Universidade de Brasília (UnB)

Instituto de Letras (IL)

Departamento de Línguas Estrangeiras e Tradução (LET)

Bacharelado em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação (LEA-MSI)

Julia Vilela Salviato

Geração semi-automática de audiodescrição: utilização de Inteligência Artificial na narração

Brasília, Distrito Federal, Brasil

dezembro/2023

Julia Vilela Salviato

Geração semi-automática de audiodescrição: utilização de Inteligência Artificial na narração

Trabalho de Conclusão de Curso apresentado ao Departamento de Línguas Estrangeiras e Tradução (LET) da Universidade de Brasília (UnB) como requisito parcial para obtenção de grau de Bacharel em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação.

Orientadora : Profª. Drª. Helena Santiago Vigata

Brasília
2023

RESUMO

Tendo em vista o alto custo de inclusão de audiodescrição (AD) nas produções, a automação tem sido apresentada como a solução mais viável para o problema de custo-benefício, pois é um processo mais rápido e mais econômico na realização da narração e na pós-edição das produções. Este trabalho de conclusão de curso se propõe a utilizar a Inteligência Artificial (IA) como sintetizador de voz na narração de AD do episódio "A Revolta dos Carecas", da Turma da Mônica, com o objetivo de analisar o comportamento e eficiência da narração feita pela IA, na tentativa de averiguar se ela se aproxima da narração original feita por um ser humano. Para sua realização, o roteiro de AD do episódio foi inserido no programa Genny para ser lido pela IA, que atendeu satisfatoriamente a maioria dos requisitos analisados, a saber: entonação/emoção, ritmo/fluidez, velocidade e voz natural. Dito isto, o único requisito que não atendeu às expectativas foi a voz natural, uma vez que a IA utilizada se mostrou mais eficiente na leitura com vozes em inglês do que em português.

Palavras-chave: Narração de audiodescrição; Automação; Inteligência Artificial; *Genny*

ABSTRACT

Given the high cost of including audio description in productions, automation has been proposed as the most viable solution to the cost-benefit problem, as it is a faster and more economical process for carrying out narration and post-editing productions. This undergraduate thesis proposes to use Artificial Intelligence (AI) as a voice synthesizer in the audio description (AD) narration of the episode "A Revolta dos Carecas", by Turma da Mônica, with the aim of analyzing the behavior and efficiency of the narration made by the AI, in an attempt to find out whether it comes close to the original narration made by a human being. To carry it out, the episode's AD script was inserted into the Genny program to be read by the AI, which satisfactorily met most of the requirements analyzed, namely: intonation/emotion, rhythm/fluidity, speed and natural voice. That said, the only requirement that did not meet expectations was the natural voice, since the AI used proved to be more efficient in reading with voices in English than in Portuguese.

Keywords: Audio description narration; Automation; Artificial Intelligence; Genny

RESUMEN

Considerando el alto costo de incluir audiodescripción (AD) en las producciones, la automatización se ha propuesto como la solución más viable para el problema costo-beneficio, ya que es un proceso más rápido y económico para realizar la narración y posedición de las producciones. Por lo tanto, este trabajo de fin de grado propone utilizar la Inteligencia Artificial (IA) como sintetizador de voz en la narración AD del episodio "A Revolta dos Carecas", de Turma da Mônica, con el objetivo de analizar el comportamiento y la eficiencia de la narración realizada por la IA, en un intento de averiguar si se acerca a la narración original realizada por un ser humano. Para llevarlo a cabo, se introdujo el guion de AD del episodio en el programa Genny para ser leído por la IA, el cual cumplió

satisfactoriamente con la mayoría de los requisitos analizados, a saber: entonación/emoción, ritmo/fluidez, velocidad y voz natural. Dicho esto, el único requisito que no cumplió con las expectativas fue la voz natural, ya que la IA utilizada demostró ser más eficiente en la lectura con voces en inglés que en portugués.

Palabras clave: Narración de audiodescripción; Automatización; Inteligencia artificial; *Genny*

RÉSUMÉ

Compte tenu du coût élevé de l'inclusion de l'audiodescription (AD) dans les productions, l'automatisation a été présentée comme la solution la plus viable au problème coût-bénéfice, puisqu'il s'agit d'un processus plus rapide et plus économique pour réaliser des productions de narration et de post-édition. Ce projet de fin d'études propose d'utiliser l'Intelligence Artificielle (IA) comme synthétiseur vocal dans la narration AD de l'épisode « A Revolta dos Carecas », de Turma da Mônica, dans le but d'analyser le comportement et l'efficacité de la narration réalisée par l'IA, pour tenter de savoir si elle se rapproche de la narration originale faite par un être humain. Pour le réaliser, le script AD de l'épisode a été inséré dans le programme Genny pour être lu par l'IA, ce qui répondait de manière satisfaisante à la plupart des exigences analysées, à savoir : intonation/émotion, rythme/fluidité, rapidité et voix naturelle. Cela dit, la seule exigence qui n'a pas répondu aux attentes était la voix naturelle, puisque l'IA utilisée s'est avérée plus efficace dans la lecture avec des voix en anglais qu'en portugais.

