

CineAD: Um Sistema de Geração Automática de Roteiros de Audiodescrição

Virginia Pinto Campos
Trabalho de Mestrado
Programa de Pós Graduação em
Informática (PPGI)
Universidade Federal da Paraíba
João Pessoa - PB
virginia@lavid.ufpb.br

Tiago Maritan U. de Araújo
(Orientador)
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
tiagomaritan@lavid.ufpb.br

Guido L. de S. Filho
(Coorientador)
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
guido@lavid.ufpb.br

ABSTRACT

Audio description (AD) is a fundamental accessibility feature for people who are blind or have low vision. This paper presents a system for automatic generation of audio description, called CineAD. The system detects gaps in speech and generates these descriptions from the analysis of the original script and subtitles. To evaluate the solution, we automatically generate the audio description script of two movies. The preliminary results indicate that the solution is able to generate descriptions of the most important events of the film and the size of the descriptions is a major challenge related to the automatic creation of AD scripts.

RESUMO

A audiodescrição (AD) é um recurso de acessibilidade fundamental para que pessoas cegas ou com baixa visão possam ter acesso ao cinema. Este artigo apresenta um sistema de geração automática de roteiros de audiodescrição de filmes, denominado CineAD. O sistema detecta os intervalos entre falas candidatas a receber a AD, e gera essas descrições a partir da análise do roteiro original e da legenda. Para testar a solução, os roteiros de audiodescrição de dois filmes foram gerados automaticamente. Os resultados preliminares mostram que a solução tem potencial para gerar as descrições dos eventos mais importantes do filme e que o tamanho das frases das descrições é um dos principais desafios relacionados a criação automática dos roteiros de AD.

Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues—*Handicapped person/Special Needs*

Keywords

audio description, screenplay, accessibility

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).

A audiodescrição é um recurso de acessibilidade desenvolvido para melhorar o acesso de pessoas cegas à essas informações visuais, permitindo que elas participem de experiências intrinsecamente visuais, como o cinema, através de descrições de imagens ou narração de ações. Para descrever esses filmes, os roteiros de audiodescrição são geralmente produzidos de forma manual, o que acaba demandando muito tempo [9]. Segundo Lakritz e Salway [9] é necessário, em média, cerca de 60 horas para um profissional descrever um filme com 2 horas de duração.

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.

Neste contexto, o objetivo 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.

O restante deste artigo está organizado da seguinte forma: na seção 2, é apresentada a base conceitual para a au-

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

Figura 1: Trecho do roteiro AD do curta Águas de Romanza (Araújo, 2010) [1].

diodescrição no cinema e os trabalhos relacionados ao tema. Na seção 3, o sistema proposto e seus componentes são descritos detalhadamente. A seção 4 contém os resultados preliminares obtidos. Por fim, a seção 5 apresenta as conclusões e propostas de trabalhos futuros.

2. AUDIODESCRIÇÃO NO CINEMA

O recurso de acessibilidade voltado para as pessoas cegas ou com baixa visão que cobre o meio visual do cinema é a audiodescrição. De acordo com Benecke [2], a AD 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.

Essa tradução transforma toda a informação visual em palavras, ou seja, audiodescreve tudo o que é visto. Diferentemente de outras tecnologias que promovem a acessibilidade, a AD não é um recurso a ser utilizado isoladamente, mas sim, vinculado ao produto visual [11]. Ademais, a aplicação da audiodescrição não se restringe ao cinema, sendo possível utilizá-la também em outros ambientes, como, por exemplo, no teatro, em óperas, museus, etc.

Segundo Nunes *et al.* [11], para se fazer a audiodescrição é preciso elaborar um roteiro adaptado ao tipo de conteúdo visual que se deseja audiodescrever. O audiodescritor roteirista é o responsável por criar um roteiro preciso com as inserções de tempo e orientações para leitura, sem conter descrições desnecessárias, mas apresentando todas as informações relevantes para a compreensão do filme. Ressalte-se também que é importante que poucas palavras sejam utilizadas nas descrições para evitar um excesso de informações no áudio.

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

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 [9]. O uso de tecnologia pode auxiliar significativamente na inserção e elaboração da audiodescrição em filmes. Em razão disso, alguns estudos vem sendo feitos para incorporar técnicas computacionais nos processos envolvidos.

