Acapela* TTS. A pesquisa contou com a participação de 236 pessoas com deficiências visuais. Os participantes assistiram amostras de vídeo e responderam questionários relacionados à compreensão das narrações TTS, nível de conforto na experiência, preferências no tipo da narração e impacto no entendimento que a

qualidade da voz pode provocar.

Os resultados mostraram que a maior parte dos participantes acharam a audiodescrição TTS uma experiência confortável, agradável ou neutra. Em relação ao entendimento, os entrevistados relataram que de 52% a 79% das descrições TTS foram compreendidas. Na audiodescrição com narração humana, por outro lado, o nível de compreensão foi de 96%. Na fase das entrevistas, foram selecionados 8 participantes que consideraram, em sua maioria, as narrações TTS toleráveis ou até mesmo confortáveis para algum gênero específico de conteúdo.

De forma geral, os estudos mostraram que as descrições sintetizadas foram aceitas, independentemente da língua utilizada, mas se adequam melhor aos vídeos informativos, onde a compreensão é um fator crítico.

3.6 TOWARDS COMPUTER-VISION SOFTWARE TOOLS TO INCREASE PRODUCTION AND ACCESSIBILITY OF VIDEO DESCRIPTION FOR PEOPLE WITH VISION LOSS

Gagnon *et al.* (2009) apresentam um projeto de desenvolvimento de ferramentas de software para auxiliar a produção de audiodescrição. O objetivo é fornecer soluções computacionais que possam detectar automaticamente o conteúdo visual e ajudar o audiodescritor no processo de criação das descrições.

O trabalho tem como público-alvo dois tipos de usuário: os audiodescritores e os usuários finais com deficiência visual. O estudo envolveu três partes principais:

- 1) Reuniões com audiodescritores e análise de filmes audiodescritos para obter uma compreensão da prática dentro da indústria;
- 2) Reuniões com os usuários finais para entender suas necessidades e compará-las com as práticas da indústria;
- 3) Seleção e integração de ferramentas computacionais para a extração automática de conteúdo visual, tendo em conta as necessidades dos audiodescritores e usuários finais, bem como a viabilidade técnica.

Na primeira parte da pesquisa, audiodescritores foram convidados a descrever o processo de trabalho utilizado por cada um na elaboração da AD, como também os elementos visuais que eles consideravam mais importante e quanto tempo era necessário para descrevê-los. As informações coletadas ajudaram a identificar as especificações do sistema computacional de auxílio à produção da AD, apresentadas a seguir: detecção de

rosto (e a frequência em que aparecem), detecção de espaços de silêncio, construção de um player que permita aos usuários repetir as descrições ou obter informação adicional, fornecimento de um resumo do filmes contendo detalhes de personagens para facilitar o planejamento da produção da audiodescrição e a descrição os movimentos de pessoas.

A análise de onze filmes de gêneros variados feita pelos participantes indicou que as categorias de informação mais encontradas, em ordem decrescente de ocorrências são: ação, movimento de personagens, ocupação dos personagens, cenário, expressões faciais e corporais, informação textual incluída na imagem e, por fim, informação sobre a atitude dos personagens.

A segunda parte da pesquisa investigou como as diferentes informações da audiodescrição atingem as necessidades dos usuários com deficiência visual. A análise das informações coletadas dos usuários destacou as seguintes informações técnicas: a existência de conflitos entre a audiodescrição e a trilha sonora original do filme, a importância da escolha certa de voz para a narração, e a possibilidade de personalização do nível de detalhamento da audiodescrição.

A terceira etapa da pesquisa consiste do projeto e implementação de um software para descrição do conteúdo visual de filmes, utilizando as informações obtidas nas etapas anteriores com audiodescritores e usuários com deficiências visuais. Devido às limitações técnicas, o objetivo do sistema é prover a sumarização de alguns elementos do filme, a fim de diminuir o esforço empregado pelo profissional no processo de criação da audiodescrição fornecendo dois produtos finais: um arquivo com a audiodescrição narrada por um sintetizador, ou um roteiro com os tempos e descrições que pode ser editado através de um player.

O sistema propõe as seguintes funções: detecção de lugares importantes e recorrentes (lugar-chave), detecção de faces (face-chave) e reconhecimento de texto importante para a compreensão da história (texto-chave), e ação de pessoas. A detecção de transições de tiro e segmentos silenciosos também fazem parte do sistema desenvolvido, mas não são descritas no artigo. Essas detecções automáticas do conteúdo visual são feitas através de uma coleção de filtros especializados que extraem informações de alto nível, afim de descrevê-los.

Na detecção de lugar-chave, a abordagem utilizada é baseada na busca de ligações entre os principais quadros do filme, usando análise probabilística latente semântica (pLSA) para permitir a extração e grupos de descritores locais que podem representar elementos característicos de um lugar-chave. Para testar esta abordagem, foram utilizados dois filmes, com a aparição de quatro locais principais em 822 e 412 cenas, respectivamente, que foram

previamente identificados de forma manual. A solução extraiu agrupamentos de cenas que representam locais com taxa de acerto de 78% no primeiro filme e 68% no segundo.

Assim como na detecção de lugares, o desafio da detecção de face é agrupar as faces dos principais atores de um filme dentre as várias imagens de faces presentes nele. Para isso, são aplicados algoritmos para formar agrupamentos de rostos similares baseados na observação de características e semelhanças entre as faces. A abordagem foi testada com um filme contendo 14 atores que foram previamente identificados de forma manual. Para analisar o resultado, foi considerado a pureza dos agrupamentos, que significa a porcentagem de imagens de rosto representando o mesmo rosto em um grupo. Os resultados mostraram que 100% dos grupos tem a pureza superior a 20%; e 60% dos grupos têm pureza superior a 50%.

A detecção de texto consiste de três etapas principais: a detecção de texto, a segmentação do texto e o reconhecimento óptico de caracteres (OCR). De forma geral, este componente funciona da seguinte forma: o detector de texto localiza regiões de interesse no quadro do filme; estas regiões passam pelo processo de segmentação para remover o fundo da imagem e ruídos; e o reconhecimento óptico de caracteres é realizado por um software OCR comercial, que retorna um conjunto de caracteres reconhecidos.

A detecção de ações humanas, como andar, correr e etc., é realizada através do acompanhamento de movimentos extraídos da sequência do vídeo. A partir de características selecionadas da imagem inicial, o movimento é acompanhado através de pontos dispersos quadro-a-quadro. Essas informações são utilizadas para acompanhar os movimentos de câmera e, com a utilização de algoritmos de agrupamento, localizar o plano de fundo da cena e encontrar pessoas ou objetos em movimento.

Para realizar e coordenar as tarefas descritas acima, três ferramentas foram criadas de acordo com as etapas da detecção do conteúdo visual. O *Audio-vision manager (AVM)* gerencia o processamento, a aquisição de parâmetros, a ordem de execução, as dependências entre os módulos e integridade dos dados. De forma geral, o AVM extrai a informação visual bruta. Em seguida, o *Audio-Visio generator (AVG)* converte a informação extraída pelo AVM para descrições legíveis com a adição de marcações de tempo para cada descrição. Por fim, o *Audio-vision Player (AVP)* faz a apresentação da audiodescrição, além de permitir que o usuário modifique o nível de descrição do vídeo.

O trabalho ainda está em fase de desenvolvimento e os resultados preliminares mostraram que a abordagem ainda não atende as necessidades dos usuários cegos de forma

satisfatória. Os autores, no entanto, mencionam que a viabilidade técnica do processo é claramente demonstrada.

3.7 THE SEMI-AUTOMATIC GENERATION OF AUDIO DESCRIPTION FROM SCREENPLAYS

Este trabalho apresenta uma abordagem semiautomática de geração de audiodescrição através da análise do roteiro cinematográfico (LAKRITZ; SALWAY, 2002). Esta geração é tratada como um problema de sumarização de texto em que frases relevantes são identificadas no roteiro e depois adaptadas para atender as diretrizes da audiodescrição.

Uma análise de roteiros cinematográficos e de audiodescrição foi realizada com o objetivo de verificar como os dois tipos de texto descrevem o mesmo evento, além das variações de linguagem presentes em cada um deles. Nessa análise, foram utilizados 70 roteiros cinematográficos e os respectivos roteiros de audiodescrição.

A análise baseou-se na frequência de ocorrência das palavras contidas no roteiro cinematográfico, com a criação de 3 listas:

- 1) 30 palavras mais frequentes em qualquer roteiro de AD;
- 2) nomes de personagens do filme, e
- 3) 20 palavras mais frequentes específicas do domínio, utilizando as palavras de maior ocorrência no roteiro.

Os resultados da análise indicaram as seguintes conclusões:

- a) em média, um roteiro contém cerca de 60% da informação necessária para uma audiodescrição, embora não necessariamente seja expressa na forma adequada;
- b) em 80% das descrições presentes em ambos roteiros, a descrição do roteiro cinematográfico contém ao menos uma palavra mais frequente;
- c) 50% das sentenças do roteiro cinematográfico que contém ao menos uma palavra mais frequentes está presente no roteiro de audiodescrição;
- d) 20% das sentenças do roteiro cinematográfico que contém ao menos uma palavra mais frequente usa uma linguagem diferente na audiodescrição.

Diante desses resultados, o trabalho propõe um método para gerar automaticamente a versão inicial da audiodescrição em dois estágios: a) extração de frases candidatas para a audiodescrição a partir de um roteiro, utilizando a frequência de ocorrência das palavras

como medida de importância; b) adaptação da linguagem da frase, quando necessário, para atender às diretrizes de estilo da audiodescrição.

O sistema possui 4 módulos: 1) Localizador de lacunas entre diálogos; 2) Extrator de informações pertinentes do roteiro; 3) Conversor para linguagem de AD; 4) Compressor da sentença para caber na lacuna. No entanto, apenas os módulos 2 e 3 foram desenvolvidos. A extração das frases do roteiro é feita selecionando as sentenças que contém ao menos uma palavra contida em qualquer uma das três listas das palavras mais frequentes. A conversão da linguagem, por outro lado, envolve a edição, reordenamento e exclusão de elementos das sentenças através da aplicação de heurísticas.

Para avaliar o sistema, foram geradas as descrições de três filmes de gêneros diversos. Os resultados mostraram que 80% das sentenças podem ser recuperadas a partir de um roteiro cinematográfico, com uma taxa de precisão de 50%. Quanto à conversão, 66% das sentenças com formas inadequadas à audiodescrição foram mapeadas à uma forma mais adequada através da aplicação do conjunto de heurísticas.

O sistema foi avaliado por profissionais descritores, que indicaram que o processo de geração semiautomática de audiodescrição é possível e aplicável, no entanto, a solução não aumentou a eficiência no processo de criação do roteiro AD, uma vez que a saída produzida não contém as marcações de tempo (ou pontos de sincronização) relacionando a AD ao vídeo. Desta forma, a identificação manual dos intervalos onde a AD deve ser inserida e a busca pela descrição correta para cada intervalo de tempo acabam demandando muito esforço.

3.8 CONSIDERAÇÕES FINAIS

Este capítulo apresentou trabalhos relacionados ao uso computacional no processo de elaboração da audiodescrição de vídeos. Os trabalhos foram selecionados a partir de uma revisão sistemática da literatura. De forma geral, as técnicas são aplicadas em várias etapas do processo, como a apresentação da audiodescrição, a narração das descrições e geração do roteiro.

Um ponto em comum presente nos trabalhos Encelle, Beldame e Prié (2013) e Chapdelaine e Gagnon (2009) é a disponibilização de diversos níveis de detalhamento da audiodescrição, com versões padrão e estendida. Outro ponto é o desenvolvimento de

players adaptados que facilitam a exibição dos vídeos com audiodescrição, recurso este apresentado nos trabalhos de Chapdelaine e Gagnon (2009) e Gagnon *et al.* (2009).

Quanto à narração das descrições, O’Connell *et al.* (2010) e Kobayashi *et al.* (2010) fazem o estudo do uso de sintetizadores de voz na audiodescrição e a aceitabilidade dos usuários. No entanto, todos os trabalhos utilizavam a voz sintetizada de forma integrada à sua solução ou em experimentos para coletar informações dos usuários, com exceção do trabalho de Lakritz e Salway (2002).

Em particular, na geração automática do roteiro de audiodescrição, os trabalhos possuem algumas limitações que não satisfazem as necessidades atuais da indústria. Dos artigos selecionados, apenas Gagnon *et al.*, 2009) e Lakritz e Salway (2002) apresentam uma abordagem de geração automática/semiautomática das descrições, sendo o primeiro com o uso de técnicas de reconhecimento de imagens que apresentaram uma alta taxa de erro, e o segundo utilizando o roteiro cinematográfico como base para a criação da audiodescrição, porém sem as marcações de tempo.

Diante deste cenário, este trabalho se propõe a desenvolver uma solução automática de geração de roteiros de audiodescrição para otimizar o processo de elaboração atual da AD, denominada *CineAD*. Para isso, o *CineAD* foi desenvolvido para utilizar o roteiro cinematográfico e a legenda dos filmes na criação das descrições vinculadas ao tempo do vídeo. A Tabela 3 apresenta a classificação do sistema em relação às categorias e critérios estabelecidos no protocolo da revisão sistemática e em comparação com os trabalhos descritos neste capítulo.

Tabela 3: Classificação de *CineAD* em relação às categorias.

Ref	Tipo de geração	Tipo de entrada	Tipo de saída	Canal de geração de áudio	Domínio da audiodescrição	Método de avaliação
ENCELLE; BELDAME; PRIÉ, 2013	Manual	Vídeo	Texto e áudio	Não classificado	Domínio geral	Com usuários
KOBAYASHI et al., 2010	Semi-automática	Roteiro de AD	Áudio	Não classificado	Domínio geral	Com usuários
CHAPDELAINE; GAGNON, 2009	Não classificado	Não classificado	Áudio	Não classificado	Domínio geral	Com usuários
O’CONNELL et. al., 2010	Não classificado	Não classificado	Não classificado	Não utiliza	Não classificado	Com usuários
GAGNON et al., 2009	Automática	Vídeo	Vídeo, áudio e roteiro	Canal separado	Domínio específico	Não especificado

LAKRITZ; SALWAY, 2002	Semi- automática	Roteiro Cinematográfico	Roteiro	Não utiliza	Domínio específico	Especiali- zada
CineAD	Automática	Roteiro Cinematográfico	Roteiro	Não utiliza	Domínio específico	Com usuários, especializa da

A geração do sistema *CineAD* é do tipo automática e utiliza o roteiro cinematográfico como entrada e elemento principal para geração das descrições. O sistema não utiliza um canal de áudio, uma vez que o objetivo é a criação do roteiro de audiodescrição, sendo este roteiro o tipo da saída fornecida pelo sistema. Quanto ao domínio da audiodescrição, o sistema trabalha com domínio específico: filmes. Por fim, o método de avaliação utilizado no trabalho consistiu de testes com usuários e com profissionais especializados. Os experimentos serão descritos no Capítulos 5, enquanto que o sistema *CineAD* e seus componentes serão descritos no capítulo a seguir.