Mots-clés : Narration en audiodescription ; Automatisation ; Intelligence artificielle ; *Genny*

LISTA DE ABREVIATURAS E SIGLAS

AD	Audiodescrição
AENOR	*Asociación Española de Normalización y Certificación*
Audetel	*Audio Description of the Television*
CAT	Comitê de Ajudas Técnicas
DARPA	*Defense Advanced Research Projects Agency*
DSP	*Digital Signal Processor*
DVD	Disco Digital Versátil
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
MEC	Ministério da Educação
NLP	*Natural Language Processing*
NTV	*Nippon TV*
PDS	Processamento Digital de Sinais
PLN	Processamento de Linguagem Natural
TTS	*Text-to-speech*

LISTA DE FIGURAS

LISTA DE TABELAS

Sumário

1 Introdução

A audiodescrição (AD) tem um papel importante no campo da acessibilidade audiovisual, pois ela permite que pessoas cegas ou com baixa visão sejam incluídas, não só no mundo cinematográfico, como também nas demais esferas da sociedade, tendo acesso a informações visuais em geral graças à tradução de imagens em palavras. Tendo em vista o alto custo da inclusão de audiodescrição nas produções, a automação tem sido apresentada como uma solução viável para o problema de custo-benefício, além de ser um processo mais rápido na realização de narração e pós-edição das produções. Por falta de um programa que produza um **roteiro** de audiodescrição automático disponível no mercado no momento da elaboração deste projeto, o presente trabalho se concentra apenas na **leitura** (sintetização de voz) de um roteiro de audiodescrição utilizando Inteligência Artificial, por isso é uma geração audiodescritiva semi-automática. Além disso, o trabalho também tratará de assuntos como sistema *text-to-speech* (TTS) e técnicas *deep learning* aplicadas na Inteligência Artificial (IA) para o avanço da narração automática.

A escolha do tema surgiu após eu ter cursado a matéria Modalidades da Tradução Audiovisual (TAV) com o professor Charles Rocha Teixeira, do Bacharelado em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação da Universidade de Brasília. Nessa matéria aprendi os parâmetros da legendagem, legendagem acessível para surdos e ensurdecidos e também os da audiodescrição. Assim, tive a oportunidade não só de aprender a teoria como também a prática dessas modalidades. Inicialmente, sabia que gostaria de desenvolver meu TCC nesse campo de estudo, mas não tinha uma ideia concreta com a qual trabalhar. Entretanto, depois do *boom* da Inteligência Artificial nas redes sociais devido à sua utilização na elaboração de obras de arte e imagens muito realistas apenas com comandos textuais, como é o caso do modelo de IA chamado Dall-E, que utiliza uma combinação de técnicas de aprendizado de máquina e redes neurais para transformar os comandos de texto em representações visuais, eu tive a inspiração de juntar a audiodescrição com a Inteligência Artificial e criar minha proposta.

Num primeiro momento, cogitei a possibilidade de utilizar a IA para criar de forma automática um roteiro de audiodescrição, apenas com as imagens e falas do filme, contudo, quando pesquisei por programas de IA que possibilitassem isso, constatei que, apesar de existirem alguns protótipos desenvolvidos com esse intuito, eles ainda não estão disponíveis no mercado no momento da elaboração deste projeto. Além disso, pude observar a falta de

pesquisas a respeito da automação da narração de AD. Sendo assim, mudei a etapa em que usaria a IA; ao invés de usá-la na roteirização, utilizei-a na narração da audiodescrição como sintetizador de voz, já que existem alguns programas de IA que possibilitam a leitura de textos.

Como pontuado anteriormente, o alto custo e o longo tempo necessários para elaborar e inserir a audiodescrição em obras cinematográficas e vídeos em geral são apresentados como motivos para a falta de acessibilidade para cegos e pessoas com baixa visão em grande parte dos filmes, séries, novelas, vídeos no YouTube e em outras plataformas. Sendo assim, esta pesquisa pretende contribuir para a área de estudos relacionada à tradução audiovisual, pois tenta propor, ao utilizar a Inteligência Artificial como sintetizador de voz, uma alternativa para uma maior inserção de audiodescrição em vídeos, tendo em vista um custo menor e tempo de preparo mais rápido. E assim, mostra a sua relevância não só ao contribuir com o campo teórico em questão, mas também possibilita que pessoas cegas e com baixa visão tenham cada vez mais acesso aos vários vídeos disponíveis em diversas plataformas.

O objetivo geral do trabalho é analisar o desempenho da narração de AD feita por uma IA, levando em consideração os parâmetros para a narração citados neste trabalho, na tentativa de constatar o comportamento e eficiência da IA e averiguar se a narração com voz sintética se aproxima da narração original feita por um ser humano.

Com intuito de atingir o objetivo geral deste trabalho, os seguintes objetivos específicos foram definidos:

- Definir o que é a audiodescrição e os parâmetros na narração;
- Realizar um levantamento bibliográfico a respeito da geração automática/semi-automática de audiodescrição;
- Realizar um levantamento bibliográfico a respeito de sistema *text-to-speech* e *deep learning* no avanço da IA como software sintetizador de voz;
- Averiguar o comportamento e eficiência da IA escolhida, após sua aplicação como narrador de audiodescrição, com o intuito de responder se as novas tecnologias já são suficientes para automatizar a narração de audiodescrição.

É importante ressaltar que esta proposta não tem por objetivo descredibilizar ou substituir o trabalho de audiodescritores. Apesar de a automação ainda ser um assunto delicado, pois envolve questões polêmicas, como a substituição de profissionais da área pela tecnologia e a aprovação dos usuários de AD, este trabalho se concentra apenas na averiguação do desempenho das vozes sintéticas face ao avanço da IA. Além disso, a

utilização da IA é uma discussão inevitável e necessária, uma vez que, por algum tempo, não se tinha o cuidado com a narração em si, havia apenas a preocupação com o roteiro de AD. Ou seja, a narração humana também precisou ser discutida e passar por melhorias, assim como o uso da IA como sintetizador de voz também precisa ser discutido e estudado.