Gagnon *et al.* [5] apresentam um projeto de desenvolvimento de ferramentas de software para auxiliar a produção de audiodescrição. O objetivo é a detecção automática do conteúdo visual através de uma coleção de filtros especializados que extraem informações de alto nível, afim de descrevê-los. O software processa automaticamente um filme, extrai informações (tiro de transições, textos, lugares, rostos e a ação de atores) e atribui um texto para cada elemento visual adicionado que, opcionalmente, pode ser lido por um sintetizador de voz. 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.

Quanto à locução, Kobayashi *et al.* [8] apresentam um estudo sobre o uso de sintetizadores de voz para gerar a narração das descrições. Os estudos mostraram que as descrições sintetizadas foram aceitas, independentemente da língua utilizada, mas se adequam melhor aos vídeos informativos, onde o entendimento é o fator crítico. Para vídeos de entretenimento, as narrações humanas são mais recomendadas, pois permitem que as pessoas cegas tenham uma experiência mais natural e agradável.

Kobayashi *et al.* [7] apresentam uma plataforma que usa síntese de voz para adicionar audiodescrição à vídeos on-line. Essa plataforma possui uma ferramenta de apoio à criação de roteiros AD, que fornece uma interface visual para editar as frases e especificar na linha do tempo quando cada descrição deve ser lida. Neste editor de roteiros é possível modificar os parâmetros da voz, tais como o sexo e a velocidade. Os testes com usuários cegos 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.

Uma geração semiautomática de audiodescrição é abordada em Lakritz e Salway [9]. O sistema realiza a extração das informações pertinentes do roteiro do filme e as convertem para uma linguagem mais apropriada para audiodescrição. A extração de informações é feita a partir das palavras mais frequentes no roteiro e a conversão da linguagem é obtida através da aplicação de heurísticas. Os resultados mostraram que esta abordagem é capaz de extrair 80% das informações importantes e de converter 66% das frases inadequadas à audiodescrição. O sistema foi avaliado por profissionais descritores, que indicaram que 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

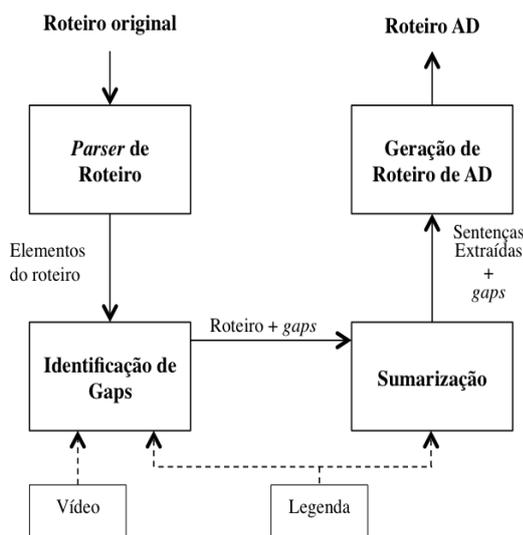


Figura 2: Visão Esquemática do Sistema CineAD

esforço. Um dos objetivos do CineAD é preencher esta lacuna. O sistema será descrito a seguir.

3. CINEAD

CineAD tem como objetivo reduzir os problemas de acessibilidade à filmes cinematográficos pelas pessoas cegas ou com baixa visão. Para isso, a solução propõe a geração automática de roteiros de audiodescrição. Na nossa abordagem, a audiodescrição é criada a partir da análise de dois elementos principais: o roteiro do filme e a legenda. A visão esquemática do sistema é ilustrada na Figura 2. Todos os componentes foram implementados na linguagem Java.

O funcionamento geral do sistema ocorre 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, a sumarização faz a extração das sentenças mais importantes do roteiro e, desta forma, resume o roteiro original e descarta as informações secundárias que não são necessárias à 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, nos gaps detectados na etapa de Identificação de Gaps. Os componentes da solução serão detalhados nas subseções 3.1, 3.2, 3.3 e 3.4.

3.1 Parser de Roteiro

A leitura automática de roteiros exige que o roteiro seja fornecido em um formato digital específico que permita a identificação e rotulagem de seus elementos. Neste trabalho, o formato *.celtx*, definido pela ferramenta CELTX, é utilizado, mas é possível estender esse módulo para suportar outros tipos de formato.

