

# Avaliação da Audiodescrição Gerada por Sistema

Uma Análise sob a Perspectiva do Usuário

Rafael F. Cordeiro<sup>†</sup>

Centro de Informática – Campus I  
Universidade Federal da Paraíba  
João Pessoa – PB, Brasil  
[cordeirof.rafael@gmail.com](mailto:cordeirof.rafael@gmail.com)

Tiago M. U. de Araújo

Centro de Informática – Campus I  
Universidade Federal da Paraíba  
João Pessoa – PB, Brasil  
[tiagomaritan@lavid.ufpb.br](mailto:tiagomaritan@lavid.ufpb.br)

Yuska P. C. Aguiar

Departamento de Ciências Exatas  
– Campus IV  
Universidade Federal da Paraíba  
Rio Tinto – PB, Brasil  
[yuska@dcx.ufpb.br](mailto:yuska@dcx.ufpb.br)

Virginia Pinto Campos

Universidade Federal do Rio Grande  
do Norte  
Natal – RN, Brasil  
[vcampos@dca.ufrn.br](mailto:vcampos@dca.ufrn.br)

## RESUMO

A audiodescrição (conhecida como AD) é uma faixa de áudio narrativa que auxilia as pessoas com deficiência visual na compreensão de conteúdos audiovisuais. O profissional responsável por gerar a AD é o audiodescritor. Este tem como tarefa transformar as informações visuais de vídeos ou apresentações em áudio para os espectadores. Este processo de transformação exige tempo e possui um alto custo. Como suporte para estes profissionais, tendo em vista a redução de tempo e custo para realização das audiodescrições, sistemas automáticos geradores de audiodescrição têm sido desenvolvidos. O objetivo deste trabalho de pesquisa consiste em avaliar um sistema gerador automático de audiodescrição de filmes. A metodologia proposta avalia a qualidade da AD gerada sob a perspectiva do usuário deficiente visual. O nível de acessibilidade da AD gerada automaticamente, a partir de dois modelos distintos, e de uma AD gerada por um audiodescritor foi avaliado a partir da aplicação de questionário sobre a compreensão dos vídeos exibidos. Os resultados mostraram que o modelo de AD automática baseado em roteiro cinematográfico apresentou resultados mais próximos à AD humana, enquanto o modelo baseado em reconhecimento de objetos contribuiu de forma mínima à compreensão, com resultados inferiores ao cenário sem audiodescrição.

## PALAVRAS-CHAVE

audiodescrição; acessibilidade; experiência do usuário.

Permission to reproduce or distribute, in whole or in part, material extracted from this work, verbatim, adapted or remixed, as well as the creation or production from the content of such work, is granted without fee for non-commercial use, provided that the original work is properly credited. IHC 2019 - WORKSHOP DE TESES E DISSERTAÇÕES, Outubro 2125, 2019, Vitória, Brasil. In Anais Estendidos do XVIII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais. Porto Alegre: SBC. © 2019 by the author(s), in accordance with the terms of the Creative Commons Attribution-NonCommercial 4.0 International Public License (CC BY-NC 4.0).

## INTRODUÇÃO

Em 2000 o governo brasileiro decretou a lei 10.098/2000 para promoção da acessibilidade a partir de normas e critérios de acessibilidade para as Pessoas com Deficiência (PcD) [7]. No Brasil, segundo o último Censo Demográfico [4], são 6,5 milhões de pessoas com deficiência visual, sendo 582 mil cegas e 6 milhões com baixa visão. Esta parte expressiva da população enfrenta diversas dificuldades em seu cotidiano. O acesso à informação é uma barreira real - quando o conteúdo não é apresentado de forma acessível (Braille e/ou ledor) - que atinge a vida das pessoas com deficiência visual em diferentes segmentos da sociedade (comunicação, educação, lazer, cultura, etc.).

Para suprir esta necessidade, o capítulo VII da lei 10.098/2000 compromete a eliminação de barreiras na comunicação e estabelece mecanismos para tornar acessíveis os sistemas de comunicação, garantindo às pessoas com deficiência o direito de acesso à informação, à comunicação, à educação, à cultura, ao esporte e ao lazer [7].

Um dos meios de possibilitar o acesso às informações audiovisuais é através da audiodescrição (AD), que consiste no detalhamento e apresentação de informações visuais relevantes na forma de áudios descritivos [2]. Ao decorrer do vídeo (ou de uma apresentação) a AD auxilia o espectador a compreender o ambiente, as ações e a história em que a cena está inserida. Assim, pessoas cegas ou com baixa visão podem ir ao cinema e teatros com a oportunidade de ampliar o conhecimento advindo do conteúdo das imagens ou cenas ali apresentadas [3].