4 SOLUÇÃO PROPOSTA - *CINEAD*

Nesse capítulo, os componentes e o funcionamento da solução proposta serão apresentados. Conforme mencionado no Capítulo 1, a solução, denominada CineAD, tem como objetivo reduzir os problemas de acessibilidade no cinema que as pessoas cegas usualmente enfrentam. Para isso, a solução é composta por um conjunto de componentes responsáveis por gerar, de forma automática, roteiros de audiodescrição para cinema, através da análise do conteúdo da legenda e do roteiro original do filme.

O funcionamento geral do sistema é apresentado na Seção 4.1. As demais seções (Seção 4.2 à 4.5) descrevem cada um dos componentes integrantes da solução separadamente, destacando as suas entradas e saídas dentro do fluxo de processamento do sistema.

4.1 CINEAD

O *CineAD* tem como objetivo reduzir os problemas de acessibilidade à filmes cinematográficos enfrentados pelas pessoas cegas ou com baixa visão. Para isso, a solução propõe a geração automática de roteiros de audiodescrição. A abordagem consiste da criação da audiodescrição a partir da análise de dois elementos principais: o roteiro do filme e a legenda. A visão esquemática do sistema é ilustrada na Figura 3. Neste trabalho, todos os componentes foram implementados na linguagem Java.

O sistema funciona da seguinte forma. Inicialmente, o roteiro original do filme é analisado e seus principais elementos são extraídos, como títulos de cena, ações, personagens e outros. O componente de Identificação de Gaps realiza a detecção dos intervalos de tempo entre os diálogos do filme, o que os caracterizam como possíveis gaps sem falas, candidatos para futuras inserções de audiodescrição. Em seguida, o componente de Sumarização faz a extração das sentenças mais importantes do roteiro e, desta forma, resume o roteiro original, descartando as informações secundárias, consideradas menos importantes para a audiodescrição. Por fim, o componente de Geração de Roteiro de AD gera o roteiro de audiodescrição alocando as sentenças que foram extraídas na etapa de sumarização dentro dos gaps detectados na etapa de Identificação de Gaps. Os componentes da solução serão detalhados a seguir nas Seções 4.2, 4.3, 4.4 e 4.5.

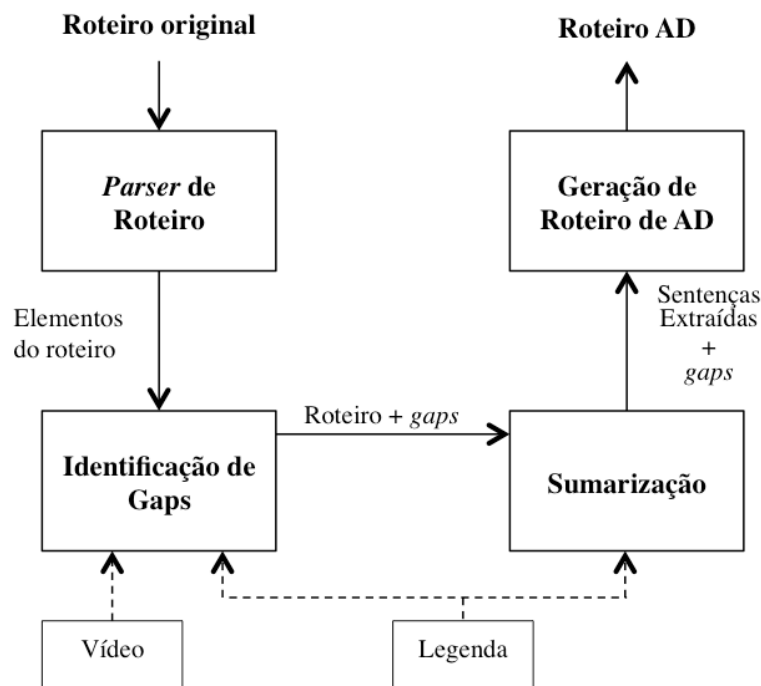


Figura 3: Visão esquemática do sistema *CineAD*

4.2 PARSE DE ROTEIROS

Este componente é responsável pela leitura e extração dos elementos contidos no roteiro, como os títulos de cena, as ações, os diálogos, os personagens, dentre outros. Para facilitar a realização dessa tarefa, é importante que um formato digital, que permite a identificação e rotulagem dos elementos do roteiro, seja utilizado.

Neste trabalho, o formato *.celtx*, definido pela ferramenta CELTX, foi utilizado, mas é possível estender esse módulo para suportar outros tipos de formato. Uma das motivações para a escolha do CELTX, foi o fato dela ser uma ferramenta *open source*.

O CELTX é um processador e editor de textos multiplataforma, livre, de pré-produção, para escrever roteiros audiovisuais, como filmes, peças teatrais, etc. (CELTIX, 2014). O programa CELTX utiliza o formato *.celtx*, que é uma pasta compactada com quatro arquivos: dois no formatos *rdf*, contendo metadados, e dois no formato *html*, contendo roteiros e informações extras como anotações, fichas, entre outras. No CELTX, é possível

escrever roteiros e identificar cada elemento com um rótulo específico. Na Figura 4 é apresentado o ambiente de edição e um exemplo da estrutura do roteiro no programa.

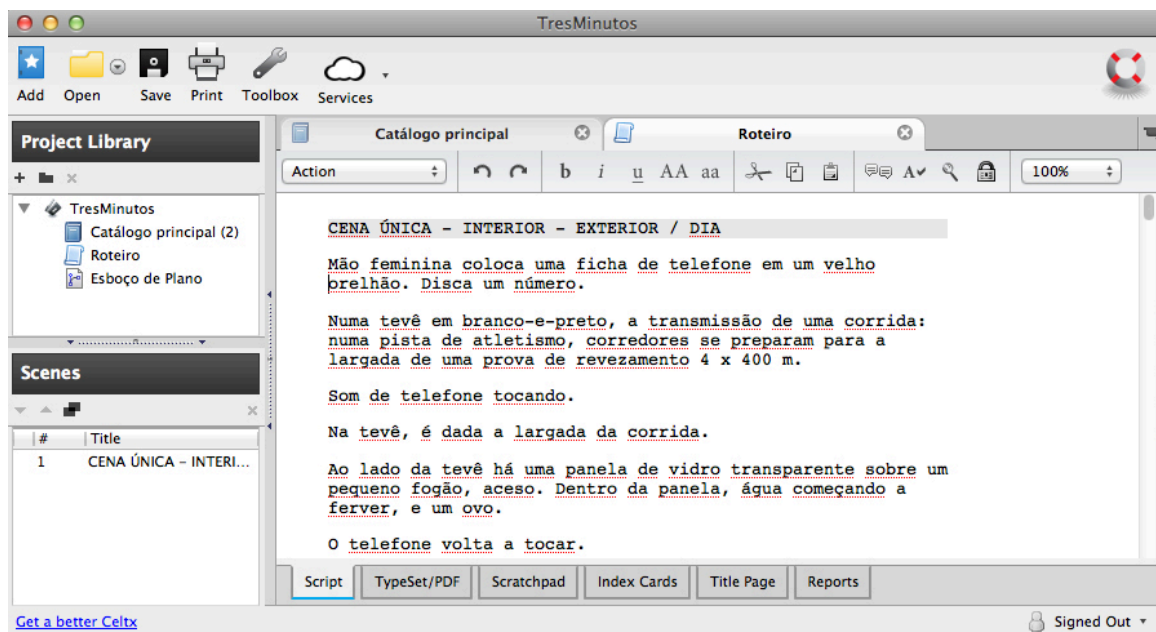


Figura 4: Ambiente de edição do CELTX.

O componente *Parser* de Roteiro faz a leitura dos arquivos contidos no formato *.celtx* e extrai os elementos do roteiro, como títulos de cena, personagens, diálogos e ações. Para realizar a leitura, foram utilizados os recursos de manipulação de *rdf* do *Apache Jena* (JENA, 2014), que é um framework em Java que dá suporte a utilização da Web Semântica e que contém bibliotecas para uso destes tipos de arquivo.

Como resultado deste processo de leitura e extração, um conjunto de elementos ordenados, cronologicamente, e de acordo com a sequência do roteiro original é gerado como saída deste componente. Neste trabalho, a audiodescrição realizada pelo sistema descreve apenas as ações do filme, ou seja, os elementos do roteiro rotulados como “ação”.

4.3 IDENTIFICACAO DE GAPS

O componente de Identificação de *Gaps* tem como objetivo detectar os intervalos de tempo do filme que não contém diálogos de personagens. Esses *gaps* são espaços no áudio original candidatos a receber trechos de audiodescrição, uma vez que eles não interferem nas falas do filme. Para a identificação desses intervalos, a legenda é utilizada, pois contém os

textos associados aos diálogos, além das marcações de início e fim de cada diálogo. Cada *gap* identificado possui uma marcação de tempo inicial e final associada aos intervalos sem falas, funcionando, dessa forma, como pontos de sincronização para inserção dos trechos de AD, conforme ilustrado na Figura 5. Arquivos de legendas no formato *.srt* e vídeos em formato *.mp4* foram utilizados nesta implementação.

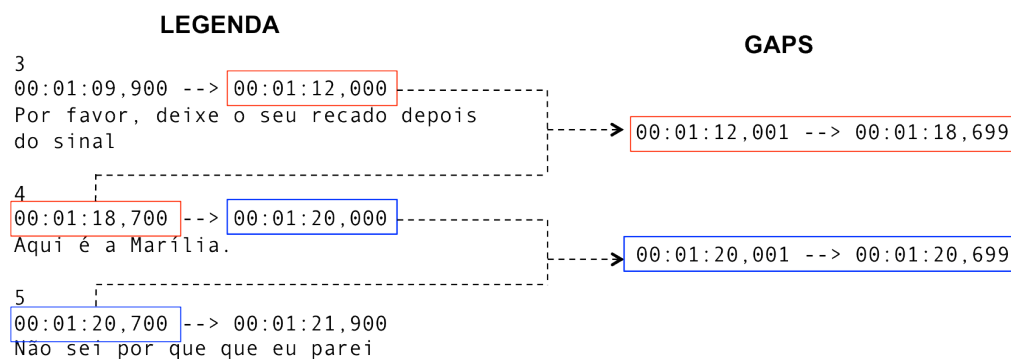


Figura 5: Detecção de intervalos sem diálogos (*gaps*) a partir da legenda.

Uma dificuldade desta abordagem é a detecção da marcação de tempo final do último *gap*, pois essa informação não está na legenda e o final da mesma não necessariamente coincide com o encerramento do filme. Dessa forma, para se obter o último *gap* (intervalo entre o último diálogo e o fim do filme), a duração do vídeo é consultada e representa o tempo final do último *gap*.

4.4 SUMARIZAÇÃO

Diferentemente do roteiro cinematográfico que contém toda a descrição de um filme, o roteiro de audiodescrição não comporta todas as informações presentes no roteiro original, uma vez que ele está limitado aos intervalos em que não há diálogos. Desta forma, torna-se necessário resumir as informações do roteiro original para inseri-las na audiodescrição. Para isso, o componente de sumarização realiza a extração das sentenças mais importantes do roteiro e descarta as outras sentenças, considerando-as menos relevantes.

Uma das técnicas de sumarização é a extração de sentenças que consiste na concatenação de várias sentenças extraídas do material que se deseja sumarizar, na forma em

que aparecem no texto e sem sofrer alterações (NENKOVA, MCKEOWN, 2011). Segundo Edmundson (1969), as informações importantes em um texto podem ser identificadas com base na frequência das palavras-chave contidas nele. Portanto, encontrando as palavras mais frequentes, é possível detectar as sentenças mais importantes em um documento.

Esta abordagem foi utilizada, por exemplo, na extração de sentenças para a geração semiautomática de audiodescrição proposta por Lakritz e Salway (2002). Essa extração baseada em frequência de palavras retornou 80% das descrições importantes, o que demonstra que o método é capaz de extrair informações relevantes para a audiodescrição. Considerando esses resultados, a solução proposta neste trabalho utiliza um método similar ao apresentado em Lakritz e Salway (2002) para extração de sentenças.

A extração de frases no *CineAD* é feita da seguinte forma. Inicialmente, são criadas três listas de palavras com maior frequência:

- 1) as palavras mais frequentes em roteiros de audiodescrição;
- 2) as palavras mais frequentes no roteiro que será audiodescrito; e
- 3) uma lista com o nome de todos os personagens.

Salway, Vassiliou e Ahmad (2005) listaram as 30 palavras mais comuns em roteiros de audiodescrição, identificadas através da análise de roteiros de AD de filmes de domínio variados. As palavras destacadas pelo trabalho, após tradução e adaptação para a língua portuguesa, formam a primeira lista do componente de sumarização (ver APÊNDICE A). É importante mencionar que essa lista de palavras não é modificada independentemente do domínio do filme.

A segunda lista contém as 30 palavras mais frequentes do roteiro que se deseja gerar a audiodescrição. Para isso, primeiramente, é necessário realizar o processo de remoção de *stopwords*, como por exemplo, artigos, pronomes, interjeições, advérbios, preposições, etc. As *stopwords* são palavras consideradas irrelevantes e devem ser eliminadas pois geralmente não representam o conteúdo do texto. Na etapa de criação desta lista, uma biblioteca de mecanismos de procura de texto escrita em Java, a *Apache Lucene* (LUCENE, 2014a), foi utilizada para retornar os termos que mais ocorrem no documento. A *Lucene* possui um analisador específico para a língua portuguesa contendo todos os *stopwords* do idioma.