2 Revolução da Tecnologia da Informação e as ferramentas assistivas

De acordo com Manuel Castells (1999), sociólogo e professor universitário espanhol, foi na década de 1970 que a Revolução da Tecnologia da Informação surgiu nos Estados Unidos. Estas intensas e numerosas mudanças tecnológicas nas áreas da microeletrônica, computação, telecomunicações e radiodifusões e optoeletrônica difundiram transformações de ordem cultural, social e econômica no cenário mundial. Ainda segundo Castells (1999, p. 69), o que caracteriza essa revolução tecnológica "não é a centralidade de conhecimentos e informação, mas a aplicação desses conhecimentos e dessa informação para a geração de conhecimentos e de dispositivos de processamento/comunicação da informação". As invenções que ganharam destaque neste período de segundo pós-guerra foram o computador programável e o transistor. Contudo, a criação da Internet pela Agência de Projetos de Pesquisa Avançada do Departamento de Defesa dos Estados Unidos (DARPA) foi, sem dúvida, a coroa da revolução, "o mais revolucionário meio tecnológico da Era da Informação" (Castells, 1999, p. 82).

Essa tal revolução proporciona, hoje, o desenvolvimento de tecnologias que promovem a inclusão digital a partir de ferramentas assistivas: audiodescrição, legenda acessível, leitores de tela, mudança no esquema de cores de sites etc. O Comitê de Ajudas Técnicas (CAT), explica o que são estas ferramentas assistivas da seguinte forma:

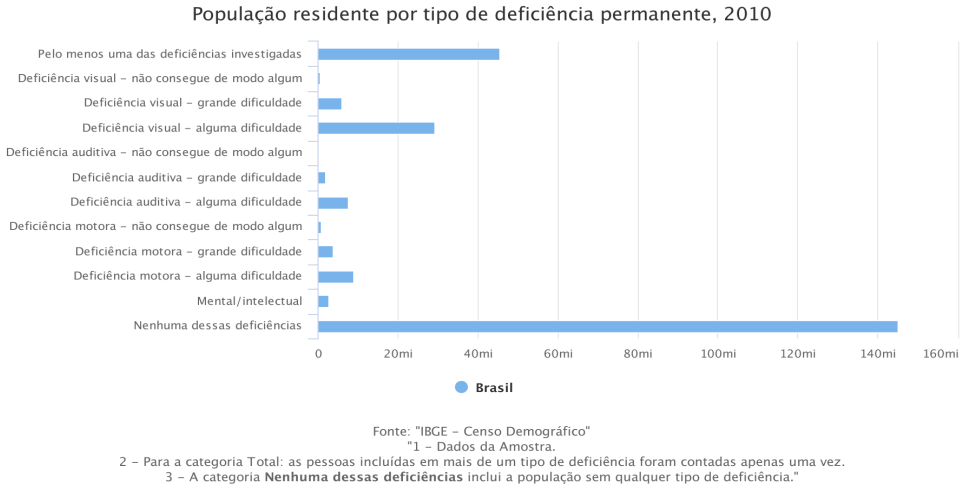
> Tecnologia Assistiva é uma área do conhecimento, de característica interdisciplinar, que engloba produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social (Brasil, 2007).

A audiodescrição, ferramenta foco desta pesquisa, é o processo em que uma informação visual é descrita oralmente; ela promove a comunicação de uma informação e ainda possibilita que esses dispositivos de processamento/comunicação da informação sejam utilizados no seu processo de criação da narração.

3 Contextualização da deficiência visual no Brasil

De acordo com o censo levantado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2010[1], conforme se pode constatar na Figura 1, pelo menos 506.377 pessoas estavam na categoria de pessoa com deficiência visual, isto é, com perda total na visão, 6.056.533 responderam ter grande dificuldade para enxergar e 29.211.482 disseram ter alguma dificuldade.

FIGURA 1- POPULAÇÃO RESIDENTE POR TIPO DE DEFICIÊNCIA PERMANENTE, 2010



Fonte: IBGE, censo de 2010

A deficiência visual compreende o espectro que vai desde a cegueira até a visão subnormal (Brasil, 2000), a cegueira, ou perda de visão total, pode ser adquirida ou congênita, enquanto a visão subnormal, ou baixa visão, engloba patologias como miopia, estrabismo, astigmatismo, ambliopia e hipermetropia.

O Dia Nacional do Cego foi criado pelo então presidente da República Jânio Quadros, e é comemorado no Brasil desde o dia 13 de dezembro de 1961. A data, segundo o Ministério da Educação (MEC), tinha por objetivo colocar pessoas com deficiência visual em destaque e conscientizar a sociedade a respeito do preconceito e discriminação para com eles. Contudo, só isso não é o suficiente para promover a inclusão dessas pessoas, por isso, em julho de 2015

[1] Até o momento, os resultados do último censo, realizado em 2022, não foram publicados na íntegra. Apenas estão disponíveis os resultados das categorias de Indígenas, Quilombolas, População e domicílios e Prévia da População dos Municípios. Sendo assim, até o momento da realização do presente trabalho, não foi possível o acesso às informações a respeito de pessoas com deficiência visual no Brasil em 2022.

entrou em vigência a Lei Brasileira de Inclusão da Pessoa com Deficiência que, além de muitas outras disposições, apresenta alguns artigos válidos de serem retratados aqui.