O CELTX é um processador e editor de textos multi-plataforma, livre, de pré-produção, para escrever roteiros audiovisuais, como filmes, peças teatrais, etc. [3]. 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. Os rótulos mais utilizados são: título de cena, ação, personagem, diálogo, planos e transição entre cenas. A Figura 3 apresenta um exemplo da estrutura do roteiro e a identificação dos rótulos de cada elemento.

O componente Parser de Roteiro faz a leitura dos arquivos contidos no formato *.celtx* e extrai os elementos do roteiro. Para realizar a leitura, foram utilizados os recursos de manipulação de *rdf* do Apache Jena [6], que é um framework em Java que dá suporte a utilização da Web Semântica.

A saída deste componente é o conjunto de elementos ordenados, cronologicamente, de acordo com a sequência do roteiro original. A audiodescrição realizada pelo sistema descreve, neste primeiro momento, apenas as ações do filme, ou seja, os elementos do roteiro rotulados como “ação”.

3.2 Identificação 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 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, ou seja, funcionando como ponto de sincronização para inserção dos trechos de AD. Nesta implementação, foram utilizados arquivos de legendas no formato *.srt* e vídeos em formato *.mp4*.

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 obter o último intervalo sem falas (entre o último diálogo e o fim do filme), a duração do vídeo é consultada e serve como o tempo final do último gap.

3.3 Sumarização

Diferentemente do roteiro original 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 espaços de tempo 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.

Segundo Edmundson [4], as informações importantes em um texto podem ser identificadas com base na frequência das palavras-chave contidas nele. Portanto, encontrando as

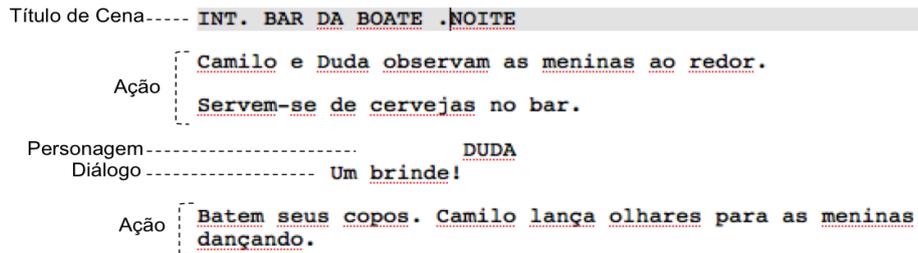


Figura 3: Estrutura do roteiro e identificação dos elementos (Trecho do roteiro do filme A Cartomante, 2004)

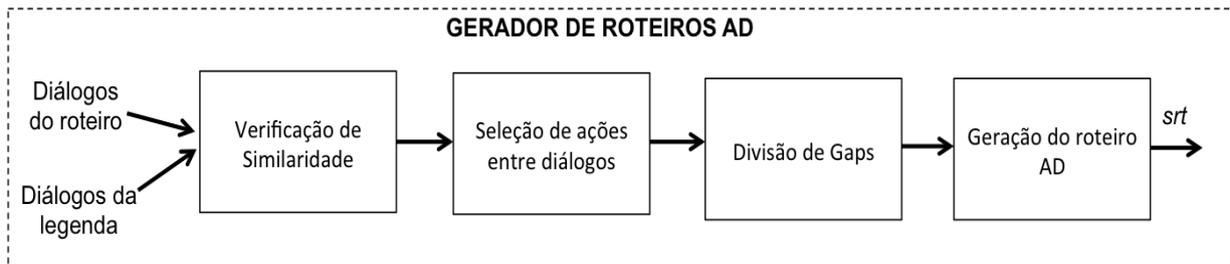


Figura 4: Fluxo do componente gerador de roteiro AD

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 [9]. Nesse trabalho, a 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 [9] 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: as palavras mais frequentes em roteiros de audiodescrição, as palavras mais frequentes no roteiro que será audiodescrito e uma lista com o nome de personagens. Salway, Vassiliou e Ahmad [12] 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 formam a primeira lista do componente de sumarização. A segunda lista contém as 30 palavras mais frequentes do roteiro que se deseja gerar a audiodescrição. Por fim, a terceira lista é formada pelos nomes de todos os personagens do roteiro-alvo. Na etapa de criação dessas listas, uma biblioteca de mecanismos de procura de texto escrita em Java, a *Apache Lucene* [10], foi utilizada para retornar os termos que mais ocorrem no documento.