O processo de criação da AD é tradicionalmente realizado por um profissional, o audiodescritor. Segundo [5], a produção manual da AD de um longa-metragem de 2 horas de duração pode demandar 60 horas de trabalho. Já em relação ao custo, um roteiro de AD fica em torno de 40 reais por minuto de vídeo no Brasil [9]. Levando em consideração uma emissora de televisão com uma grade de conteúdos atualizada diariamente ou um cinema com uma variedade rotativa de filme, o custo e tempo necessários

para desenvolvimento e disponibilização de acessibilidade para os conteúdos são altos.

Uma das alternativas para diminuir o tempo e custo de produção de AD é sua geração automática por sistemas computacionais. Neste caso, o sistema gera automaticamente a AD e a introduz no contexto do vídeo. Esta audiodescrição deve possibilitar a compreensão dos conteúdos em vídeos de forma semelhante (próxima) à descrição realizada por um audiodescritor humano. Vale salientar que este tipo de sistema não exclui o papel do audiodescritor humano do processo de criação de AD. Este profissional, essencial no processo de descrição, utiliza a solução computacional para otimizar seu trabalho. Ao ser liberado da carga total de trabalho, o audiodescritor assume a função de ajustar a versão gerada pelo sistema e realizar melhorias na mesma.

No entanto, para que este cenário seja efetivo, é necessário que os sistemas gerem AD de qualidade. Portanto, a definição de diretrizes para criação de audiodescrição, assim como a avaliação da AD gerada automaticamente, pode potencializar o uso das soluções automáticas (ou semiautomáticas). Dessa forma, o escopo desta pesquisa envolve a identificação e a definição de um conjunto de diretrizes para apoiar o desenvolvimento e a avaliação de audiodescrição gerada por sistema, a partir da perspectiva do usuário com deficiência visual.

## TRABALHOS RELACIONADOS

Determinadas pesquisas tem como objetivo desenvolver sistemas geradores de audiodescrição e avaliar a qualidade da AD gerada. O desenvolvimento de um sistema automático gerador de audiodescrição cujo objetivo é auxiliar na descrição das imagens em vídeos voltados à vigilância é descrito em [9]. Tal sistema recebe o vídeo a ser descrito, realiza a detecção de objetos e pessoas nas cenas, e por fim disponibiliza a AD gerada em formato de áudio. Para verificar a qualidade da AD gerada, os autores adotaram um método de avaliação automático através de *script*. Os elementos verificados foram: quantidade de detecções de objetos e pessoas que o sistema realizou; a taxa de acerto; número de objetos relatados nas descrições; tamanho da frase e tempo de narração. A partir dessas métricas, foi avaliado se a AD gerada corresponde ao esperado ou não.

Uma revisão sistemática de literatura sobre as técnicas de geração automática e semiautomática de audiodescrição foi realizada por [6]. Investigou-se quais técnicas de desenvolvimento estão sendo relatadas na literatura, quais estratégias de avaliação estão sendo adotadas para analisar a qualidade de AD geradas automaticamente. O autor selecionou 13 artigos e identificou que 8/13 artigos possuem audiodescrição automatizadas, apenas um artigo apresenta semiautomatizada e os outros 4/13 são manuais ou não especificado. No quesito relacionado à avaliação destes sistemas, foi detectado que 10/13 artigos selecionados coletaram *feedback* sobre as ADs através de testes com usuários. Apenas um artigo realizou testes automáticos e 2/13 trabalhos não especificaram como foram feitas as avaliações. No fim, o autor

ressalta a importância de realizar os testes com usuários para coletar informações relevantes sobre como melhorar as AD geradas por sistemas.

## SOLUÇÃO PRETENDIDA

A diversidade de vertentes encontradas na literatura para avaliar a qualidade da audiodescrição gerada por sistema, impossibilita realização de cruzamento de testes distintos para averiguar qual sistema proporciona a qualidade esperada pelo público alvo. Diante disto, esta pesquisa tem como objetivo principal definir um conjunto de diretrizes para guiar a avaliação da qualidade da audiodescrição gerada (semi) automaticamente a partir da perspectiva de usuários reais (pessoas com deficiência visual) no contexto de Testes de Usabilidade.