Devido à variação dos tempos verbais e da conjugação da língua portuguesa, a *Lucene* considera apenas o radical das palavras para identificar as palavras mais frequentes. Isso garante que as diferentes variações da palavra sejam consideradas. Por exemplo, para o verbo *correr*, apenas o radical “*corr*” será levando em conta, e com isso qualquer variação da palavra (*corremos*, *correu*, *corrida* e *etc.*) pode ser considerada. Esse mesmo

comportamento é também é aplicado através do *Lucene* na primeira lista de palavras mais frequentes.

Por fim, a terceira lista é formada pelos nomes de todos os personagens do roteiro-alvo. Essas informações são obtidas através da leitura dos metadados do arquivo *.celtx* que contém a quantidade de personagens, o nome e descrição de cada participante da história do roteiro.

Após a criação das listas, todas as ações do roteiro são fragmentadas em frases, de modo que cada ação seja subdividida em suas frases constituintes. A seleção das sentenças para o roteiro de audiodescrição é realizada da seguinte forma. Para cada sentença é verificada a presença de pelo menos uma das palavras contidas nas 3 listas. Caso a sentença tenha, no mínimo, uma palavra das listas de palavras mais frequente, ela será uma sentença candidata ao roteiro de audiodescrição. A Figura 6 apresenta um exemplo deste processo de extração das sentenças candidatas.

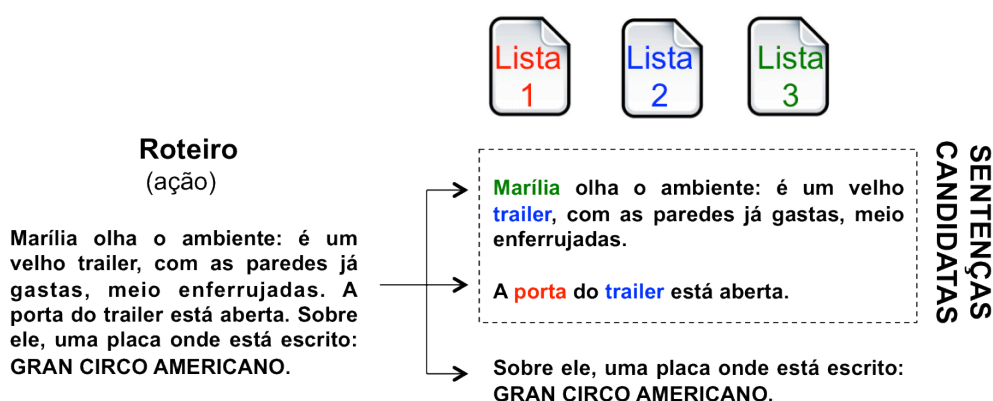


Figura 6: Processo de Extração das sentenças candidatas.

Em seguida, para cada sentença, é calculada uma média ponderada de acordo com a quantidade de palavras que possui de cada uma das três listas. Essa média será um *score* que define a sua importância, a partir da quantidade e do tipo das palavras mais frequentes do roteiro contidas na frase. A Tabela 4 apresenta o peso atribuído a cada lista.

Tabela 4: Peso das Lista de Palavras Frequentes.

Lista	Peso
1. Palavras comuns em AD	2
2. Palavras comuns do domínio	1
3. Nome dos Personagens	3

Esses pesos foram escolhidos com base na frequência e importância das palavras no roteiro. A terceira lista, nomes de personagens, recebeu o maior peso (peso 3) devido a sua importância para o enredo do filme, pois, quando um personagem é referenciado dentro do roteiro, geralmente ela descreve uma ação realizada por, ou ocorrida com, ele. A segunda lista com as palavras de domínio do filme recebeu o menor peso (peso 1), devido a alta frequência de ocorrências dentro do roteiro. Por fim, a lista com as palavras comuns em AD recebeu o peso 2.

Com a definição dos pesos, a média ponderada de cada sentença é calculada da seguinte forma:

$$Score = \frac{(qtde_{lista1} \times 2) + (qtde_{lista2} \times 1) + (qtde_{lista3} \times 3)}{quantidade\ total\ de\ palavras\ da\ sentença}$$

Onde $qtde_{lista1}$, $qtde_{lista2}$ e $qtde_{lista3}$ representam, respectivamente, a quantidade de palavras das listas 1, 2 e 3 presentes na sentença em questão.

O resultado deste processo é um conjunto de sentenças candidatas extraídas do roteiro, cada uma com seu respectivo *score*. Esse *score* é utilizado pelo componente seguinte para geração do arquivo com o roteiro de AD.

4.5 GERACAO DE ROTEIRO DE AUDIODESCRICOÃO

O componente Gerador de Roteiro cria o roteiro com a audiodescrição e os seus tempos específicos com base nas informações produzidas pelos componentes anteriores. Mais especificamente, o objetivo do componente é a alocação das sentenças extraídas na etapa de sumarização dentro dos intervalos sem falas (*gaps*) identificados pelo componente de Identificação de Gaps. O componente é executado em quatro etapas:

- 1) verificação de similaridade entre as falas da legenda e os diálogos do roteiro;
- 2) extração de ações entre os diálogos;
- 3) divisão dos gaps para inserção das ações; e
- 4) geração do arquivo com o roteiro AD.

O fluxo da geração é apresentado na Figura 7, e o funcionamento de cada uma das etapas desse fluxo é apresentado nas Seções 4.5.1, 4.5.2, 4.5.3 e 4.5.4.

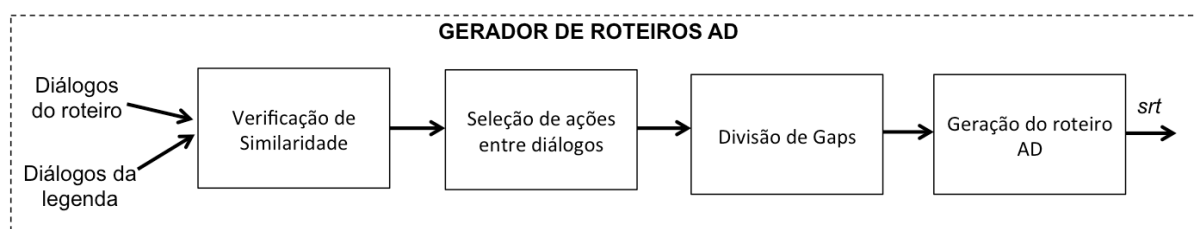


Figura 7: Fluxo da geração do roteiro AD.

4.5.1 Verificação de Similaridade

A primeira etapa verifica a similaridade da legenda com os diálogos do roteiro, fazendo com que as sentenças (ou bloco de sentenças) da legenda sejam associadas a um diálogo de personagem contido no roteiro. Essa verificação é necessária porque um diálogo no roteiro pode estar associado a várias sentenças (ou blocos de sentenças) da legenda. Na Figura 8 é ilustrado um exemplo deste caso.

Para relacionar as sentenças da legenda com cada diálogo, um cálculo da similaridade é realizado utilizando a biblioteca *Apache Lucene* (LUCENE, 2014). A *Lucene* faz a indexação dos textos originais e pesquisa, dentro do índice, os resultados com maior similaridade com a consulta desejada. Como resultado desta etapa, é possível ter pares de informações de diálogos legenda-roteiro associadas ao tempo específico do vídeo. Com isso, os diálogos do roteiro passam a ter marcações de tempo.

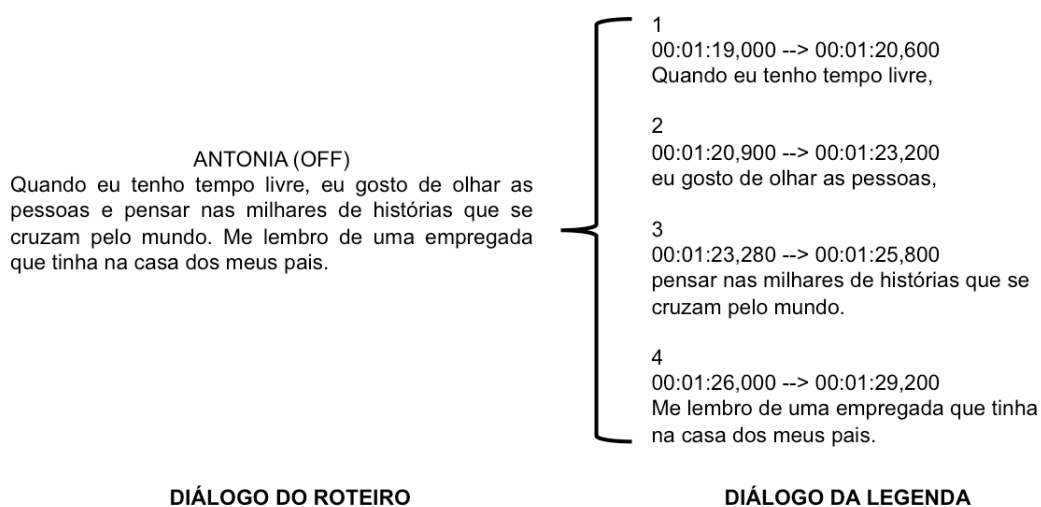


Figura 8: Exemplo da similaridade entre diálogos do roteiro e legenda (Trecho do roteiro do filme *A Cartomante*, 2004).

4.5.2 Seleção de ações entre diálogos

Após o mapeamento dos diálogos, a etapa seguinte utiliza as ações selecionadas pelo componente de sumarização e que estão cronologicamente entre diálogos no roteiro. É importante ressaltar que, entre duas falas do filme, um conjunto de ações importantes para a história podem ocorrer, como mostrado na Figura 9. Dessa forma, nesta etapa, as sentenças candidatas que estão entre dois diálogos são selecionadas, ou seja, é feita uma associação dos gaps identificados e as sentenças candidatas contidas nele, cronologicamente. Essa etapa utiliza como base as informações produzidas pelo componente de sumarização e os gaps sem falas identificados pelo componente Identificador de Gaps.

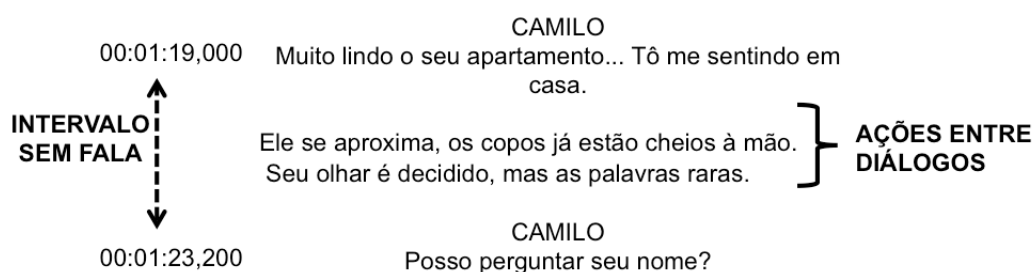


Figura 9: Exemplo de ações entre diálogos do roteiro (Trecho do roteiro do filme *A Cartomante*, 2004)

4.5.3 Divisão de Gaps

Em seguida, na etapa de divisão do gaps, o tempo disponível para a audiodescrição é dividido entre as sentenças candidatas. Entretanto, é possível que nem todas as sentenças selecionadas pelo sumarizador façam parte do roteiro AD. Nunes *et al.* (2011) ressaltam que é importante que poucas palavras sejam utilizadas na audiodescrição para evitar um excesso de informações no áudio. Portanto, a duração do *gap* definirá quantas sentenças candidatas farão parte do roteiro final.

Para realizar a divisão, o componente consulta a duração do intervalo que será utilizado para inserir as ações selecionadas na etapa anterior. O intervalo de tempo é então dividido de acordo com a quantidade de palavras que é possível narrar em sua duração. Araújo (2006) cita que, em média, as pessoas são capazes de ler 180 palavras por minuto em uma legenda, ou seja, cerca de 3 palavras por segundo. O *CineAD* utiliza este parâmetro para fazer a divisão dos *gaps*.

Desta forma, a quantidade de palavras para cada *gap* é definida a partir da duração do intervalo. Por exemplo, um *gap* de 20 segundos pode conter até 60 palavras. Dessa forma, as frases narradas neste intervalo deverão ter, acumuladas entre si, um total de, no máximo, 60 palavras. Qualquer sentença candidata que ficar fora deste limite será excluída do roteiro de AD.

A definição das sentenças candidatas que serão inseridas nos *gaps* é feita a partir do *score* calculado pelo componente de sumarização. Este *score* define o nível de importância que aquela sentença tem dentro do roteiro, a partir das palavras mais frequentes contidas nele. Com isso, é possível fazer um ranking das sentenças candidatas de acordo com o *score*, ordenando as frases conforme a sua importância dentro do roteiro. A divisão do *gap* ocorre através da inserção das frases mais importantes, com os *scores* mais altos, que acumuladas, contém o número máximo de palavras que poderão ser narrada dentro do *gap*, levando em conta o parâmetro de 3 palavras por segundo.

Um fator importante nesta etapa é que as sentenças são inseridas no *gap* de acordo com sua ordem cronológica na história do filme, apresentada no roteiro, portanto, a ordem do ranking do *score* serve apenas para seleção das sentenças mais importantes.

Desta forma, o sistema garante uma sincronização relativa das ações com o tempo no vídeo, o que faz com que a descrição apareça de forma sincronizada com a ocorrência do evento e, ao mesmo tempo, sem sobrecarregar o áudio com muitas descrições a serem narradas.

4.5.4 Geração do roteiro de audiodescrição

Após a alocação das ações no respectivos *gaps*, o roteiro de audiodescrição é gerado. A última etapa do componente de Geração de roteiro de AD consiste da criação de um arquivo no formato *.srt*, comumente utilizado para representar legendas, contendo o roteiro de AD. A escolha deste formato baseou-se nos requisitos que o sistema exige, sendo necessário o destaque do texto descritivo e do intervalo de tempo que deverá ser utilizado para a narração de cada trecho de audiodescrição. Desta forma, o roteiro de AD contém a audiodescrição em texto e as marcações de tempo (pontos de sincronização) para sua apresentação. Esse arquivo *srt* poderá servir como base para narradores humanos, audiodescritores ou para sistemas de voz sintetizada.