O art. 42 estabelece que:

> "[...] a pessoa com deficiência tem direito à cultura, ao esporte, ao turismo e ao lazer em igualdade de oportunidades com as demais pessoas, e que lhe é garantido o acesso a bens culturais em formato acessível, a programas de televisão, cinema, teatro e outras atividades culturais e desportivas em formato acessível; e a monumentos e locais de importância cultural e a espaços que ofereçam serviços ou eventos culturais e esportivos [...]" (Brasil, 2015).

O parágrafo 6º do art. 44 determina que as salas de cinema devem oferecer, em todas as sessões, recursos de acessibilidade para a pessoa com deficiência.

4 Breve histórico da audiodescrição

A audiodescrição já existia informalmente, contudo, conforme Philip Piety (2004), surgiu formalmente em 1975 na Universidade de São Francisco, Estados Unidos, em uma dissertação de mestrado de Gregory Frazier, intitulada *The Autobiography of Miss Jane Pitman: An All-audio Adaptation of the Teleplay for the Blind and Visually Handicapped*. Entretanto, foi somente na década de 1980 que a audiodescrição realmente marcou presença por meio da ativista que ficou cega aos 30 anos, Margaret Rockwell, considerada como a mãe da audiodescrição, e seu marido, Cody Pfanstiehl. Rockwell fundou, em 1974, o Metropolitan Washington Ear, que promovia o serviço de leitura para cegos, e em 1980 ela e seu marido implementaram o programa de audiodescrição em peças teatrais no teatro Arena Stage Theater, Washington DC. De acordo com Nunes *et al*. (2010), a audiodescrição chega ao Japão por meio da rede de televisão NTV, que incluiu o serviço em sua programação em 1983. Na Europa, a audiodescrição não tardou; em 1989, ela chegou na França por meio da associação Valentin Haüy; já no Reino Unido, em 1992, com o projeto *Audetel* que reunia emissoras, fabricantes de tecnologia e organizações que representam os interesses dos cegos e pessoas com baixa visão (Salway). Na Espanha, em 1994, o projeto espanhol denominado *Sonocine* se transformou no sistema *Autodesk* e passou a ser usado em algumas áreas audiovisuais diversas, tais como teatro, vídeos, televisão, museus e exposições (Nunes *et al*., 2010).

No Brasil, a audiodescrição estreou no ano de 2003 com o filme *Assim Vivemos,* exibido no Festival Internacional de Cinema (Nunes *et al*., 2010). Em 2005, o primeiro filme audiodescrito, *Irmãos de Fé*, foi lançado em formato DVD. Já no ano de 2008, a marca

pioneira em propaganda com AD foi a Natura (Franco; Silva, 2010). Em 2007, foi exibida a peça teatral *Andaime*, em São Paulo, a primeira que contou com o recurso de acessibilidade. Uma das figuras muito importantes da AD brasileira é Lívia Motta, que foi a responsável pela exibição da primeira peça e ópera audiodescrita no Brasil, além de publicar, juntamente com Paulo Romeu Filho, o primeiro livro brasileiro sobre audiodescrição, *Audiodescrição: Transformando Imagens em Palavras* (2010).

Quanto ao público infantil, apesar de não haver milhares de obras audiovisuais com AD no Brasil, os episódios do desenho Turma da Mônica no seu canal oficial do YouTube já possuem AD disponível, um dos motivos para a escolha deste desenho para o trabalho. Além disso, atualmente existem algumas pesquisas brasileiras a respeito de como elaborar o roteiro de AD e a narração própria para crianças. O artigo "Filmes infantis audiodescritos no Brasil: Uma Análise dos Filmes A Turma da Mônica 2 e Hotel Transilvânia", escrito por Charles Rocha Teixeira, Sofia Ferreira Alves Fiore e Bárbara Carvalho (2013), apresenta parâmetros para a AD baseado no RNIB Sunshine House School, um guia elaborado exclusivamente para crianças. Da mesma forma, a estudante graduada pela UnB em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação, Bianca Nathália da Silva Pereira (2020), realizou um compilado de informações a respeito da audiodescrição infantil e de como fazê-la. É válido apontar que a preocupação com o público infantil não se resume a apenas estes trabalhos citados, outros estudiosos e profissionais da área possuem pesquisas similares.

Esse panorama geral e muito breve não corresponde a todos os eventos da AD no Brasil e ao redor do mundo. Segundo a Fundação Dorina Nowill Para Cegos (2020/2021), Estados Unidos, Reino Unido, França, Espanha, Alemanha e Uruguai são os países que mais investem em audiodescrição na televisão, no cinema e no teatro.

5 Fundamentação teórica

Nesta seção serão desenvolvidos alguns conceitos importantes para o entendimento completo do problema de pesquisa, como os de audiodescrição, acessibilidade, tradução intersemiótica, *text-to-speech* e *deep learning*, presentes nos subitens: Conceituando a audiodescrição, Parâmetros da narração de AD, Geração semi-automática de AD, Utilização de Inteligência Artificial como sintetizador de voz na AD: *text-to-speech e deep learning* e Narração humana vs. narração sintética.

5.1 Conceituando a audiodescrição

Primeiramente, vale salientar que a conceituação da audiodescrição neste artigo se dará dentro do contexto de pessoas apenas com deficiência visual e com relação à produção eletrônica de imagens em movimento, como novelas, filmes, séries e vídeos em geral.