Após a criação das listas, todas as ações do roteiros são fragmentadas em frases, de modo que cada ação seja subdividida em suas frases constituintes. A seleção das sentenças para o roteiro de audiodescrição dá-se 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 do roteiro de audiodescrição. A etapa seguinte é a geração do arquivo com o roteiro AD.

3.4 Geração de roteiro AD

O componente Gerador de Roteiro cria o roteiro com a audiodescrição e os seus tempos específicos. O objetivo do componente é a alocação das sentenças, extraídas na etapa de sumarização, nos intervalos sem falas identificados pelo componente de Identificação de Gaps. O componente é executado em quatro etapas: verificação de similaridade entre as falas da legenda e os diálogos do roteiro; extração de ações entre os diálogos; divisão dos gaps para inserção das ações; e geração do arquivo com o roteiro AD. O fluxo da geração é apresentado na Figura 4.

A primeira etapa verifica a similaridade da legenda com os diálogos do roteiro, fazendo com que cada texto da legenda seja associado a um diálogo de personagem contido no roteiro. Essa verificação é necessária porque várias falas da legenda correspondem ao mesmo diálogo presente no roteiro. A Figura 5 mostra um exemplo deste caso. Para saber qual legenda está relacionada com cada diálogo é feito o cálculo da similaridade utilizando a biblioteca *Apache Lucene* [10]. 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.

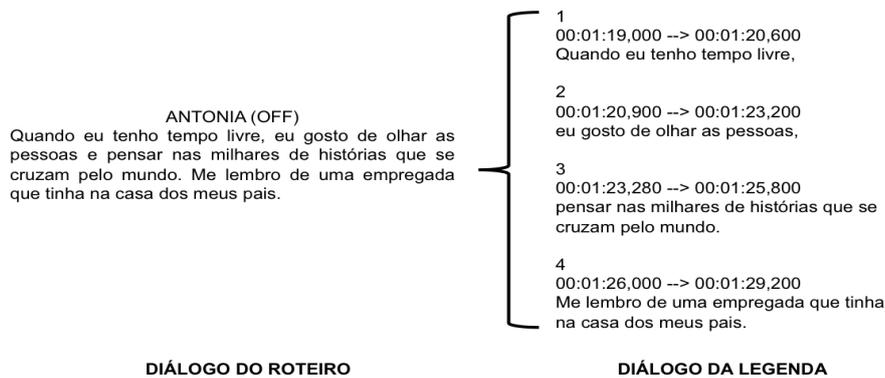


Figura 5: Exemplo da similaridade entre diálogos do roteiro e legenda (Trecho do roteiro do filme A Cartomante, 2004)

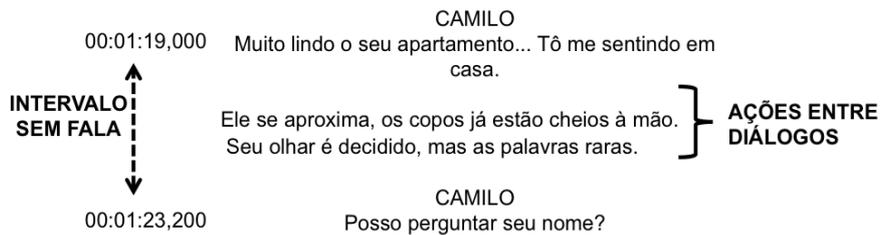


Figura 6: Exemplo de ações entre diálogos do roteiro (Trecho do roteiro do filme A Cartomante, 2004)

Após o mapeamento dos diálogos, a etapa seguinte utiliza todas 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 6. Dessa forma, nesta etapa, as frases mais importantes das ações que estão entre dois diálogos são selecionadas. Essa etapa utiliza como base as informações produzidas pelo componente de sumarização e os gaps sem falas identificados pelo componente Identificador de Gaps.

Em seguida, na etapa de divisão do gaps, 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 igualmente entre o número de sentenças de ações, isto é, caso se tenha um gap de 40 segundos e 4 frases descrevendo as ações daquele intervalo, cada frase terá 10 segundos para ser exibida, seguindo a cronologia do roteiro. 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.

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 na criação de um arquivo no formato *srt*, comumente utilizado para 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 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 ou para sistemas de voz sintetizada.