Um exemplo de um roteiro de AD gerado pelo *CineAD* com o relacionamento entre os elementos da legenda e do roteiro original é apresentado na Figura 10. A coluna da esquerda mostra os intervalos de tempo e as falas da legenda. A coluna central apresenta o roteiro cinematográfico original com ações e diálogos, enquanto que a coluna da direita mostra o roteiro de audiodescrição gerado automaticamente. É possível observar que as falas da legenda estão relacionadas com os diálogos do roteiros da mesma forma que as descrições do roteiro AD estão associadas às ações do roteiro original.

LEGENDA	ROTEIRO	ROTEIRO AD
37 00:02:40,300 --> 00:02:42,300 Tu sempre diz que isso é bobagem,	Recuando, saindo pela porta.	
38 00:02:42,400 --> 00:02:43,400 que isso é pro público.	RECADO (VS - cont) Tu sempre diz que isso é bobagem, que isso é pro público. Mas eu quero ser que nem o público. Será que é tão difícil de entender isso? Será que é tão difícil entender que eu só queria ter uma vida...	11 00:02:38,500 --> 00:02:40,300 Recuando, saindo pela porta.
39 00:02:44,300 --> 00:02:46,600 Mas eu quero ser que nem o público.		12 00:02:54,000 --> 00:02:57,498 A porta do trailer está aberta.
40 00:02:47,200 --> 00:02:49,600 Será que é tão difícil de entender isso?	Agora vemos o ambiente todo, de fora: é um velho trailer, com as paredes já gastas, meio enferrujadas. A porta do trailer está aberta. Sobre ele, uma placa onde está escrito: GRAN CIRCO AMERICANO.	
41 00:02:50,800 --> 00:02:52,200 Será que é tão difícil entender	RECADO (VS - cont) (engasgada, quase chorando) O que é que eu fiz da minha vida?	13 00:02:59,400 --> 00:03:02,816 Ruído no telefone.
42 00:02:52,300 --> 00:02:54,000 que eu só quero ter uma vida...		14 00:03:02,816 --> 00:03:06,232 A uns vinte metros do trailer, um velho orelhão, agora vazio.
43 00:02:57,500 --> 00:02:59,400 O que é que eu fiz da minha vida?	Ruído no telefone. Cai a ligação. Sinal de ocupado. A uns vinte metros do trailer, um velho orelhão, agora vazio.	

Figura 10: Exemplo de geração de roteiro AD (Trecho do roteiro do filme *3 Minutos*, 1999).

De acordo com a Figura 9, também é possível observar que os intervalos de tempo do roteiro AD são inversos aos da legenda, levando em consideração o tempo de início e de término do diálogo do roteiro. Além disso, as ações que estão presente no roteiro, mas não estão na audiodescrição foram eliminadas ou pelo componente de sumarização, por não conter alguma palavra mais frequente, ou na etapa de divisão de gaps, por não haver tempo disponível para sua narração.

4.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou uma solução para geração automática de audiodescrição para filmes, denominada *CineAD*, e suas principais características. Os componentes *Parser* de Roteiros, Identificador de *Gaps*, Sumarizador e Gerador do Roteiro de Audiodescrição foram descritos de forma detalhada, como também as estratégias empregadas na sua implementação. A solução proposta realiza a análise do roteiro original e da legenda do filme, retornando um roteiro com marcações de tempo contendo as descrições das ações.

O capítulo seguinte apresenta um conjunto de experimentos realizados para avaliar o *CineAD* com relação a qualidade da audiodescrição gerada e nível de compreensão dos usuários cegos.

5 EXPERIMENTOS

O experimento conduzido para avaliar a solução proposta será apresentado neste capítulo. O objetivo do experimento é avaliar a solução proposta sob duas perspectivas: de usuários com deficiência visual e de profissionais audiodescritores.

A primeira parte consiste da avaliação de filmes com audiodescrição gerada pela solução proposta por usuários deficientes visuais com relação ao nível de compreensão do conteúdo. Desta forma, é possível investigar se os usuários são capazes de compreender os filmes com a audiodescrição gerada pelo sistema e, como consequência, se é possível reduzir as suas barreiras de acesso à informação. A segunda parte consiste da avaliação do roteiro de audiodescrição gerado pelo sistema por audiodescritores para destacar os pontos positivos e negativos da solução sob a perspectiva de profissionais da área.

O processo envolvido no experimento envolveu as etapas de definição e planejamento (apresentadas no Apêndice B), análise estatística e interpretação dos resultados. As etapas de execução do experimento, de análise e interpretação dos resultados serão apresentadas nas Seções 5.2 e 5.3. Além destas etapas, foi também realizada uma análise técnica dos roteiros de audiodescrição para os filmes que foram utilizados durante as duas partes da avaliação. A análise técnica será descrita na Seção 5.1.

5.1 ANÁLISE TÉCNICA DOS ROTEIROS

Esta análise tem o objetivo de apresentar informações sobre os roteiros de audiodescrição gerados pelo sistema e que foram utilizados nas duas partes da avaliação. Os vídeos selecionados foram os seguintes: um curta-metragem (filme A), uma cena de um longa-metragem (filme B) e um curta-metragem (filme C). A Tabela 5 apresenta as características dos vídeos utilizados.

Tabela 5. Características dos vídeos de teste

Itens	Filme A	Filme B	Filme C
Duração (mm:ss)	05:29	01:43	05:43
nº de cenas	6	1	1
nº de frases de ações	35	16	83

n° de diálogos	22	2	11
% de diálogos no vídeo	23,8	12,6	32,7
% sem diálogos do vídeo	76,2	87,4	67,3

O filme A contém 6 cenas. Em seu roteiro, estão presentes 35 frases de ações e 22 diálogos. O tempo do vídeo ocupado por diálogos corresponde a 23,8%, o que implica que 76,2% do filme não apresenta falas. O filme B, por outro lado, contém uma única cena, 16 frases de ações e 2 diálogos no roteiro, que ocupam 12,6% do tempo do vídeo. Por fim, o filme C contém 1 cena, 83 frases de ações e 11 diálogos no roteiro. A duração dos diálogos corresponde a 32,7% do tempo do vídeo, restando 67,3% do tempo sem falas. Destaque-se que todos os vídeos apresentam mais informações visuais do que diálogos.

A geração automática do roteiro de audiodescrição dos três filmes utilizados nos testes resultaram nos dados mostrados na Tabela 6.

Tabela 6. Dados da geração do roteiro AD

Itens	Filme A	Filme B	Filme C
n° de sentenças candidatas	30	16	66
n° de sentenças selecionadas para a AD	20	15	45
n° de gaps sem falas detectados	22	6	50
n° de gaps descartados	12	4	45
n° de gaps preenchidos	10	2	5
% de audiodescrição no vídeo	45,41	85,48	44,63
Média de palavras por sentença	10,35	10,26	9,46

No filme A, o sumariador considerou 30 sentenças importantes dentro do roteiro. Destas sentenças candidatas, 20 foram inseridas no roteiro de audiodescrição. As demais foram descartadas por não haver tempo disponível para a narração. Com isso, foram selecionadas apenas aquelas com o maior *score* e que respeitam os limites do *gap* de acordo com o parâmetro de 3 palavras por segundo. Foram detectados 22 gaps, ou seja, 22 intervalos entre as falas da legenda. Na geração do roteiro de audiodescrição, 12 gaps foram descartados e apenas 10 gaps sofreram inserções de descrições em seus intervalos. Esse descarte acontece quando não existe tempo disponível para narrar a descrição associada, ou

quando não existem ações entre os diálogos descritas no roteiro cinematográfico, seguindo a ordem cronológica. Um exemplo comum deste segundo tipo de situação é quando um longo diálogo é articulado pausadamente pelo personagem, gerando lacunas entre as falas, porém sem a ocorrência de novas ações. O tempo total das descrições geradas no roteiro de AD consome cerca de 45,41% da duração do vídeo. As sentenças selecionadas tem o tamanho médio de 10,35 palavras.

No filme B, 16 sentenças foram escolhidas como candidatas e apenas uma foi excluída do roteiro de audiodescrição por ter o menor *score* e não ter o tempo disponível para a narração. Na geração do roteiro de AD, foram descartados 4 gaps e as sentenças selecionadas foram alocadas em 2 gaps. O tempo total da audiodescrição ocupa 85,48% do filme. As sentenças possuem o tamanho médio de 10,26 palavras.

Por fim, o filme C apresentou 66 sentenças candidatas, sendo 45 selecionadas para o roteiro de audiodescrição. Quanto aos *gaps*, 50 foram detectados, porém apenas 5 foram utilizados. O tempo total das descrições geradas no roteiro de AD consome cerca de 44,63% da duração do vídeo. As sentenças selecionadas tem o tamanho médio de 9,46 palavras.

Os roteiros de audiodescrição descritos nesta seção fizeram parte do experimento de avaliação da solução proposta. A análise e interpretação dos resultados obtidos em cada parte das avaliações serão apresentadas a seguir.

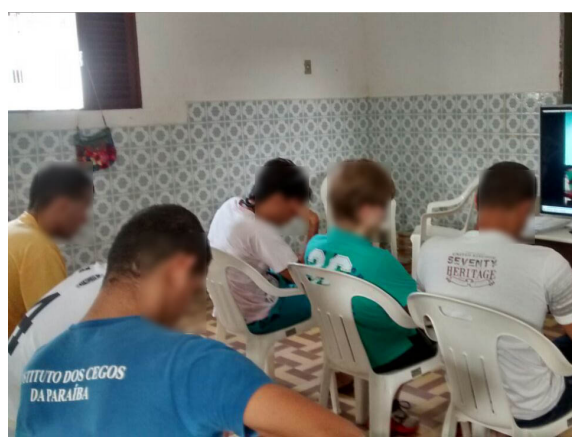
5.2 AVALIAÇÃO DO NÍVEL DE COMPREENSÃO DOS CONTEÚDOS

A primeira parte do experimento avalia o nível de compreensão dos conteúdos com usuários deficientes visuais. O experimento foi realizado nos dias 31 de Outubro de 2014 e 03 de novembro de 2014 no Instituto dos Cegos da Paraíba e contou com a participação de 12 pessoas com deficiências visuais (cegueira total ou parcial).

Estes usuários foram divididos em dois grupos para assistir os filmes selecionados para o experimento (apresentados na Seção B.2.1): um grupo para avaliar os filmes sem audiodescrição e outro para avaliar os filmes com a audiodescrição gerada pelo sistema. Os usuários foram instruídos a responder, inicialmente, a primeira parte do questionário sobre dados pessoais e em seguida assistir os filmes selecionados. Após a exibição de cada filme, os usuários foram convidados a responder as questões de avaliação referentes a cada conteúdo apresentado. Na Figura 11 são apresentadas algumas imagens do experimento.



(a)



(b)

Figura 11: Experimento com usuários. (a) Usuários avaliando o conteúdo com audiodescrição. (b) Usuários avaliando o conteúdo sem audiodescrição.

O grupo de usuários era formado por 10 homens e 2 mulheres classificados nas seguintes faixas etárias: 25% abaixo de 18 anos; 25% entre 18 e 30 anos; 25% entre 31 e 40 anos; e 25% acima de 40 anos. Quanto ao grau de deficiência visual, 33,3% dos usuários possuem cegueira total, enquanto que 66,7% possuem cegueira parcial. O grau de escolaridade dos usuários é apresentado na Figura 12.

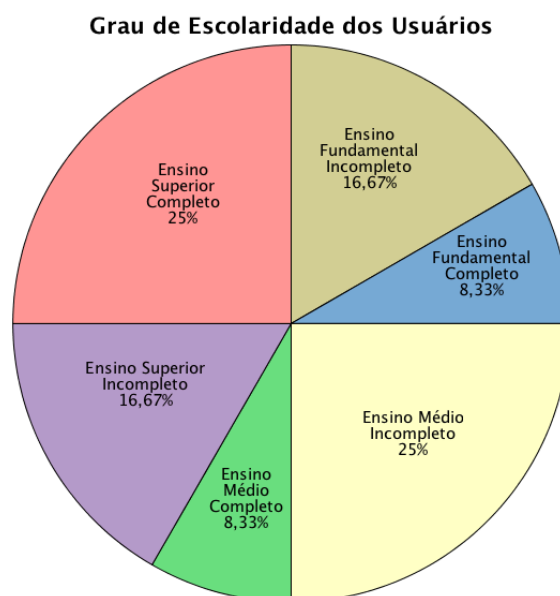


Figura 12: Grau de escolaridade dos usuários da avaliação.

Na Figura 13 são apresentados os resultados dos testes de nível de compreensão por vídeo/filme e no total. Os resultados dos testes do nível de compreensão dos filmes mostra que os usuários que assistiram o vídeo com a audiodescrição gerada pela solução proposta obtiveram uma média de acertos maior que os usuários do grupo sem a audiodescrição.