## METODOLOGIA

A metodologia adotada compreende quatro etapas: (i) definição do sistema gerador e AD e seus modelos de geração; (ii) design do experimento para realização da avaliação da qualidade de AD a partir de testes de usabilidade com usuários com deficiência visual para comparar o nível de compreensão dos participantes quando vídeos são exibidos em cenários distintos de acessibilidade; (iii) execução do experimento para avaliar se as ADs geradas auxiliam na compreensão audiovisual; e, (iv) análise dos resultados para definição do conjunto de diretrizes.

Atualmente a pesquisa se encontra na primeira versão dos testes experimentais, onde são feitas as análises dos dados coletados para realizar o levantamento das melhorias no design do experimento e na AD gerada pelo sistema selecionado. A pesquisa foi aprovada pelo Comitê de Ética e Pesquisa da Universidade Federal da Paraíba, através do Processo N°11355218.4.0000.5188

## Definição do sistema e dos modelos de AD

A solução selecionada para esta pesquisa foi o CineAD, que é um sistema desenvolvido pelo Núcleo de Pesquisa e Extensão LAViD (Laboratório de Aplicações de Vídeo Digital) da Universidade Federal da Paraíba. O CineAD tem como objetivo, através da geração automática da AD, reduzir as barreiras enfrentadas pelas pessoas com deficiência visual em relação à acessibilidade de filmes cinematográficos [8].

O CineAD possui dois formatos de captação para geração de audiodescrição automática. O modelo I utiliza o roteiro cinematográfico como base para geração da AD, onde descreve personagens, ações, ambientes e objetos. O modelo II, utiliza como base a identificação de objetos nas cenas apresentadas no vídeo, cuja AD descreve apenas objetos em cena. Neste projeto foram analisados os dois modelos disponibilizados pelo CineAD.

## Design do experimento

Os protocolos de usabilidade determinam as etapas, processos e atividades que devem ser planejadas e conduzidas durante a

avaliação baseada na observação da interação do usuário [10]. Considerando o tipo de produto sob avaliação (solução de audiodescrição) e o público de usuário com deficiência visual, se faz necessário adaptar protocolo normalmente utilizado em testes de usabilidade de produtos de software genéricos.

Como também todo o material utilizado no experimento seja acessível (fichas, questionário, roteiros de teste, termo de consentimento, etc.) para os usuários com deficiência visual. Além disso, o ambiente onde será realizado o experimento deve ter uma acústica que minimize o barulho externo, evitando interferência por ruído durante a exibição do vídeo e respectiva AD.

O Protocolo Experimental de Observação da Interação (PEOI) [11], adaptado para esta pesquisa, tem como propósito auxiliar na investigação da relação entre as características de um indivíduo, sua percepção da situação e o seu desempenho na realização de tarefas durante a realização de testes de usabilidade. A de PEOI se justifica pela possibilidade de adaptação do mesmo para contextos de acessibilidade, como anteriormente realizado para deficientes visuais [1] e para pessoas com Síndrome de Down [10].

Para tornar possível a análise do impacto da audiodescrição gerada pelo CineAD, para ambos os modelos, na compreensão de conteúdos audiovisuais, mais especificamente filmes, os participantes do teste assistiram a um conjunto de curta-metragem. Os participantes foram agrupados em 4 grupos (A, B, C e D) e os vídeos foram disponibilizados considerando 4 cenários:

- AD-H: Vídeo com audiodescrição humana;
- AD-R: Vídeo com audiodescrição automática baseada em roteiro cinematográfico via CineAD (Modelo I);
- AD-O: Vídeo com audiodescrição automática baseada em reconhecimento de objetos via CineAD (Modelo II);
- S/AD: Vídeo sem audiodescrição- S/AD.

A escolha destes cenários foi idealizada para analisar se as ADs automáticas do CineAD (baseada em roteiros e a baseada em identificação de objetos) favorecem a compreensão dos conteúdos das mídias de vídeos quando comparadas aos cenários extremos: ausência de solução de audiodescrição (pior cenário) e solução de audiodescrição humana (cenário ideal). Para tornar isto possível, os voluntários selecionados foram separados em quatro grupos com quatro participantes cada. Cada grupo de voluntário possui acesso a cada um dos cenários apenas uma vez. O objetivo desta regra é evitar que apenas um grupo avalie um determinado cenário de acessibilidade. Portanto, definiu-se uma rotação para a apresentação dos vídeos nos respectivos cenários definidos e para os diferentes grupos (Tabela 1).