4. RESULTADOS PRELIMINARES

A fim de realizar uma avaliação preliminar do CineAD, a audiodescrição de dois vídeos foi gerada usando o sistema. O primeiro filme foi um curta-metragem (filme A), e o segundo, um trecho de um longa-metragem (filme B). A Tabela 1 apresenta as características dos vídeos utilizados.

O filme A contém uma única cena. Em seu roteiro, estão presentes 83 frases de ações e 11 diálogos. O tempo do vídeo ocupado por diálogos corresponde a 32,67%, o que implica que 67,33% não apresenta falas. O filme B, por outro lado, contém 3 cenas, 41 frases de ações e 20 diálogos no roteiro. A duração dos diálogos corresponde a 23,13% do tempo do vídeo, enquanto que 76,87% aos intervalos sem falas. Destaque-se que ambos os vídeos apresentam mais informações visuais do que diálogos.

Os roteiros de audiodescrição gerados nos testes descrevem os principais eventos dos filmes. Na Figura 7, é apresentado como exemplo, um trecho de um dos filmes com o relacionamento entre os elementos da legenda, do roteiro original e do roteiro AD. A coluna da esquerda mostra os intervalos de tempo e as falas da legenda. A coluna central apresenta o roteiro original com ações e diálogos, enquanto

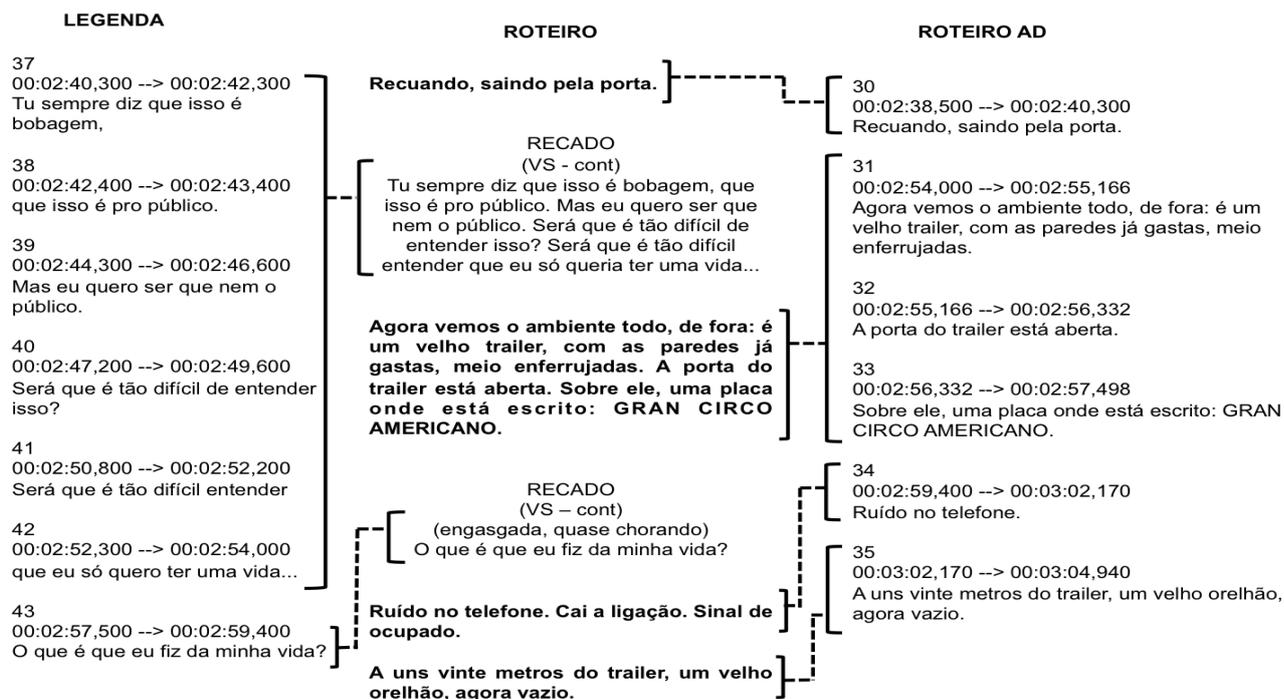


Figura 7: Exemplo de geração de roteiro AD (Trecho do roteiro do filme 3 Minutos, 1999)

Tabela 1: Características dos vídeos de teste

Itens	Filme A	Filme B
Tamanho (mm:ss)	05:42	04:31
nº de cenas	1	3
nº de frases de ações	83	41
nº de diálogos	11	20
% de dialogos no vídeo	32.67	23.13
% sem dialogos no vídeo	67.33	76.87

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, as descrições do roteiro AD estão associadas às ações do roteiro original.