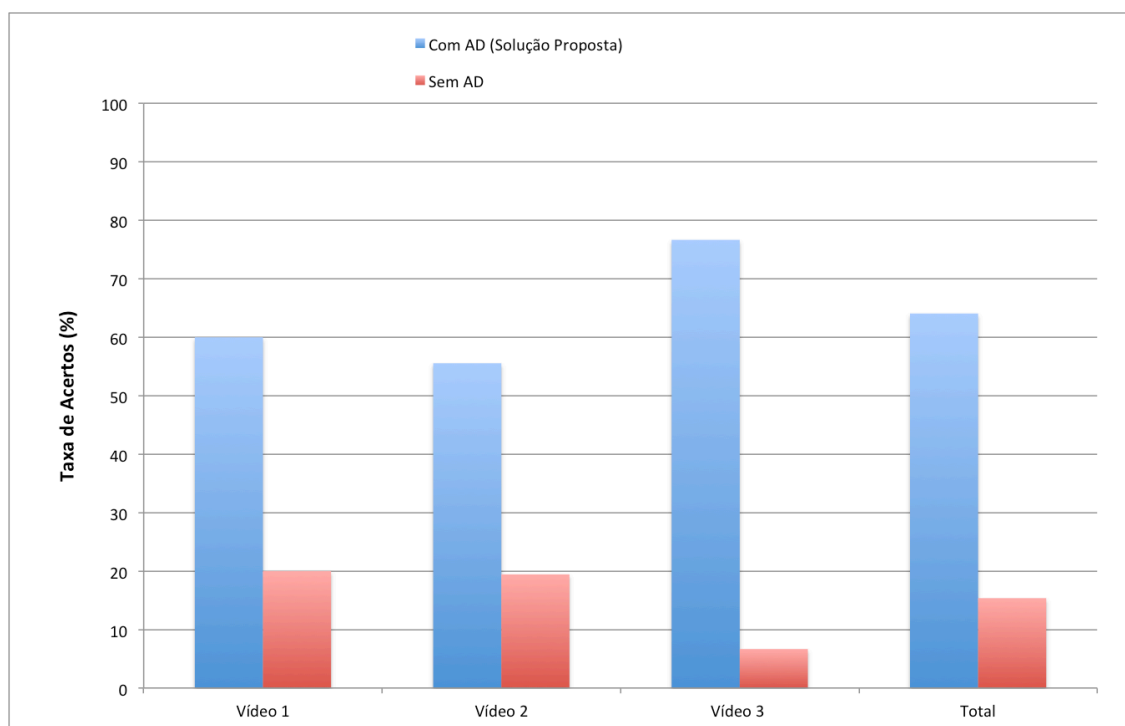


Figura 13: Resultados dos testes do nível de compreensão dos conteúdos

De acordo com a Figura 13, considerando todos os filmes, a taxa média de acertos do grupo que assistiu os vídeos sem audiodescrição foi de 15,37% com desvio padrão de 8,47%, enquanto que o grupo que assistiu esses conteúdos com a audiodescrição gerada pelo CineAD obteve uma taxa média de acertos de 64,07% com um desvio padrão de 18,37%. Nos três filmes exibidos, a audiodescrição gerada pela solução proposta obteve uma maior taxa de acertos.

A Figura 14 apresenta uma representação dos dados de resultados em um gráfico *Box Plot*. De acordo com a Figura 14, os usuários que assistiram os filmes com a audiodescrição gerada pela solução proposta obtiveram um melhor desempenho. Para o grupo com os filmes sem audiodescrição, os valores de mediana, do primeiro e terceiro quartil foram de 12,22%, 12,22% e 17,78%, respectivamente. Um *outlier* positivo foi identificado na amostra, representando um usuário que obteve 31,11% de acertos. Essa amostra não foi removida da análise porque a experiência pessoal de cada indivíduo com deficiência visual influencia diretamente na sua capacidade de percepção e, desta forma, é considerado válido para a avaliação.

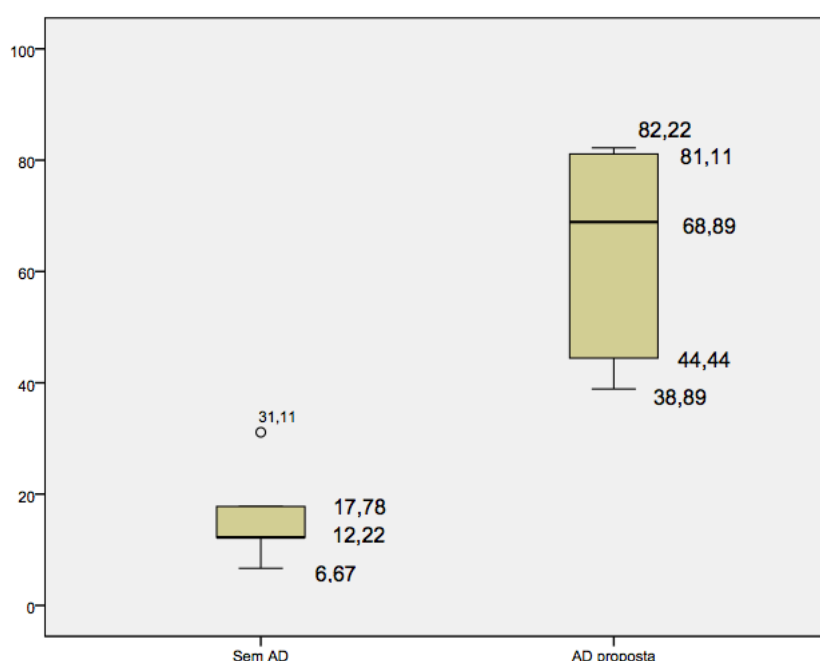


Figura 14: Gráfico *Box Plot* (caixa) dos testes do nível de compreensão dos conteúdos

Para o grupo que assistiu os filmes com audiodescrição gerada pelo CineAD, os valores de mediana, do primeiro e terceiro quartil foram, respectivamente, de 68,89%, 44,44% e 81,11%. Isso significa que 50% dos usuários obtiveram uma taxa de acerto entre

81,11% e 44,44%, com a menor taxa de acerto de 38,89% e a maior de 82,22%. Ademais, nenhum *outlier* foi identificado, o que significa que todos os usuários obtiveram taxa de acertos entre os valores 38,89% e 82,22%.

Além dos resultados já apresentado, foi aplicado também um teste-t para avaliar se a diferença de desempenho entre os grupos é representativa estatisticamente. Para isso, foi considerado um intervalo de confiança de 95% e 10 graus de liberdade (número total de usuários menos 2; $12 - 2 = 10$). Os resultados do teste-t são apresentado na Tabela 7.

Tabela 7. Resultado do teste-t para a Hipótese

Grau de Confiança (p-valor)	0,95
Grau de Liberdade	10
t-valor	5,896
$t_{0,095,10}$	2.228

O t-valor obtido para o teste foi de 5,896. De acordo com Wohlin *et al.* (2000), o valor crítico para o intervalo de confiança de 95% e 10 graus de liberdade é 2,228 ($t_{0,095,10}$). Uma vez que o t-valor obtido é maior que o valor crítico em questão, é possível concluir que a hipótese nula H_0 pode ser rejeitada. Com esta conclusão, pode-se afirmar que os filmes com a audiodescrição da solução proposta são mais bem compreendidos em relação aos filmes sem audiodescrição.

Além das questões sobre o nível de compreensão dos filmes, outros aspectos relacionados a solução proposta foram avaliados nessa primeira parte do experimento. No final do questionário, os usuários atribuíram conceitos, dentro de uma escala de 1 (Confuso) a 6 (Claro), sobre os seguintes aspectos:

- 1) Facilidade no entendimento dos filmes e
- 2) Auxílio da audiodescrição na compreensão dos filmes.

Este segundo aspecto não foi avaliado com o grupo que assistiu os vídeos sem audiodescrição. Os resultados desta avaliação são apresentados na Tabela 8.

Tabela 8. Resultados da avaliação

Aspectos Avaliados	Com AD da solução proposta		Sem AD	
	Média	Desvio Padrão	Média	Desvio Padrão
1. Facilidade no entendimento dos filmes	3,67	0,81	2,3	1,5
2. Auxílio da audiodescrição na compreensão dos filmes	4	0,89	—	—

O grupo com a audiodescrição gerada pelo solução avaliou a facilidade no entendimento dos filmes com um valor médio de 3,67 com desvio padrão de 0,81, enquanto que o grupo sem audiodescrição avaliou com um valor médio de 2,3 com desvio padrão de 1,5. Quanto ao auxílio da audiodescrição na compreensão dos filmes, o grupo que assistiu os vídeos com a audiodescrição gerada pelo sistema atribuiu o valor médio de 4 com um desvio padrão de 0,89.

De acordo com a Tabela 8, os valores médios obtidos sobre a facilidade no entendimento do filme entre os dois grupos diferem em apenas 1,37, apesar da diferença entre as taxas de acertos dos grupos avaliados (15,37% para o grupo sem audiodescrição e 64,07% para o grupo com audiodescrição gerada pelo sistema). De forma a analisar a compatibilidade entre esses dados, a correlação entre as seguintes variáveis foi calculada: a) taxa de acertos e b) grau de satisfação (facilidade no entendimento dos filmes). Os valores do coeficiente de correlação de *Pearson* e do coeficiente de correlação de postos de *Spearman* para as duas variáveis são, respectivamente, 0,353 e 0,265, o que representa uma fraca correlação entre as duas variáveis. Como consequência, é possível concluir que existe uma correlação fraca entre o nível de satisfação com a solução e o nível de compreensão dos conteúdos.

Segundo Wohlin *et al.* (2000), uma das possíveis explicações para esse resultado é que os humanos têm medo de serem avaliados e quando passam por um processo de avaliação tentam parecer melhor do que realmente são, o que pode atrapalhar a saída do experimento. Outra explicação é que alguns usuários possuem alguma resistência com relação ao uso de tecnologias assistivas e de sintetizadores de voz, que foram utilizados para narrar as descrições geradas neste experimento. Contudo, uma análise mais aprofundada precisa ser realizada para avaliar esses aspectos, e é uma proposta de trabalho futuro.

5.3 AVALIAÇÃO COM ESPECIALISTAS

A segunda parte do experimento consistiu da avaliação com especialistas (audiodescritores). Esta avaliação foi realizada através de um formulário *on line* que ficou com acesso disponível durante período de 15 a 31 de Dezembro de 2014. A divulgação da avaliação foi feita através de uma comunidade virtual privada de audiodescritores e contou com a participação de 5 audiodescritores. Na Tabela 9, são apresentados alguns dados gerais sobre os participantes, como faixa etária, tipo de conteúdo que audiodescreve, experiência em audiodescrição, entre outros.

Tabela 9. Dados gerais sobre os participantes

Participante	Faixa Etária	Estado	Tipo de conteúdo visual	Tempo de experiência
A	Acima de 45 anos	RJ	Filme, Teatro, Televisão, Seminários, Palestras.	de 5 a 10 anos
B	Entre 20 e 35 anos	PE	Filme, Palestras	de 1 a 5 anos
C	Entre 35 e 45 anos	RS	Filme, Teatro, Dança, Ilustrações, Museus	de 5 a 10 anos
D	Entre 35 e 45 anos	PE	Filme, Teatro, Ópera, Dança, Eventos Esportivos, Feiras de Arte, Casamentos	de 1 a 5 anos
E	Entre 20 e 35 anos	PB	Filme, Teatro	Menos de 1 ano

Todos os participantes informaram que tinham experiência na audiodescrição de filmes. Dentre os audiodescritores, dois utilizam o roteiro cinematográfico durante a criação da audiodescrição porque contém informações importantes acerca do filme e auxilia nas escolhas tradutórias. Os demais (3 participantes) não utilizam o roteiro devido à dificuldade de acesso e, normalmente, a necessidade de se ter contado direto com o diretor do filme, o que nem sempre é possível.

Os participantes foram convidados a avaliar em uma escala de 1 a 6 os seguintes questionamentos:

- 1) O roteiro de AD gerado pelo sistema aumenta a eficiência do processo de descrição?
- 2) Qual a classificação da adequação da velocidade da audiodescrição?

- 3) Quanto ao sincronismo, como a simultaneidade entre os eventos do filmes e as descrições é avaliada?
- 4) As lacunas entre os diálogos identificadas pelo sistema são satisfatórias?
- 5) Como você avalia a qualidade das descrições geradas?
- 6) O roteiro de audiodescrição contém informações relevantes e precisas?

Na Tabela 10, são apresentados os resultados desta avaliação.

Tabela 10. Resultados da avaliação com especialistas (Valores na escala de 1 a 6)

	Questionamento	Média	Desvio Padrão
1	Aumento na eficiência do processo de descrição	3,4	1,51
2	Classificação da adequação da velocidade da audiodescrição	4,2	1,48
3	Sincronismo entre eventos do filmes e as descrições	3,2	1,3
4	Lacunas entre os diálogos são satisfatórias	4,6	1,67
5	Qualidade das descrições geradas	2,8	1,3
6	Roteiro de AD com informações relevantes e precisas	3,2	1,79

Quanto ao aumento na eficiência do processo de descrição, o valor médio recebido na avaliação foi de 3,4 (dentro da escala de 1 a 6) com desvio padrão de 1,51. A maioria dos audiodescritores informaram que o roteiro gerado pela solução serve de suporte ao profissional por apresentar detalhes relevantes de acordo com a visão do roteirista do filme, mas concordam que necessita de alguma modificações antes da geração da versão final da audiodescrição. Um audiodescritor avaliou com a nota máxima por considerar que o roteiro aumenta bastante a eficiência no processo de criação, pois apresenta a “minutagem” e destaca elementos relevantes na narrativa. Entretanto, um avaliador atribuiu a nota 2 por considerar que o sistema descreve alguns elementos desnecessários.

Sobre a velocidade da audiodescrição, a média obtida foi de 4,2 com desvio padrão de 1,48. De forma geral, os avaliadores consideraram satisfatórias, mas identificaram que em alguns momentos as descrições estão longas, o que necessita acelerar o tempo da narração

para caber na lacuna disponível. Segundo eles, isto não chega a prejudicar o entendimento, mas torna a experiência um pouco desconfortável.

A média obtida na avaliação do sincronismo entre eventos do filmes e as descrições foi de 3,2 com desvio padrão de 1,3. Apenas um audiodescritor comentou este tópico, informando que em alguns pontos a audiodescrição se apresenta dessincronizada.

A identificação das lacunas (*gaps*) entre os diálogos recebeu o valor médio de 4,6 com desvio padrão de 1,67, sendo o quesito de maior nota da avaliação. Segundo os comentários dos avaliadores, a identificação dos *gaps* é o ponto principal do sistema.

Quanto à qualidade das descrições geradas, a média obtida foi de 2,8 com desvio padrão de 1,3. Os audiodescritores consideraram a qualidade variável com alguns trechos bons, porém com outros que incluem descrições desnecessárias, como por exemplo as descrições relacionadas a sons do áudio original do filme.

Em relação à presença de informações relevantes e precisas no roteiro gerado pela solução, os audiodescritores avaliaram com uma média de 3,2 com desvio padrão de 1,79. De forma geral, os comentários indicam que a solução proposta gera um roteiro de audiodescrição relevante, pois utiliza o roteiro original do filme como base. Entretanto, segundo os avaliadores, alguns elementos importantes são ignorados e podem gerar um retrabalho para o audiodescritor, que terá que editar e ajustar o roteiro antes de gerar a versão final da audiodescrição.