A audiodescrição, *grosso modo*, se trata de uma ferramenta assistiva para pessoas com deficiência visual[2], entretanto, para que se compreenda sua definição e importância, é preciso analisar o termo *acessibilidade*, cuja aparição não se mostra muito antiga no mundo. De acordo com Sassaki (2004), esse termo apareceu nos Estados Unidos, em 1940, para "designar a condição de acesso das pessoas com deficiência", tendo sua origem com o "surgimento dos serviços de reabilitação física e profissional". Conforto e Santarosa (2002) assinalam que: "acessibilidade passa a ser entendida como sinônimo da aproximação, um meio de disponibilizar a cada usuário interfaces que respeitem suas necessidades e preferências". No Brasil, a Lei Brasileira de Inclusão da Pessoa com Deficiência considera o termo acessibilidade como:

> [...] possibilidade e condição de alcance para utilização, com segurança e autonomia, de espaços, mobiliários, equipamentos urbanos, edificações, transportes, informação e comunicação, inclusive seus sistemas e tecnologias, bem como de outros serviços e instalações abertos ao público, de uso público ou privados de uso coletivo, tanto na zona urbana como na rural, por pessoa com deficiência ou com mobilidade reduzida [...] (Brasil, 2015).

Pincelado o conceito de acessibilidade, entendemos por que a audiodescrição é considerada como uma ferramenta assistiva. Franco e Silva (2010) consideram que a AD "consiste na transformação de imagens em palavras para que informações-chave transmitidas visualmente não passem despercebidas e possam também ser acessadas por pessoas cegas ou com baixa visão".

Conforme o *Guia para produções audiovisuais acessíveis*, desenvolvido pelo Ministério da Cultura/Secretaria do Audiovisual, a AD pode ser definida como:

> "[...] uma modalidade de tradução audiovisual, de natureza intersemiótica, que visa tornar uma produção audiovisual acessível às pessoas com deficiência visual. Trata-se de uma locução adicional roteirizada que descreve as ações, a linguagem corporal, os estados emocionais, a ambientação, os figurinos e a caracterização dos personagens [...]" (Naves et al., 2016, p. 9).

[2] De acordo com a ENAP (2020, p.10), a audiodescrição também beneficia pessoas com outras deficiências, como Síndrome de Down e dislexia, e idosos.

A audiodescrição passou a integrar uma categoria da tradução, sendo caracterizada com natureza intersemiótica. Entende-se por tradução intersemiótica, segundo Jakobson (1959), a "interpretação dos signos verbais por meio de sistemas de signos não-verbais", ou seja, é quando se traduz uma obra para um outro tipo de textualidade ou plataforma midiática (Silva, 2018). A título de exemplo, pode-se citar a tradução intersemiótica como uma poesia traduzida em uma pintura ou, como no caso da audiodescrição, a tradução de imagens em palavras. Cabe ressaltar que a AD não estava incluída na definição de Jakobson quando esta foi concebida; passou a integrar esta categoria com a posterior ampliação dela por outros autores.

5.2 Parâmetros da narração de AD

Bittner (2012) argumenta que orientações também devem ser direcionadas aos narradores de AD e aos designers de som, não apenas aos roteiristas da audiodescrição. Sendo assim, esta pesquisa, ao propor analisar e entender parâmetros da narração de AD, se apoia em sua totalidade no âmbito da narração, visto que há muito material voltado apenas para os roteiristas, deixando os narradores de lado. Dessa forma, recomendações ou regras sobre como elaborar o roteiro da audiodescrição (o que, como e quando descrever) não serão encontradas na revisão literária a seguir.

Conforme o *Guia para produções audiovisuais acessíveis* (Naves et al., 2016, p.10), a narração deve ser "fluida e não monótona, sem vida". A obra ainda defende que uma narração neutra pode comprometer o fluxo do filme, uma vez que a AD precisa compor o significado já existente da obra audiovisual em questão e não destoar com sua carga emocional. Por exemplo, uma "narração mais pausada, com entonação melancólica, de uma cena dramática, pode contribuir para a dramaticidade".

Outro ponto abordado é que não é recomendável sobrepor a AD aos diálogos, sons importantes e a trilha sonora, salvo casos em que uma ação relevante acontece ao mesmo tempo que diálogos/sons/trilha sonora. Nesta situação, a AD se torna essencial para o entendimento da cena, mas precisa ser feita de forma concisa para interferir o mínimo possível no resto da informação sonora. Toda vez que a informação visual for mais relevante que a informação sonora, a AD deve sobrepor-se à informação sonora, seja ela em forma de diálogos, efeitos sonoros ou música.

Com relação ao público infantil, o guia estabelece que a narração deve ser mais lúdica, a fim de não cansar a criança usuária de AD.

Da mesma forma, Cristóbal Cabeza-Cáceres (2013) entende que a velocidade da narração afeta tanto a compreensão como a fruição fílmica e conclui que uma narração lenta demais ou rápida demais pode levar à rejeição por parte do público. Com relação à entonação, ele acredita que esse fator não influencia na compreensão do filme, mas sim na fruição; dessa forma, deve compor a obra. Além disso, Cabeza-Cáceres constata que estudos com abordagem narratológica discursiva enfatizam a necessidade de considerar o AD como parte integrante do produto e como um elemento que deve se encaixar e funcionar dentro da sua própria narrativa, tratando assim de conceitos como coerência e coesão discursiva dentro do produto final, além da necessidade de fazer uma análise narratológica prévia que permita abordar a criação do roteiro audiodescritivo.