Vídeos	Grupos Participantes			
	A	B	C	D
Ré mebol	S/ AD	AD - H	AD - R	AD - O
O pedido	AD - H	AD - R	AD - O	S/ AD
Morango	AD - R	AD - O	S/AD	AD - H
Mercúrio	AD - O	S/ AD	AD - H	AD - R

Tabela 1 Cenário de rotação de acessibilidade nos vídeos

Os grupos de voluntários formados são constituídos de maneira homogênea (em relação à idade, nível de cegueira, escolaridade e experiência prévia com AD). A formação dos grupos foi sistematizada a partir de um questionário pré-teste aplicado com os voluntários, a partir do qual foi possível definir a alocação destes nos grupos correspondentes. Os vídeos selecionados são curta-metragens, com duração máxima de 10 minutos. Esta configuração se justifica pelo fato do conteúdo abordar uma situação autocontida, com início, meio e fim. A duração máxima de 10 minutos é importante para manter a concentração e o foco dos participantes durante as sessões de teste. Detalhes sobre o perfil dos participantes podem ser acessadas no link <http://bit.ly/2ZtcNOL>

## Execução do experimento

Participaram do experimento 16 voluntários com deficiência visual ou baixa visão, recrutados no Instituto dos Cegos da Paraíba Adalgisa Cunha e na Fundação Centro Integrado de Apoio à Pessoa com Deficiência da Paraíba (FUNAD-PB), local de realização dos testes. Os participantes assinaram o Termo de Consentimento Livre e Esclarecido. As sessões de teste foram individuais e não paralelas, conduzidas por dois avaliadores e com duração média de 30 minutos.

A coleta de dados foi realizada a partir da aplicação de uma entrevista semiestruturada com os participantes. Esta ocorreu após a exibição de cada vídeo, de forma individual e não paralela. Foram coletadas informações sobre a compreensão dos vídeos (história, cenário e objetos dos vídeos) para os diferentes cenários considerados. A entrevista foi organizada de forma a contemplar dois objetivos macro. O primeiro relativo ao nível de compreensão dos conteúdos exibidos (e respectiva AD), incluindo aspectos de compreensão geral do vídeo, objetos do cenário presentes, características e/ou comportamentos dos personagens. As quatro questões apresentadas foram objetivas, a exemplo de: “Onde o casal está conversando?” e as suas alternativas “a) em casa b) no clube c) no parque d) não sei informar”. A compreensão do vídeo é avaliado de acordo com a taxa média de acerto das questões objetivas. Em seguida, questões subjetivas foram apresentadas para os participantes exporem suas considerações e sugestões de melhoria para as ADs exibidas. O acesso ao roteiro da entrevista semiestruturada e as respostas podem ser acessados na íntegra a partir do link: <http://bit.ly/2UdeMp1>

Estas perguntas foram realizadas até mesmo para os filmes que não possuem AD, a fim de não induzir o usuário a identificar qual vídeo tem ou não o recurso de acessibilidade de AD.

## RESULTADOS PARCIAIS

A partir dos dados coletados durante o experimento foi possível realizar a análise da taxa média de acertos das questões objetivas de cada cenário proposto. Com o Gráfico 1, é possível identificar que o cenário com a AD baseada em Roteiros possui a taxa de acertos (45,83%) mais próxima da AD humana (54,17%), que é a

considerada como cenário ideal. Já a AD baseada na identificação de objetos ficou próxima da taxa de acertos (25,00%) do cenário sem AD, sendo assim uma contribuição mínima na compreensão dos vídeos.

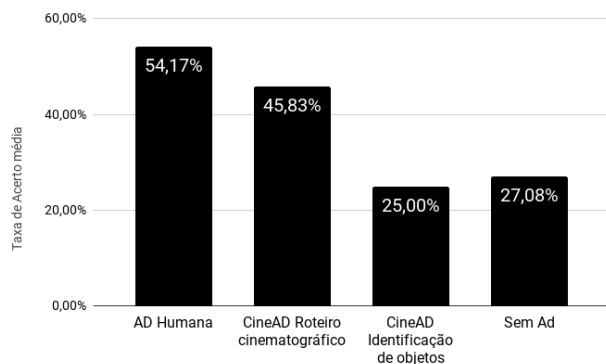


Gráfico 1 Taxa média de acerto x Tipo de cenário

Durante a entrevista semiestruturada alguns voluntários sugeriram melhorias para as ADs geradas pelo CineAD. Os participantes evidenciaram a importância de descrever uma quantidade maior de cenas e detalhar aspectos que auxiliam o ouvinte na compreensão do contexto da cena. Em específico, para o CineAD Modelo II (baseado em identificação de objetos), deve descrever ambientes, personagens e ações, não apenas objetos.

Os voluntários também identificaram que, em certos momentos do vídeo, a voz da audiodescrição sobrepõe às dos personagens, sendo difícil de compreender o que era fala e o que era descrição. Como também apontam a semelhança da narração ou dos personagens com a voz sintetizada. As soluções propostas para resolver essas questões são, configurar a sincronização da AD e selecionar um tom de voz distinto do áudio original.