Além dos aspectos apresentados na Tabela 10, os usuários também responderam outros questionamentos sobre a utilidade do sistema dentro do processo de criação de audiodescrição e os seguintes resultados foram obtidos:

- 100% dos audiodescritores consideraram que as descrições geradas servem como referência durante o trabalho de criação da audiodescrição, porém algumas alterações no roteiro serão necessárias;
- 3 audiodescritores usariam o sistema com o objetivo de agilizar o trabalho e o recomendariam para outros profissionais;
- 2 audiodescritores não utilizariam o sistema por considerar uma mecanização da atividade e por conter certas limitações. Por estes motivos, não recomendariam para outros profissionais.

Por fim, os audiodescritores citaram pontos do sistema que poderiam melhorar, além de outras funcionalidade que gostariam de ver na solução. Essas sugestões estão listadas a seguir:

- Excluir informações sobre efeitos sonoros óbvios presentes no áudio original do filme;
- Melhorar a escolha dos descritivos contidos no roteiro;
- Utilizar o tempo verbal no presente;
- Ajustar a sincronia entre a ocorrência dos eventos do filme e o momento da narração da descrição;
- Ajustar a velocidade da narração para evitar mudanças bruscas;
- Possibilitar a edição do roteiro gerado pelo sistema com opções de ajuste do texto das descrições, tempo e velocidade;
- Possibilitar que o audiodescritor grave o áudio (voz humana) da narração durante a edição do roteiro;

5.4 CONSIDERAÇÕES FINAIS

Este capítulo apresentou o experimento de avaliação do sistema proposto neste trabalho. A avaliação envolveu a definição, planejamento e execução de um experimento em duas partes: com usuários deficientes visuais e com profissionais audiodescritores.

A primeira parte envolveu usuários com deficiências visuais para testar a hipótese sobre o nível de compreensão de usuários cegos com filmes audiodescritos pelo sistema apresentada no planejamento (Apêndice B). A segunda parte foi a avaliação feita com especialistas para identificar os pontos positivos e negativos da solução proposta, sob a perspectiva de um profissional.

Os resultados do experimento mostraram que a hipótese nula formulada pôde ser rejeitada. Isso indica que o sistema conseguiu melhorar o nível de compreensão dos filmes em relação aos filmes sem a audiodescrição, que a situação mais corriqueira atualmente. Os especialistas consideraram que o roteiro de audiodescrição gerado pela solução proposta serve como referência, porém necessita de alterações.

O capítulo a seguir apresenta as conclusões sobre o presente trabalho, como também as propostas de trabalhos futuros.

6 CONCLUSÃO

Este trabalho apresentou um sistema de geração automática de roteiro de audiodescrição para filmes, o CineAD. Tal solução contempla recursos de acessibilidade, fornecendo a descrição de conteúdos visuais para as pessoas cegas ou com baixa visão.

Alguns trabalhos, como Gagnon et al. (2009) e Lakritz & Salway (2002) foram propostos com o propósito de gerar automaticamente a audiodescrição de filmes, mas possuem algumas lacunas a serem resolvidas. Diante disso, a abordagem que o sistema CineAD utiliza envolve a análise do roteiro original cinematográfico e da legenda do filme, retornando um roteiro com marcações de tempo contendo as descrições das ações.

Para validar a solução proposta, um experimento foi executado de forma a obter a avaliação sob duas perspectivas: pessoas cegas ou com baixa visão e audiodescritores, com o objetivo de analisar a visão do usuário final do produto visual e do profissional que trabalha na elaboração deste recurso de acessibilidade.

Os resultados mostram que os usuários que assistiram os filmes com audiodescrição gerada pelo sistema CineAD obtiveram uma taxa de acertos de 64,07%, enquanto que os usuários que assistiram os filmes sem a audiodescrição atingiram 15,37% de acertos. A partir desses dados, pode-se afirmar que a solução proposta reduziu as barreiras de acesso enfrentada por pessoas com deficiência visual, aumentando o nível de compreensão dos filmes.

A avaliação especializada com audiodescritores mostrou que o roteiro de audiodescrição gerado pelo sistema contém informações relevantes e precisas de forma razoável. De forma geral, os audiodescritores destacaram a identificação dos *gaps* entre diálogos como principal funcionalidade do sistema e consideraram que as descrições geradas servem como referência, entretanto é preciso modificar o roteiro antes de gerar a narração da audiodescrição. Durante a análise dos resultados, também foi observada uma certa resistência dos audiodescritores à proposta de geração automática do roteiro de audiodescrição, apesar de não ser verificado explicitamente na avaliação. Entretanto, é necessário realizar uma análise mais aprofundada para verificar se existe essa resistência e quais as possíveis razões para isso.

O CineAD tem potencial para gerar as descrições dos eventos mais importantes do filme. De forma alternativa, a solução pode ser incorporada a uma ferramenta de sintetização de voz ou utilizada por um audiodescritor para a criação do áudio com as descrições.

Acreditamos que uma taxa média de 64,07% de acertos seja um resultado satisfatório para uma primeira implementação, e, dessa forma, que a solução pode efetivamente facilitar a vida das pessoas com alguma limitação visual.

6.1 TRABALHOS FUTUROS

A partir da análise dos resultados obtidos e da avaliação com os especialistas, consideramos como propostas de trabalhos futuros, a incorporação de um componente de compressão de sentenças na etapas de sumarização para reduzir o tamanho das frases. Tal incorporação facilitaria a inserção das descrições e, consequentemente, permitiria maiores ajustes de tempo para a narração da audiodescrição.

Além disso, uma possível proposta seria uma adaptação do estilo das sentenças do roteiro, que normalmente possuem um caráter mais narrativo, para um estilo descritivo adequado à audiodescrição.

Por fim, a criação de regras ou heurísticas para a remoção de elementos do roteiro que estão relacionados a sons óbvios presentes no filme poderão otimizar a descrição, uma vez que o áudio é perceptível aos cegos e, em sua grande parte, não necessitam ser descritos.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. ASSOCIAÇÃO BRASILEIRAS DE NORMAS TÉCNICAS. **NBR 9050:** acessibilidade de pessoas portadoras de deficiências a edificações, espaço, mobiliário e equipamento urbanos. Rio de Janeiro, 1994.

ALVES et al. **Propostas para um modelo brasileiro de audiodescrição para deficientes visuais.** Revista Brasileira de Tradutores Tradução & Comunicação N°.22, Ano 2011. São Paulo: Anhanguera Educacional Ltda. Disponível em: <<http://sare.anhanguera.com/index.php/rtcom/article/view/3158/1215>> Acesso em: 23 out. 2014.

ANCINE/OCA. Disponível em: <http://issuu.com/oca_ancine/docs/anu__rio__estat__stico__do__cinema__bra> Acesso em: Jan. 2015.

ARAÚJO, V. L. S. **A Formação de Audiodescritores no Ceará e em Minas Gerais: Uma Proposta Baseada em Pesquisa Acadêmica.** In: Livia Maria Villela de Mello Motta, Paulo Romeu Filho. Audiodescrição: Transformando Imagens em Palavras. São Paulo: Secretaria de Estado dos Direitos da Pessoa com Deficiência, 2010.

ARAÚJO, V. L. S. O processo de legendagem no Brasil. Revista do GELNE, Fortaleza, v. 1/2, n. 1, p. 156-159, 2006.

BENECKE, B. Audio-description. In: GAMBIER, Y. (ed.), 2004. Meta. Volume 49, nº.1, p. 78-80.

BRAGA, K. B. *Cinema Acessível para Pessoas com Deficiência Visual: a Audiodescrição de O Grão de Petrus Cariry.* 2011. 153 f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada do Centro de Humanidades, Universidade Estadual do Ceará, 2011.

BRASIL. Lei nº 10.098, de 19 de dezembro de 2000. Estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências. Diário Oficial da União, Brasília, DF, 19 dez. 2000.

BRASIL. Portaria nº 188, de 24 de março de 2010. Estabelece prazos para recursos de acessibilidade para pessoas com deficiência, na programação veiculada nos serviços de radiodifusão de sons e imagens e de retransmissão de televisão,. Diário Oficial da União, Brasília, DF, 19 dez. 2010.

CAMPOS, Flavio de. Roteiro de cinema e televisão: a arte e a técnica de imaginar, perceber e narrar uma história. 2ª ed. Rio de Janeiro: Jorge Zahar Editor, 2009.

CELTX. Disponível em: <<https://www.celtx.com/index.html>> Acesso em: Jan. 2014.

CHAPDELAINE, C.; GAGNON, L. Accessible Videodescription On-Demand. ASSETS '09, Pennsylvania, p 221-222, oct. 2009.

DEMIR, S.; OLIVER, D.; SCHWARTZ, E. Interactive SIGHT Demo: Textual Summaries of Simple Bar Charts. ASSETS '10, Florida, p. 267-268, oct. 2010.

DIAS, C.. **Usabilidade na web**: Criando portais mais acessíveis. 2a edição Rio de Janeiro: Alta Books, 2007. 296 p.

DULYAN, A.; EDMONDS, E. AUXie: Initial Evaluation of a Blind-Accessible Virtual Museum Tour. OZCHI, Brisbane, p. 272-275, nov. 2010.

EDMUNDSON, H. P. (1969) "New Methods in Automatic Extracting", In: Journal of the ACM, Vol 16, Issue 2, Pages 264-285, 1969.

ENCELLE, B.; BELDAME, M. O.; PRIÉ, Y. Towards the usage of pauses in audio described videos. 22nd International World Wide Web Conference, Rio de Janeiro, may 2013.

FERRES, L.; SEPÚLVEDA J. F. Improving Accessibility to Mathematical Formulas: The Wikipedia Math Accessor. W4A '11, Andhra Pradesh, mar. 2011

FIELD, Syd. Manual do Roteiro: os fundamentos do texto cinematográfico. Rio de Janeiro: Objetiva, 2001.

FLOR, C. S. **Diagnóstico da acessibilidade dos principais museus virtuais disponíveis da internet**. 129 f. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.

FRANCO, E. P. C., SILVA, M. C. C. C., **Audiodescrição: Breve passeio histórico**. In: Lívia Maria Villela de Mello Motta, Paulo Romeu Filho. Audiodescrição: Transformando Imagens em Palavras. São Paulo: Secretaria de Estado dos Direitos da Pessoa com Deficiência, 2010.

GAGNON, Langis et al. **Towards computer-vision software tools to increase production and accessibility of video description for people with vision loss**. Springer-Verlag, p 199-218, feb. 2009.

IBGE. Censo Demográfico 2010: Características gerais da população, religião e pessoas com deficiência. **IBGE- Instituto Brasileiro de Geografia e Estatística**, 2010. Disponível em:
<http://www.ibge.gov.br/home/estatistica/populacao/censo2010/caracteristicas_religiao_deficiencia/caracteristicas_religiao_deficiencia_tab_pdf.shtm>. Acesso em: 28/01/2014.

JAIN, R. **The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling**, Willey, Nova York, EUA, 1991.

JENA. Apache Jena 2.0. Disponível em: < <https://jena.apache.org/> > Acesso em: Jan. 2014.

KOBAYASHI, M.; NAGANO, T.; FUKUDA, K.; TAKAGI, H. **Describing Online Videos with Text-to-Speech Narration**. W4A2010, Raleigh, apr. 2010.

LAKRITZ, J.; SALWAY, A. “**The Semi-Automatic Generation of Audio Description from Screenplays**”, Dept. Of Computing Technical Report CS-06-05, 2002, University of Surrey.

LUCENE. Apache lucene 2.0. Disponível em: <<http://lucene.apache.org/>> Acesso em: Jan. 2014.

MOSKOVITCH, Y.; WALKER, B. N. Evaluating Text Descriptions of Mathematical Graphs. ASSETS '10, Florida, p. 259-260, oct. 2010.

MOTTA, Livia Maria Villela de Mello. Audiodescrição: recurso de acessibilidade para a inclusão cultural. Disponível em <<http://www.planetaeducacao.com.br/portal/artigo.asp?artigo=1210>>. Acesso em 11 ago. 2014.

MUNIZ, R. C. (2014), Uma Revisão Sistemática da Literatura sobre Técnicas de Geração Automática e Semiautomática de Audiodescrição. Monografia (Bacharelado em Ciências da Computação), Universidade Federal da Paraíba.

MURPHY, E.; BATES, E.; FITZPATRICK, D. Designing Auditory Cues to Enhance Spoken Mathematics for Visually Impaired Users. ASSETS '10, Florida, p. 75-82, oct. 2010.

NENKOVA, A.; McKEOWN, K. **Automatic Summarization..** In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts of ACL 2011* (HLT '11). Association for Computational Linguistics, Stroudsburg, PA, USA, , Article 3 , 86 pages.

NUNES, E. V.; MACHADO, F. O. e VANZIN, T. (2011). **Audiodescrição como Tecnologia Assistiva para o Acesso ao Conhecimento por Pessoas Cegas**. Ambiente virtual de aprendizagem inclusivo / organizadores: Vania Ribas Ulbricht, Tarcísio Vanzin e Vilma Villarouco. – 352 p. Florianópolis : Pandion, 2011.

O'CONNELL, T.; KOBAYASHI, M.; GOULD, B.; TAKAGI, H.; ASAKAWA, C. Are Synthesized Video Descriptions Acceptable? ASSETS '10, Florida, p. 163-170, oct. 2010.

OKTAY, B.; FOLMER, E. Syntherella: A Feedback Synthesizer for Efficient Exploration of Virtual Worlds using a Screen Reader. Graphics Interface Conference, Newfoundland, p. 65-70, may 2011.

QUEVEDO, S. R. P. de; ULBRICHT, V. R. Como os cegos aprendem. **Comunicação, tecnologia e cultura de rede**. São Paulo: Pandion Editoria, 2011.

SALWAY, A., VASSILIOU, A. e AHMAD, K. (2005). Whats Happens in Films? In Procs. IEEE International Conference on Multimedia and Expo, ICME 2005.

SANTANA, M. **A Primeira Audiodescrição na Propaganda da TV Brasileira: Natura Naturé Um Banho de Acessibilidade.** In: Livia Maria Villela de Mello Motta, Paulo Romeu Filho. Audiodescrição: Transformando Imagens em Palavras. São Paulo: Secretaria de Estado dos Direitos da Pessoa com Deficiência, 2010.