Com relação à voz usada, o American Council of the Blind (2009) orienta que a voz do narrador de AD deve ser diferente das vozes da produção e do narrador da obra, caso haja, mas não pode ser uma voz que tire a atenção do telespectador, como a voz de uma pessoa famosa facilmente reconhecível. Além disso, define a velocidade da narração em 160 palavras por minuto. Ainda com relação à velocidade, a Independent Television Commission (2000) recomenda não fazer uma descrição apressada e que cada palavra seja clara, audível e cronometrada para não ficar perto demais do diálogo que está por vir.

Por sua vez, o guia francês *Charte de l'audiodescription* (Morisset; Gonant, 2008) também defende que o narrador deve adaptar sua voz à emoção da cena e ao ritmo da ação, contudo, deixa claro que deve haver certa neutralidade, caso contrário, um narrador extremamente presente competiria com os personagens do filme. Entretanto, a questão da neutralidade levantada pelos autores diz respeito à etapa de elaboração do roteiro propriamente dito e não da narração em si.

Conforme a Asociación Española de Normalización y Certificación (AENOR, 2005), deve-se selecionar o narrador de acordo com o tipo de voz (masculina, feminina, adulto ou jovem) e o tom adequado para cada obra. Além disso, a norma espanhola também recomenda que, para obras infantis, o locutor (a) utilize entonação mais expressiva, adequada para crianças. Para as produções direcionadas a um público adulto, da mesma forma que o guia francês, esta norma defende que a narração deve ser neutra, mas com a entonação, o ritmo e a vocalização condizentes com a obra, e recomenda evitar uma entonação (prosódia) afetiva[3].

[3] **"**A prosódia afetiva é definida como o processamento e o reconhecimento de elementos emocionais e afetivos provindos das informações da entoação vocal." (Jorge, 2018)

Finalmente, é preciso destacar que os parâmetros variam de país para país e de autor para autor, ou seja, não são fórmulas universais. Por não haver uns parâmetros definidos para a narração de AD no Brasil, na análise da narração de AD com voz sintética que se desenvolverá mais adiante, serão aplicados os seguintes parâmetros de narração explicitados anteriormente:

- **entonação/emoção:** segundo o *Guia para produções audiovisuais acessíveis (2016), Charte de l'audiodescription* (2008) e AENOR (2005);
- **velocidade**: de acordo com o estabelecido pela Independent Television Commission (2000);
- **voz natural:** será analisado se a voz sintética se aproxima de uma voz natural, além de seguir recomendações da American Council of the Blind (2009) na escolha da voz ;
- **ritmo/fluidez:** como os parâmetros anteriormente citados foram definidos para narradores humanos e não para vozes sintéticas, também será levada em consideração, na análise, a questão da não robotização e fluidez na fala da voz sintética escolhida, fatores que influenciam no quesito voz natural.

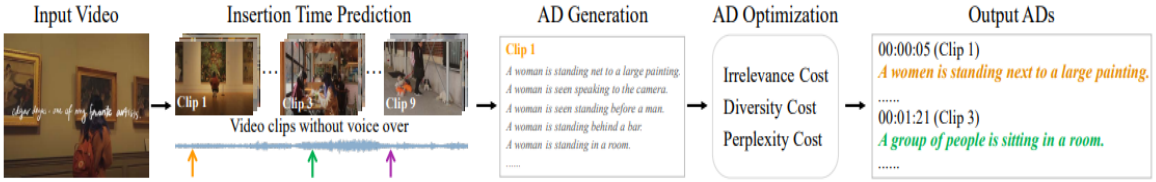
5.3 Geração semi-automática de AD

A tentativa de automatizar o processo de audiodescrição não é muito antiga. A brasileira Virginia Campos (2015) realizou um sistema chamado CineAD, que consiste em gerar automaticamente roteiros de audiodescrição de filmes a partir do roteiro do filme e da legenda. O sistema funciona da seguinte maneira:

> 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 (CAMPOS, 2015, p. 40).

O sistema de semi-automação de AD *Tiresias* desenvolvido por japoneses (Liang *et al.,* 2021) consiste em analisar o conteúdo audiovisual de um vídeo para gerar as audiodescrições. O sistema compreende três módulos: previsão de tempo de inserção de AD, geração de AD e otimização de AD. A seguinte figura ilustra como o sistema funciona na prática:

FIGURA 2 - VISÃO GERAL DO SISTEMA *TIRESIAS*



Fonte: *Toward Automatic Audio Description Generation for Accessible Videos*

Pode-se perceber que nestes dois sistemas existe automação apenas no processo de roteirização, não incluindo assim a narração de AD. Sendo assim, este trabalho se propõe a automatizar o processo de narração para que, futuramente, as duas etapas da audiodescrição, a saber, criação do roteiro e narração, sejam desenvolvidas em um único sistema que possibilite automação completa e não apenas a semi-automação em uma das etapas.

Vale lembrar que estes sistemas não estão disponíveis para uso geral.

5.4 Utilização de Inteligência Artificial como sintetizador de voz na AD: *text-to speech e deep learning*

Alan Turing, matemático britânico, também conhecido como o pai da computação, fez contribuições em várias áreas, a saber, Ciência da Computação, Ciência Cognitiva e, claro, Inteligência Artificial, assunto em que foi o pioneiro. Em 1950, com o artigo *Computing Machinery and Intelligence*, ele inventou o *Turing test* (Teste de Turing), que consistia em promover a interação de um ser humano e uma máquina. O teste tem o intuito de testar a capacidade de um computador de apresentar comportamento inteligente, assim como de um ser humano, e parte da premissa que se a pessoa não perceber que está interagindo com uma máquina, o computador passa no teste.