De acordo com a Figura 7, 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, cada gap do roteiro AD é dividido entre as frases das ações do roteiro original. O funcionamento do componente de sumarização pode ser percebido na descrição de número 34 do roteiro AD, onde são excluídas as frases consideradas menos importantes (“Cai a ligação.” e “Sinal de ocupado”). Como proposta de trabalho futuro, pretende-se avaliar a qualidade das descrições geradas com usuários cegos ou com baixa visão.

A geração automática do roteiro de audiodescrição dos dois filmes resultaram nos dados mostrados na Tabela 2.

Tabela 2: Dados da geração do roteiro AD

Itens	Filme A	Filme B
nº de gaps sem falas detectados	50	39
nº de frases de ações selecionadas	71	33
nº de gaps descartados	39	25
nº de gaps preenchidos	11	14
% de audiodescrição no vídeo	47.26	37.48
Tempo médio para cada sentença	2.27	3.07
nº médio de palavras por sentença	10.23	8.57

No filme A, foram detectados 50 gaps, ou seja, 50 intervalos entre as falas da legenda. Na geração do roteiro de audiodescrição, 39 gaps foram descartados e apenas 11 gaps sofreram inserções de descrições em seus intervalos. Esse descarte acontece quando não existem ações entre os diálogos descritas no roteiro. Um exemplo comum deste 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.

Outro dado importante é que o componente de sumarização extraiu do roteiro 71 sentenças de ações que foram inseridas nos 11 gaps. Essa distribuição do tempo para apresentação das sentenças é feito na etapa de divisão de gaps do componente Gerador de Roteiro AD. O tempo total das descrições geradas consome cerca de 47,26% da duração do vídeo. As sentenças tem o tamanho médio de 10,23 palavras e o tempo médio alocado para cada uma é de 2,27 segundos.

No filme B, 39 gaps foram identificados e 33 sentenças de

ações foram extraídas. Na geração do roteiro de AD, foram descartados 25 gaps. As sentenças extraídas foram alocadas nos 14 gaps restantes. O tempo total da audiodescrição ocupa 37,48% do filme. As sentenças possuem o tamanho médio de 8,57 palavras, com o tempo médio de 3,07 segundos para cada.

Em ambos os roteiros testados, também é possível observar que o tamanho médio das sentenças é alto em relação ao tempo médio alocado para a sua apresentação. Isso implica que as sentenças da audiodescrição estão longas para o tempo disponível para sua narração. Para respeitar esses limites de tempo, uma alternativa seria a aplicação de técnicas de compressão de sentenças, como um complemento à sumarização. O módulo de compressão poderia ser adicionado após a extração das frases, reduzindo o seu comprimento, mas respeitando a gramática e sem perder informações.

Nestes testes iniciais foram observados outros problemas. Uma dificuldade encontrada na abordagem proposta é a detecção do início da primeira cena do filme a ser descrita. Os aspectos visuais e estéticos dos trechos iniciais dos filmes variam e, por essa razão, o início do vídeo pode não ter correspondência com o início da primeira cena. Um exemplo para esta situação é quando um filme apresenta, no início do vídeo, créditos iniciais sobre a tela preta ou vinhetas introdutórias antes da primeira cena. Além disso, esta informação também não pode ser obtida a partir da legenda, pois a primeira legenda indica apenas qual é o primeiro diálogo do filme. Em consequência disso, nesta primeira versão, o tempo inicial da primeira cena do roteiro no filme é informado manualmente. Esse valor de tempo é considerado o início do primeiro gap.

Outra questão observada durante os testes refere-se ao fato de que, em alguns momentos, o roteiro pode não ter correspondência com o audiovisual do filme, ou seja, podem existir diferenças entre o enredo do roteiro e o do filme. Isso acontece porque o roteiro de pré-produção é desenvolvido para servir de guia nas etapas de filmagem, e durante as gravações, ele pode sofrer alterações. Consequentemente, se o roteiro estiver diferente, a audiodescrição gerada automaticamente estará incorreta. Diante disso, é necessário que sejam utilizados roteiro de pós-produção na geração da audiodescrição.