THOMAN, K. E.; SRIPADA, S.; NOORDZIJ, M. L. Atlas.txt: exploring linguistic grounding techniques for communicating spatial information to blind users. Springer-Verlag, p. 85-98, dec. 2010.

TORRES, E. F., MAZZONI, A. A., MELLO, A. G. **Nem toda pessoa cega lê Braille nem toda pessoa surda se comunica em língua de sinais**, São Paulo: Educação e Pesquisa, v. 33, n. 2, mai/aug 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-97022007000200013>. Acesso em: 31 ago. 2014.

WANG, Z.; LI, N.; LI, B. Fast and independent access to map directions for people who are blind. Elsevier B.V., Arizona, p. 91-106, 2012.

WOHLIN, C., P. RUNESON, M. HOST, M. C. OHLSSON, B. REGNELL & A. WESSLÉ. **Experimentation in Software Engineering: An Introduction**, Kluwer Academic Publisher, Norwell, EUA, 2000.

ANEXO A - PROTOCOLO DE REVISÃO SISTEMÁTICA

Conforme mencionado na Seção 2.2, a definição do protocolo de revisão sistemática é a primeira etapa necessária para se realizar uma revisão sistemática. Esse protocolo consiste em um plano para conduzir a revisão de modo a fornecer uma visão geral de uma área de pesquisa e permiti identificar, quantificar e analisar os tipos de pesquisas e os resultados obtidos.

Primeiramente faz-se necessário definir quais são as “questões de pesquisa” desse estudo. Considerando que o tema dessa revisão sistemática são os sistemas de geração automática e semiautomática de audiodescrição, as questões de pesquisa formuladas foram as seguintes:

- QP 1: Como estão distribuídas as publicações envolvendo geração automática e semiautomática de audiodescrição em Tecnologias da Informação e Comunicação (TIC)?
- QP 2: Quais os métodos de pesquisa utilizados nestes trabalhos e quais as principais estratégias utilizadas para abordar esse problema?
- QP 3: Quais são as formas de avaliação comumente utilizadas para analisar a qualidade de técnicas de geração automática e semiautomática de audiodescrição?

A pesquisa sobre trabalhos relacionados foi realizada através de um processo de busca automática sobre os principais portais (bibliotecas digitais) de artigos científicos, listados abaixo:

- IEEEExplore (<http://ieeexplore.ieee.org>)
- ACM Digital Library (<http://dl.acm.org>)
- ScienceDirect (<http://www.sciencedirect.com/>)
- Scopus (<http://www.scopus.com/>)
- ISI Web of Knowledge (<http://isiwebofknowledge.com/>)

Para realizar uma busca automática nesses sites, pode-se pesquisar de forma mais precisa considerando exclusivamente palavras chave, título e resumo dos artigos de modo a montar uma “*string* de busca” com as palavras chave, evitando assim que as palavras chave

aparecem soltas nos corpos dos artigos pesquisados, reduzindo a probabilidade de encontrar artigos irrelevantes. Dessa forma, para definir uma “*string* de busca” é necessário então identificar os “termos de pesquisa” que a comporão, ou seja, as palavras chaves, seus sinônimos e termos relacionados. Os termos de pesquisa, juntamente com seus sinônimos são então combinados utilizando operadores lógicos aceitos pelos portais (por exemplo, os operadores OR (ou), AND (e) e NOT (não)) para formar a *string* de busca. As palavras chave, sinônimos e termos utilizados para a composição da *string* de busca definida nesse trabalho são apresentados na Tabela 1.

Tabela 1: Termos de busca

Palavras chave	Sinônimos e termos relacionados
Audio description	Visual description, vídeo description
Automatic	
Semi automatic	
Accessibility	Assistive technology
Generation	Synthesys

Definida a lista de palavras chaves, termos relacionados e sinônimos, foi definida então a seguinte “string de busca”:

- String de busca: (synth* OR generat* OR autom* OR semi automat*) AND (audio description OR audiodescription OR video description OR visual description) AND (accessibility OR assistive technology)

Além da busca automática realizada nos portais listados, foi realizada uma busca manual com objetivo de encontrar um maior número de trabalhos a serem analisados. Essa busca foi realizada levando em consideração a análise de artigos referenciados em outros trabalhos filtrados na primeira interação, ou seja, que haviam sido analisados e estavam de acordo com o tema proposto para este trabalho. O resultado dessa etapa de busca manual foi a pré-seleção de três artigos que foram incluídos no processo de filtragem.

Depois de realizado o processo de busca, os documentos recuperados são filtrados com base nos seguintes critérios de inclusão e exclusão dos trabalhos:

- I1: Incluir todos os artigos completos relacionados ao contexto da revisão (através da leitura do título e do resumo e, em caso de informações insuficientes, da leitura completa do artigo);
- E1: Excluir todos os documentos que não são artigos científicos;
- E2: Excluir documentos com conteúdos irrelevantes para a pesquisa;
- E3: Excluir os trabalhos repetidos. Caso existam trabalhos repetidos, incluir apenas o trabalho mais recente;
- E4: Excluir trabalhos em que o acesso ao artigo seja incompleto, apresentando apenas o resumo (abstract), por exemplo. Após esse processo de inclusão e exclusão, os artigos selecionados são classificados com base nos seguintes critérios:

1. Tipo de geração de áudio descrição: Classificação enquanto a automatização do processo de geração de audiodescrição. Considerando esse critério, os artigos podem ser classificados da seguinte forma:

- (a) Automático: geração automática de audiodescrição;
- (b) Manual: geração manual de audiodescrição;
- (c) Semiautomático: geração semiautomática de audiodescrição;
- (d) Não classificado: o artigo não especifica o tipo de geração de audiodescrição.

2. Tipo de entrada: Tipo de entrada utilizado pelo sistema para gerar a audiodescrição. Considerando esse critério, os artigos podem ser classificados da seguinte forma:

- (a) Roteiro: geração de audiodescrição a partir de roteiro;
- (b) Texto: geração de audiodescrição a partir de texto;
- (c) Ambiente 3D: geração de audiodescrição a partir de ambientes tridimensionais;
- (d) Dados estatísticos: geração de audiodescrição a partir de dados estatísticos.
- (e) Vídeo: geração de audiodescrição a partir de vídeo;
- (f) Roteiro: geração de audiodescrição a partir de roteiros ou scripts;
- (g) Dados matemáticos: geração de áudio descrição a partir de dados matemáticos;
- (h) Não classificado: o artigo não especifica o tipo de entrada.

3. Tipo de saída: Tipo de saída gerado pelo sistema de áudio descrição. Os tipos mais utilizados são:

- (a) Texto: gera como saída do sistema, texto;
- (b) Áudio: gera como saída do sistema, áudio;
- (c) Impressão tátil: gera como saída do sistema, impressão tátil;
- (d) Gráfico: gera como saída do sistema, gráfico;
- (e) Vídeo: gera como saída do sistema, vídeo;
- (f) Roteiro: gera como saída do sistema, roteiro;
- (g) Não classificado: o artigo não especifica o tipo de saída.

4. Canal de geração do áudio: Qual canal é utilizado como meio de transporte dos dados relativos à saída do sistema de áudio descrição. Os tipos mais utilizados são:

- (a) Canal original: o áudio é embutido junto com a trilha sonora original (o narrador descreve os eventos que ocorrem na tela entre pausas naturais);
- (b) Canal separado: existe outro canal de áudio que produz o som produzido pelo processo de áudio descrição;
- (c) Não utiliza: não utiliza áudio como saída do sistema;
- (d) Não classificado: utiliza áudio como saída do sistema, mas não exemplifica o canal de geração.

5. Domínio da audiodescrição: Identifica se o sistema de audiodescrição foi projetado para qualquer tipo mídia (teatro, filmes, cinema, televisão, museu, etc) ou se foi projetado para uma aplicação em um domínio específico. Com base no que foi dito os artigos podem ser classificados como:

- (a) Domínio geral: audiodescrição para qualquer tipo de mídia;
- (b) Domínio específico: audiodescrição para um domínio específico;
- (c) Não classificado: O domínio da áudio descrição não é identificado ou definido pelo artigo.

6. Método de avaliação: Identifica qual o método de avaliação utilizado no processo de áudio descrição para avaliar seu desempenho, viabilidade e qualidade do áudio gerado pelo processo. Nesse contexto, a avaliação pode ser:

- (a) Automática: somente utiliza medidas objetivas obtidas a partir de testes computacionais;
- (b) Com usuários: avaliação somente subjetiva através de testes realizados com usuários do sistema;
- (c) Híbrida: avaliação tanto automática quanto com avaliação com usuários;
- (d) Especializada: avaliação feita por especialistas em audiodescrição;
- (e) Não classificado: o tipo de avaliação não é definido pelo artigo.

APENDICE A – LISTA DE PALAVRAS FREQUENTES

LISTA DAS PALAVRAS MAIS FREQUENTES EM ROTEIROS DE
AUDIODESCRIÇÃO (adaptado de Salway, Vassiliou e Ahmad (2005))

HOMEM
CABEÇA
ROSTO
OLHOS
MÃO
MÃOS
HOMENS
MULHER
OLHA
TORNA
VIRA
LEVA
ANDA
VAI
FICA
PASSOS
SORRI
OLHARES
COLOCA
ASSISTE
ABRE
PROCURANDO
PORTA
QUARTO
CARRO
JANELA
MESA
ÁGUA
CAMA
CASA

APENDICE B – PLANEJAMENTO DOS EXPERIMENTOS

Nesse apêndice serão apresentados a definição e planejamento do experimento, com a definição do propósito, o contexto, a formulação das hipóteses, definição das variáveis, sujeitos, objetos e instrumentos do experimento, dentre outros.

B.1 DEFINIÇÃO DO EXPERIMENTO

O objetivo deste experimento é avaliar a solução proposta em duas partes: 1) Nível de compreensão dos conteúdos gerados pelo sistema por pessoas deficientes visuais e 2) Avaliação com especialistas. Cada parte do experimento será descrita a seguir.

B.2 AVALIAÇÃO DO NÍVEL DE COMPREENSÃO DOS CONTEÚDOS

Os elementos que caracterizam a avaliação do nível de compreensão foram definidos da seguinte forma:

- **Objeto de estudo:** O objeto de estudo são filmes com audiodescrição gerada pela solução proposta.
- **Propósito:** O propósito é avaliar os conteúdos gerados pela solução proposta com relação a capacidade de compreensão das pessoas deficientes visuais. Para isso, foi avaliado o nível de compreensão dos conteúdos acessíveis gerados pela solução proposta em comparação com o nível de compreensão de conteúdos sem audiodescrição (apenas o filme).
- **Foco de qualidade:** O principal efeito estudado pelo experimento é o nível de compreensão dos conteúdos.
- **Perspectiva:** A partir do ponto de vista de pessoas cegas (total ou parcial) brasileiras.
- **Contexto:** Essa parte do experimento foi conduzida em um ambiente controlado no Instituto dos Cegos da Paraíba. Os sujeitos do experimento foram pessoas com deficiências visuais (cegueira total e parcial) brasileiros e os objetos foram filmes cinematográficos, de diferentes gêneros, gerados em duas versões diferentes: (1) com audiodescrição gerada pela solução proposta e (2) sem audiodescrição. Os usuários foram divididos em dois grupos de forma que cada grupo avaliasse apenas uma das versões acessíveis de cada um dos conteúdos.

B.2.1 Planejamento do Experimento

Nessa seção será apresentado o contexto do experimento, as hipóteses que foram testadas, os sujeitos, variáveis e objetos utilizados no experimento, dentre outros aspectos do experimento.

Seleção do Contexto

Essa parte do experimento foi conduzida no Instituto dos Cegos da Paraíba com um conjunto de 12 pessoas com deficiências visuais (cegos totais e parciais) selecionados por conveniência. Estes usuários foram convidados a avaliar três conteúdos multimídia gerados em duas versões diferentes: (1) com audiodescrição gerada pela solução proposta ou (2) sem audiodescrição. No planejamento inicial deste trabalho foi inserido uma terceira versão a ser avaliada com a audiodescrição gerada por um profissional. No entanto, devido à escassez de profissionais na área e, consequentemente, à dificuldade de encontrar audiodescritores para colaborar com o trabalho, a avaliação consistiu das duas versões citada anteriormente. Os usuários foram divididos em dois grupos de forma que cada usuário avaliasse apenas uma das versões dos conteúdos. A amostra de conteúdos consistia de três vídeos com filmes ou cenas de filmes de gêneros variados. Os vídeos eram curtos para que a avaliação não fosse cansativa e não influenciasse o resultado da avaliação.

Formulação das Hipóteses

Informalmente, a hipótese formulada nessa parte do experimento foi a seguinte:

1. Os conteúdos acessíveis baseados na solução proposta neste trabalho tornam os conteúdos mais bem compreendidos pelos cegos do que os conteúdos sem audiodescrição.

A partir da definição informal, foi possível derivar uma definição formal da hipótese e definir quais seriam as medidas necessárias para avaliá-la.

1. **Hipótese nula H_0 :** Os conteúdos acessíveis baseados na solução proposta (SP) são tão bem compreendidos pelos usuários quanto os conteúdos sem audiodescrição

(semAD), onde o nível de compreensão dos conteúdos foi medido utilizando um conjunto de perguntas sobre os conteúdos avaliados.

$$\mu_{compreensao}(SP) = \mu_{compreensao}(semAD) \quad (D.1)$$

Hipótese alternativa H1: Os conteúdos acessíveis baseados na solução proposta (SP) são mais bem compreendidos do que os conteúdos sem audiodescrição (semAD).

$$\mu_{compreensao}(SP) > \mu_{compreensao}(semAD) \quad (B.2)$$

Seleção das Variáveis

A variável independente desse estudo é o tipo de conteúdo acessível avaliado pelo usuário. Ela é uma variável nominal que pode assumir os seguintes valores: (SP) filmes acessíveis com audiodescrição gerada pela solução proposta ou (semAD) filmes sem audiodescrição.

As variáveis dependentes são as variáveis de resposta aos diferentes valores (tratamentos) assumidos pela variável independente. Neste estudo, a variável dependente é o nível de compreensão dos conteúdos.