Mas, afinal, o que é a Inteligência Artificial? Damaceno e Vasconcelos (2018) a definem da seguinte maneira:

> "[...] tem-se como Inteligência Artificial a confecção de máquinas como capacidade de aprender sendo estas programadas previamente, fazendo uso de algoritmos bem elaborados e complexos que proporcionem a tomada de decisões, especulações e até interações baseadas nos dados fornecidos [...]" (Damaceno; Vasconcelos, 2018, p. 12).

Sendo assim, a IA é "ensinada" a raciocinar como um ser humano ao ser alimentada com dados que irão fundamentar seu conhecimento. Em contraponto com esta definição, Miguel Nicolelis, médico e neurocientista brasileiro, afirmou, em uma fala publicada pelo

Jornal Folha de São Paulo em 08 de julho de 2023, que “a IA não é inteligente”, pois a inteligência é uma propriedade dos seres vivos e “nem artificial” por ter sido criada por seres humanos, assim, o algoritmo não é “inteligente por definição”. Seguindo esta lógica, a IA é uma máquina que só tem a capacidade de reproduzir aquilo que seus algoritmos, criados por seres humanos, estabeleceram para que ela reproduza, não tendo a capacidade em si de ser ensinada a fazer alguma coisa e sim programada.

Independente de como é definida, a IA pode ser subdivida em duas camadas: 1ª) *Machine learning* e 2ª) *Deep learning*. A primeira fornece os dados, seus algoritmos são “estruturados com equações pré-definidas para organizar e executar os dados conforme a demanda” (DAMACENO, VASCONCELOS; 2018). A segunda é a subárea que interessa a esse trabalho, pois é um tipo de *Machine learning* mais desenvolvido, isto é, é a camada da IA que capacita a máquina a realizar tarefas mais complicadas, como reconhecimento de fala e identificação de imagens.

Devido ao avanço tecnológico, atualmente existe a possibilidade de se utilizar programas de IA como sintetizadores de voz, ou seja, é possível que essa tecnologia leia um texto, como por exemplo, um roteiro de audiodescrição. Essa leitura de texto é feita por um sistema que faz a conversão *text-to-speech (TTS)*, que é entendido por:

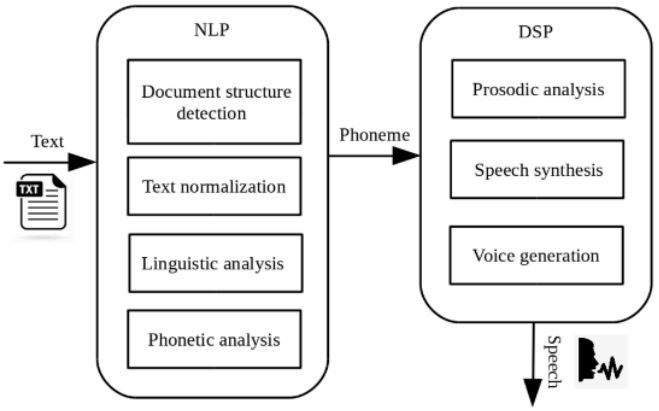
> A síntese TTS (texto para fala) é um mecanismo no qual o texto escrito é convertido em fala. Uma síntese de texto para fala pode seguir duas etapas principais, são elas a análise de texto e geração de onda de fala. Nas últimas décadas, a tecnologia de conversão de texto em fala tem surgido com a ajuda de várias tecnologias, como inteligência artificial e aprendizado de máquina [5]. Usando tecnologias de aprendizado de máquina, a síntese de fala em TTS tem apoiado a renderização artificial de sistemas de computador com fala semelhante à humana (Kumar, Koul, Singh; 2022, p. 2. Tradução própria)[4].

A figura a seguir demonstra como o sistema *text-to-speech* funciona na prática. Basicamente, o sistema TTS é separado em 2 módulos: *NPL* (PLN- Processamento de Linguagem Natural) e *DSP* (PDS- Processamento Digital de Sinais). No primeiro módulo, *NPL*, ocorre o processo de normalização de texto, ou seja, as frases de entrada são organizadas em listas de palavras gerenciáveis. Em seguida, números, abreviações, siglas e palavras específicas são transformadas em um texto completo quando necessário. É nesse módulo que a análise linguística, com relação à transcrição fonética e informações prosódicas,

[4] TTS (text-to-speech) synthesis is a mechanism in which written text is converted into a speech. A text-to-speech synthesis may follow two main steps, i.e., text analysis and speech waveform generation. Over the last few decades, Text-to-speech technology has been emerging out its best with the help of various technologies like artificial intelligence and machine learning [5]. Using machine learning technologies, speech synthesis in TTS has supported the artificial rendering of human-like speech computer systems (Kumar, Koul, Singh; 2022, p. 2)

é realizada pelo sistema. Então, esta parte fonética transcreve os símbolos ortográficos lexicais em representações fonêmicas, com intuito de fornecer a pronúncia correta das palavras. O módulo *DSP*, por sua vez, é baseado na análise prosódica que determina a entonação adequada, taxa de fala e amplitude para cada fonema na transcrição. Dessa forma, ele transforma em fala a informação simbólica que recebeu do primeiro módulo.