Um aspecto a ser melhorado, que foi identificado durante os testes, diz respeito à utilização de regras para remover ações do roteiro que estão relacionadas ao áudio do filme, citando como exemplo a frase “Ruído no telefone”, apresentada no exemplo da Figura 7. A maior parte dos sons presentes no áudio do filme não necessitam de descrição, pois pessoas cegas e com baixa visão podem percebê-los. Desta forma, a exclusão dessas descrições desnecessárias, reduziria o número de sentenças e facilitaria a inserção e a divisão dos gaps.

5. CONCLUSÕES E TRABALHOS FUTUROS

Neste artigo, foi apresentado um sistema de geração automática de audiodescrição para filmes, o CineAD. Esta solução contempla recursos de acessibilidade, fornecendo a descrição de conteúdos visuais para as pessoas cegas ou com

baixa visão.

O sistema realiza a análise do roteiro original e da legenda do filme, retornando um roteiro com marcações de tempo contendo as descrições das ações. Os resultados preliminares demonstram que o sistema 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 narrador humano para a criação do áudio com as descrições.

Como proposta de trabalho futuro, planejamos incorporar um componente de compressão de sentenças na etapa de sumarização para reduzir o tamanho das frases. Além disso, regras ou heurísticas para remover conteúdos do roteiro que estão relacionados a sons 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. Planeja-se também a integração com uma estratégia de síntese de voz para geração de áudio e a distribuição desse fluxo de áudio para os usuários. Por fim, a realização de testes com usuários cegos ou com baixa visão e com audiodescritores para avaliar a qualidade das descrições geradas automaticamente e o nível de compreensão das mesmas.

6. REFERÊNCIAS

- [1] V. L. S. Araújo. A formação de audiodescritores no ceará e em minas gerias: Uma proposta baseada em pesquisa acadêmica. In L. M. V. de Mello Motta and P. R. Filho, editors, *Audiodescrição: Transformando Imagens em Palavras*. Secretaria dos Direitos da Pessoa com Deficiência do Estado de São Paulo, São Paulo, 2010.
- [2] B. Benecke. *Audio-description*, volume 49. Meta, Reading, Massachusetts, 2004.
- [3] Celtx. Free scriptwriting and all-in-one production studios. <https://www.celtx.com/index.html>, Jan. 2014.
- [4] H. P. Edmundson. New methods in automatic extracting. *Journal of the Association for Computing Machinery*, 16:264–285, 1969.
- [5] L. Gagnon, S. Foucher, M. Heritier, M. Lalonde, D. Byrns, C. Chapdelaine, J. Turner, S. Mathieu, D. Laurendeau, N. T. Nguyen, and D. Ouellet. Towards computer-vision software tools to increase production and accessibility of video description for people with vision loss. *Univers. Access Inf. Soc.*, 8(3):199–218, July 2009.
- [6] Jena. Apache jena. <http://jena.apache.org/>, Jan. 2014.
- [7] M. Kobayashi, T. Nagano, K. Fukuda, and H. Takagi. Describing online videos with text-to-speech narration. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, W4A '10, pages 29:1–29:2, New York, NY, USA, 2010. ACM.
- [8] M. Kobayashi, T. O'Connell, B. Gould, H. Takagi, and C. Asakawa. Are synthesized video descriptions acceptable? In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '10, pages 163–170, New York, NY, USA, 2010. ACM.
- [9] J. Lakritz and A. Salway. The semi-automatic

generation of audio description. Technical report, University of Surrey, Department of Computing, 2002.

- [10] Lucene. Apache lucene 2.0. <http://lucene.apache.org/>, Jan. 2014.
- [11] Nunes, Machado, and Vanzin. Audiodescriçao como tecnologia assistiva para o acesso ao conhecimento por pessoas cegas. In T. V. Vania Ribas Ulbricht and V. Villarouco, editors, *Ambiente Virtual de Aprendizagem inclusivo*. Pandion, Florianópolis, 2011.
- [12] A. Salway, A. Vassiliou, and K. Ahmad. What happens in films? In *ICME*, pages 49–52. IEEE, 2005.