Seleção dos Sujeitos

Os sujeitos desse experimento são doze usuários com deficiência visual (cegos total ou parcial) voluntários e escolhidos por conveniência. Os usuários são membros do Instituto dos Cegos da Paraíba e possuem uma diversificada faixa etária e diferentes graus de escolaridade.

Projeto do Experimento

Segundo os padrões de projetos de experimentos apresentados por Wohlin et al. (2000), a definição, hipótese e as medidas indicam que o projeto desta parte do experimento deve ser do tipo um fator com dois tratamentos (do inglês, *one factor with two treatments*), onde o fator seria o tipo de conteúdo acessível e os tratamentos seriam as possíveis escolhas para esse fator: SP ou semAD.

Os princípios gerais de projeto de um experimento foram explorados da seguinte forma:

- **Aleatorização:** Neste estudo, cada usuário (sujeito) foi atribuído aleatoriamente para cada tratamento (SP ou semAD). Isso implica que cada sujeito avaliou apenas uma versão de cada um dos conteúdos, evitando influências da outra versão. Além disso, todos os objetos (conteúdos multimídia) foram avaliados por todos os usuários e a ordem de apresentação dos conteúdos foi a mesma para todos os usuários.

- **Balanceamento:** O experimento foi balanceado, uma vez que cada tratamento foi avaliado pelo mesmo número de usuários.

- **Bloqueio:** Nenhuma abordagem sistemática de bloqueio foi aplicada.

O teste estatístico aplicado foi o Teste t de *Student* (Jain 1991), um dos testes paramétricos mais utilizados na literatura científica para comparar dois tratamentos.

Instrumentação

Segundo o Anuário Estatístico do Cinema Brasileiro (ANCINE/OCA 2013), no ano de 2013, foram lançados filmes dos seguintes gêneros: 81,1% de ficção; 14,4% documentário; 4,5% animação. Diante disso, os vídeos selecionados como objetos deste experimento foram três filmes que pertencem ao gênero Ficção, uma vez que representam a grande maioria dos filmes lançados no último ano. A Tabela C.1 apresenta os vídeos que foram utilizados nos testes.

Tabela B.1: Vídeos utilizados nos testes

Vídeo	Duração	Descrição
Vídeo 1	5m29s	Esse vídeo é um curta-metragem do gênero comédia.
Vídeo 2	1m43s	Esse vídeo é parte de um longa-metragem e contém uma cena do filme.
Vídeo 3	5m43s	Esse vídeo é um curta-metragem do gênero drama.

Conforme pode ser observado na Tabela D.1, os vídeos são curtos para evitar que os testes sejam muito longos e os usuários se entediem durante a avaliação, influenciando na avaliação dos conteúdos apresentados.

Os dados dessa avaliação foram coletados através de questionários. O questionário usado era composto de três partes e é apresentado no Apêndice E. Na primeira parte do questionário, o usuário fornece informações pessoais como sexo, idade, grau de escolaridade, dentre outros. Na segunda parte, ele responde a 16 (dezesesseis) perguntas sobre os vídeos apresentados para avaliar o nível de compreensão dos conteúdos. Por fim, na terceira parte, ele avalia subjetivamente os conteúdos apresentados com relação ao nível de compreensão.

Para verificar o nível de compreensão, todas as questões apresentavam quatro alternativas (A, B, C ou D) relacionadas ao conteúdo apresentado, onde apenas uma das alternativas é correta. Para todas as questões, a quarta alternativa (D) representava a opção "Não sei responder", que foi incluída para evitar que os usuários escolhessem aleatoriamente uma das alternativas quando não soubessem a resposta correta. Os usuários foram orientados a selecionar apenas uma alternativa e, quando desconhecessem a resposta correta, selecionar a alternativa D. A terceira parte do questionário, o usuário avaliava os conteúdos com relação ao nível de compreensão dentre outros, através de duas questões que utiliza uma escala de 1 a 6.

Os questionários foram aplicados em 3 formatos: versão em Braille, versão com letra aumentada e entrevista oral.

B.3 AVALIAÇÃO COM ESPECIALISTAS

A segunda parte dos experimentos consistiu da avaliação do roteiro de audiodescrição gerado pela solução proposta sob a perspectiva de profissionais da área. Os elementos que caracterizam essa avaliação especializada foram definidos da seguinte forma:

- **Objeto de estudo:** O objeto de estudo é o roteiro de audiodescrição gerado pela solução proposta.
- **Propósito:** O propósito é avaliar o roteiro de audiodescrição gerado pela solução proposta com relação a capacidade de descrever os eventos do filmes. Para isso, audiodescritores profissionais analisaram um vídeo e o respectivo roteiro gerado pelo sistema, destacando os pontos positivos e negativos do roteiro.
- **Foco de qualidade:** O principal efeito estudado pelo experimento é avaliar os pontos positivos e negativos do roteiro de audiodescrição gerado pela solução.
- **Perspectiva:** A partir do ponto de vista de audiodescritores brasileiros.

- **Contexto:** Essa parte do experimento foi realizada através de um formulário *on line* que foi enviado aos audiodescritores. Cada profissional assistiu um filme e avaliou o roteiro de audiodescrição gerado pela solução proposta.

B.2.2 Planejamento do Experimento

Nessa seção será apresentado o contexto do experimento, os sujeitos e objetos utilizados no experimento, dentre outros aspectos do experimento.

Seleção do Contexto

Essa parte do experimento foi conduzida de forma *on line* com um conjunto de 5 audiodescritores voluntários. Estes profissionais foram convidados responder um formulário *on line* que continha 3 partes: apresentação da solução proposta (descrição do trabalho); apresentação do vídeo e roteiro de audiodescrição de teste; e questionário.

Esta parte da avaliação foi feita através de formulário *on line* devido a localização geográfica dos audiodescritores e à dificuldade de encontrar profissionais atuantes na área.

Seleção dos Sujeitos

Os sujeitos desse experimento são cinco audiodescritores voluntários e escolhidos por conveniência. Os usuários são profissionais que atuam na área e possuem variados tempos de experiência em audiodescrição. A seleção dos sujeitos foi um fator de dificuldade devido a escassez de profissionais da área.

Instrumentação

A avaliação foi feita através de um formulário *on line* que continha 3 partes: apresentação da solução proposta (descrição do trabalho), apresentação do vídeo e roteiro de audiodescrição de teste e questionário.

O vídeo selecionado como objeto deste experimento foi um filme que pertence ao gênero Ficção. A escolha de selecionar um vídeo foi feita a pedido dos audiodescritores para evitar que a duração dos testes se estendessem muito. A Tabela C.2 apresenta o vídeo que foi utilizado nos testes.

Tabela B.2: Vídeo utilizado nos testes

Vídeo	Duração	Descrição
Vídeo 1	5m29s	Esse vídeo é um curta-metragem do gênero comédia.

Os dados dessa avaliação foram coletados através de questionário composto de duas partes e é apresentado no Apêndice E. Na primeira parte do questionário, o audiodescritor fornece informações pessoais como sexo, idade, tempo de experiência, estado em que atua, entre outras. Na segunda parte, ele responde a 11 (onze) perguntas sobre o roteiro de audiodescrição gerado pelo sistema para destacar os pontos positivos e negativos da solução.

APENDICE C - QUESTIONÁRIO DE AVALIAÇÃO DOS CONTEÚDOS ACESSÍVEIS

Parte 1: Informações dos Usuários

Sexo: ☐ Masculino ☐ Feminino

Qual a sua faixa etária?

- ☐ Abaixo de 18 anos
- ☐ Entre 18 e 30 anos
- ☐ Entre 31 e 40 anos
- ☐ Acima de 40 anos

Qual o seu grau de deficiência visual?

- ☐ Total
- ☐ Parcial

Qual a natureza da deficiência visual?

- ☐ Congênita
- ☐ Adquirida

Qual é o seu grau de escolaridade?

- ☐ Nunca estudei
- ☐ Ensino Fundamental Incompleto
- ☐ Ensino Fundamental Completo
- ☐ Ensino Médio Incompleto
- ☐ Ensino Médio Completo
- ☐ Ensino Superior Incompleto
- ☐ Ensino Superior Completo
- ☐ Pós-graduação
- ☐ Outro. Qual? _____

Você costuma a assistir filmes?

- ☐ Sim
- ☐ Não

Você costuma ir ao cinema?

- ☐ Sim
- ☐ Não

Quanto tempo de vídeos com audiodescrição você consome por semana?

- ☐ Nenhum.
- ☐ Menos de 1 hora.
- ☐ Entre 1 e 5 horas.

- () Entre 5 e 10 horas.
- () Acima de 10 horas.

Parte 2: Compreensão dos Conteúdos

Vídeo 1:

1) O vídeo mostra um homem dirigindo um carro. Qual é o tipo de roupa que ele está vestindo?

- A) Roupas de praia.
- B) Camisa social e paletó.
- C) Camiseta e jeans.
- D) Não sei responder

2) Uma discussão acontece entre o personagem e os passageiros de um ônibus. Qual o motivo dessa discussão?

- A) Porque ele estacionou na faixa de pedestre.
- B) Porque ele tem uma atitude arrogante.
- C) Porque ele fez elogios à namorada de alguém.
- D) Não sei responder

3) Como o homem suja a sua roupa?

- A) Alguém cospe nele.
- B) Alguém joga comida nele.
- C) Alguém atira uma pedra nele.
- D) Não sei responder

4) O homem estacionou o carro no seu local de trabalho. Onde fica?

- A) Em um ministério.
- B) Na câmara legislativa.
- C) Em uma loja.
- D) Não sei responder

5) Ele deve trocar a roupa que está suja. Como ele volta para casa?

- A) Ele pega um ônibus.
- B) Ele chama um táxi.
- C) Ele dirige o carro.
- D) Não sei responder

Vídeo 2:

1) Qual a primeira coisa que a personagem faz após acordar?

- A) Toma café.
- B) Faz uma maquiagem.

- C) Vai ao banheiro.
- D) Não sei responder

2) Quantas vezes a personagem acorda?

- A) Uma vez.
- B) Duas vezes.
- C) Três vezes.
- D) Não sei responder

3) Qual o nome e idade da personagem?

- A) Ana; 18 anos.
- B) Rita; 25 anos.
- C) Maria; 14 anos.
- D) Não sei responder

4) Com a tesoura ela recorta um pedaço de:

- A) Revista
- B) Jornal
- C) Livro
- D) Não sei responder

5) Qual o signo do horóscopo da personagem?

- A) Leão
- B) Áries
- C) Peixes
- D) Não sei responder

6) A personagem deixa transparecer o seu estado de espírito. Como ela se sente?

- A) Esperançosa.
- B) Ansiosa.
- C) Tranquila.
- D) Não sei responder

Vídeo 3:

1) Onde a personagem mora?

- A) Trailer.
- B) Cabana.
- C) Circo.
- D) Não sei responder

2) Em cima do fogão ligado há uma panela. O que está ela está cozinhando?

- A) Frango.
- B) Sopa.

C) Ovo.

D) Não sei responder

3) Qual o programa que está passando na televisão?

A) Novela.

B) Corrida de atletismo.

C) Corrida de carros.

D) Não sei responder

4) De onde a personagem principal faz a ligação telefônica?

A) do orelhão público.

B) do telefone de casa.

C) do celular.

D) Não sei responder

5) O que acontece com a personagem no final do filme?

A) Vai embora para a casa da mãe.

B) Foge com o circo.

C) Volta para a casa.

D) Não sei responder

Parte 3: Avaliação dos Conteúdos

1) É fácil de entender? (6 - Claro, 1 - Confuso)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

2) A audiodescrição auxilia a compreensão?

(6 - Claro, 1 - Confuso)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

APENDICE D - QUESTIONÁRIO DE AVALIAÇÃO COM AUDIODESCRITORES

PARTE 1

1. Qual a sua faixa etária? *

- ☐)Abaixo de 20 anos
- ☐)Entre 20 e 35 anos
- ☐)Entre 35 e 45 anos
- ☐)Acima de 45 anos

2. Qual o Estado onde você atua?

3. Você trabalha na audiodescrição de qual conteúdo visual? *

Pode selecionar mais de um tipo de conteúdo visual.

- ☐)Filme
- ☐)Teatro
- ☐)Ópera
- ☐)Dança
- ☐)Televisão
- ☐)Eventos Esportivos
- ☐)Outros. Qual?

4. Há quanto tempo você trabalha/estuda audiodescrição? *

- ☐)Menos de 1 ano
- ☐)de 1 a 5 anos
- ☐)de 5 a 10 anos
- ☐)Acima de 10 anos

5. Você costuma utilizar o roteiro cinematográfico na criação da audiodescrição de filmes? Por quê? *

- ☐)Sim
- ☐)Não

PARTE 2

As perguntas a seguir estão relacionadas ao sistema proposto e ao vídeo de exemplo apresentado anteriormente. Após cada pergunta, há um espaço para observações complementares, sugestões ou críticas sobre o assunto tratado.

1. O roteiro gerado pelo sistema aumenta a eficiência do processo de descrição? *

(6 - Aumenta muito; 1- Não aumenta)

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

Observações:

2. Qual a classificação da adequação da velocidade da audiodescrição? *

(6 – Muito rápida; 1- Muito lenta)

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

Observações:

3. Quanto ao sincronismo, como você avalia a simultaneidade entre os eventos do filmes e as descrições?*

(6 – Muito bom; 1- Muito ruim)

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

Observações:

4. As lacunas entre os diálogos identificadas pelo sistema são satisfatórias? *

(6 – São satisfatórias; 1- Não são satisfatórias)

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

Observações:

5. Como você avalia a qualidade das descrições geradas? *

(6 – Muito bom; 1- Muito ruim)

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

Observações:

6. As descrições geradas servem como referência durante o trabalho de criação da audiodescrição? *

() Sim

() Não

Observações:

7. O roteiro de audiodescrição contém informações relevantes e precisas? *

(6 – Muito bom; 1- Muito ruim)

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

Observações:

8. Você usaria o sistema para auxiliar no processo de trabalho como audiodescritor? Por quê? *

() Sim

() Não

Por quê?

9. Você recomendaria o sistema para outros audiodescritores? *

() Sim

() Não

Observações:

10. Em que pontos você acha que o sistema poderia melhorar? *

11. Tem algum recurso/funcionalidade que você gostaria de ver nessa solução?