Durante a aplicação do questionário pré-teste realizado para descrever o perfil dos voluntários e separá-los em grupos, quando perguntado se “Você tem acesso a audiodescrição?”, a resposta se tornou vaga porque os usuários consideravam uma única experiência com AD, como sendo acesso habitual à audiodescrição. Sendo assim, não foi possível avaliar se a familiaridade com o recurso auxilia na compreensão dos vídeos com audiodescrição. A solução proposta seria reformular a pergunta para “Com que frequência você tem acesso a AD?” e as alternativas da resposta para “Nunca tive acesso, Raramente, Regularmente, Com frequência e No dia a dia”.

Após as implantações das melhorias recomendadas será realizado uma nova rodada de experimento com novos voluntários, com o objetivo de analisar se os pontos identificados neste experimento foram solucionados e se a AD gerada pelo CineAD aumentou o nível de compreensão dos usuários.

## CONTRIBUIÇÕES ESPERADAS

Este trabalho apresentou uma metodologia de avaliação de qualidade de audiodescrição gerada automaticamente por

sistemas, especificamente CineAD, sob o ponto de vista de pessoas com deficiência visual. Quando concluída a fase de análise da perspectiva do usuário, será desenvolvida uma metodologia de testes computacionais para avaliação da AD automática sem a presença de usuários. Para isso, práticas e métricas serão identificadas, além de métodos automáticos de avaliação de audiodescrição. A partir desses dois parâmetros (com ou sem usuário) espera-se ser capaz de criar um conjunto de diretrizes que permita avaliar a audiodescrição gerada por um sistema.

## AGRADECIMENTOS

Agradecemos à CAPES pela bolsa concedida para realização da pesquisa, aos voluntários participantes do experimento, ao Instituto dos Cegos da Paraíba Adalgisa Cunha e à Fundação Centro Integrado de Apoio à Pessoa com Deficiência (FUNAD) pela colaboração e disponibilização de infraestrutura.

## REFERÊNCIAS

- [1] ACO Lima, MFQ Vieira, R.S Ferreira, YPC Aguiar, MP Bastos and SLM Lopes Júnior. Evaluating system accessibility using an experimental protocol based on usability. *International Conferences on Interfaces and Human Computer Interaction* 2018, 2018. p. 85-92.
- [2] E. P. C Franco and M. C. C. C SILVA. Audiodescrição: breve passeio histórico. *Audiodescrição: transformando imagens em palavras*. São Paulo: Secretaria dos Direitos da Pessoa com Deficiência do Estado de São Paulo, p. 23-42, 2010.
- [3] F. J. de Lima. Áudio-descrição: orientações para uma prática sem barreiras atitudinais. *Revista Brasileira de Tradução Visual*, v. 2, n. 2, 2010.
- [4] IBGE. Censo demográfico 2010. IBGE: Instituto Brasileiro de Geografia e, 2010.
- [5] J. Lakritz and A. Salway, The Semi-Automatic Generation of Audio Description from Screenplays, Dept. Of Computing Technical Report CS-06-05, 2002, University of Surrey. <http://www.bbrel.co.uk/pdfs/CS-06-05.pdf>
- [6] R. M. Correia. Uma Revisão Sistemática da Literatura sobre Técnicas de Geração Automática e Semi-Automática. João Pessoa: UFPB, 2014.
- [7] V. C. D. S. P. Brasil, (2000). Lei nº 10.098, de 19 de dezembro de 2000. Diário Oficial da União, <https://www2.camara.leg.br/legin/fed/lei/2000/lei-10098-19-dezembro-2000-377651-publicacaooriginal-1-pl.html>
- [8] VP Campos. Um sistema de geração automática de roteiros de audiodescrição. 2015. 89 f. Dissertação (Mestrado em Informática) - Universidade Federal da Paraíba, João Pessoa, 2015.
- [9] VP Campos, LM Gonçalves and TMU De Araújo. Applying audio description for context understanding of surveillance videos by people with visual impairments. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2017. p. 1-5.
- [10] W.S. Pereira et al. Usuários com Síndrome de Down: Como realizar Testes de Usabilidade?. *RENOTE*, v. 15, n. 2. <https://doi.org/10.22456/1679-1916.79274>
- [11] Y. P. C Aguiar. (2012). Protocolo Experimental para Observação da Interação um Arcabouço para o Estudo do Comportamento Humano. Campina Grande: UFCG.