FIGURA 3 - ARQUITETURA PADRÃO DO SISTEMA *TEXT-TO-SPEECH*



Fonte: *International Journal of Speech Technology*, 2020

Como observado, esse sistema também faz uso de Processamento de Linguagem Natural, uma vertente da IA que ajuda máquinas a processar e manipular a linguagem humana em diversos níveis. O aperfeiçoamento do sistema TTS ao longo dos anos se deu com os avanços das técnicas de *Deep learning*; assim, a aplicação de *Deep learning* ao sistema TTS faz com que o desempenho da leitura artificial produzida por máquinas seja cada vez melhor. Desse modo, ao se juntar o *Deep learning* e o sistema TTS, tem-se uma Inteligência Artificial capaz de transformar um texto digital em onda sonora, ou melhor, reproduzir um texto em fala.

5.5 Narração humana vs. narração sintética

A voz humana é o instrumento que possibilita o ser humano estabelecer uma comunicação e transmitir por meio dela emoções, ênfases, ideias e as intenções claras e subentendidas do ato comunicativo. Assim, por se tratar de uma ferramenta natural do corpo humano, a narração produzida por pessoas é, obviamente, natural. Nela não existe mecanicidade, falta de fluidez ou ausência de entonação. Em razão disto, por muito tempo a narração feita com voz sintética caiu em descrédito, pois não possuía características que a

aproximasse de uma voz natural. Sempre muito robotizada, sem fluidez na fala e sem a entonação correta na leitura de palavras e frases, o que conferia, por vezes, um sentido diferente do que se pretendia.

Entretanto, uma das melhorias que a IA trouxe na sintetização de voz, no cenário mundial atual, foi a evolução nas vozes. Agora, elas são mais realistas, podendo até mesmo ser a reprodução da voz de um famoso ou da própria voz, possibilidade que alguns programas já oferecem. Além disso, a fluidez teve um avanço significativo, uma vez que se percebe a fala menos travada e a entonação correta das palavras, o que confere mais naturalidade para a narração. Com relação à entonação da narração, alguns programas de IA, tais como LOVO.ai e Play.ht, oferecem a alternativa de ser escolhida a emoção da narração; dessa maneira, você pode, por exemplo, decidir quando a voz deve ser mais animada ou melancólica, assim como ajustar ênfases e pausas. A velocidade é outro atributo que as IAs já disponibilizam para ser adequada conforme a necessidade da narração.

Portanto, o que se observa na lista de atributos oferecidos pelos programas de IA atualmente é uma aproximação significativa das vozes naturais. Este recurso pode, inevitavelmente, revolucionar a quantidade de obras com informações visuais, não só em formato de vídeo como também exposições artísticas e peças teatrais, com audiodescrição disponível, dado que realizar uma narração de AD usando IA como sintetizador de voz é mais viável em custo, elaboração, mixagem e um processo mais rápido com relação ao tempo necessário para executar todas as etapas.

6 Metodologia

Esta pesquisa tem caráter exploratório, pois além de estabelecer um panorama geral dos parâmetros da narração de audiodescrição, visa igualmente explorar o campo tecnológico para a geração semi-automática da audiodescrição, utilizando-se da Inteligência Artificial como sintetizador de voz. Dessa forma, emprega a metodologia qualitativa a partir de pesquisa bibliográfica e análise do comportamento e eficiência da IA na narração de audiodescrição.

Com relação aos procedimentos, o primeiro passo é a escolha de um vídeo, seja série ou filme, que já possua audiodescrição feita por um ser humano. Em seguida, o roteiro de audiodescrição do episódio será inserido no programa de IA para a produção de uma narração de audiodescrição semi-automática. Desse modo, será feita a análise do comportamento e eficiência da narração feita pela IA, na tentativa de averiguar se ela se aproxima da narração

original feita por um ser humano. Para isto, questões como entonação/emoção, ritmo/fluidez, velocidade e voz natural serão analisadas, levando em consideração os parâmetros de AD descritos anteriormente.

7 Experimento

O vídeo escolhido para testar o comportamento de uma IA como um sintetizador de voz foi o episódio “A revolta dos carecas", da Turma da Mônica, extraído do canal da Turma da Mônica no YouTube. Para a decisão de utilizar este episódio no experimento, foi levado em consideração o fato de ser um desenho que já possuía narração de AD feita por um ser humano[5] e, assim, o programa de IA trabalharia em cima do roteiro de AD já existente. Além disso, o desenho é dinâmico e com muitas falas, o que é interessante para analisar o desempenho da IA frente ao *time-code* estabelecido no roteiro.

Já o programa de IA utilizado para este experimento foi o *Genny*, da empresa LOVO. Ele está disponível para uso no site *Lovo.ai* e é uma espécie de editor de vídeo online. O programa promete vozes naturais e humanas, podendo criar conteúdo com mais de 400 vozes em 100 idiomas e vários sotaques. Além disso, as vozes podem expressar mais de 25 emoções, enquanto o tom e a ênfase das frases também podem ser ajustadas.

7.1 “A revolta dos carecas”, *Turma da Mônica*

Este episódio faz parte do Cine Gibi 7 e possui em torno de 7 minutos de duração. Também conta com dublagem em espanhol e inglês. Além da audiodescrição, o episódio apresenta outra ferramenta de acessibilidade, a tradução em libras, no canal da Turma da Mônica no YouTube.

Neste episódio, Cebolinha e Cascão vão até a casa de seu amigo Xaveco com o intuito de chamar sua irmã Xabéu, pois ela era muito bonita, para protagonizar um filme sobre os personagens secundários da Turma da Mônica. Contudo, Xaveco gostaria de participar deste filme, então, finge ser a Xabéu para convencer seus amigos de que ele era o personagem secundário certo. Após um certo tempo de conversa entre os três amigos, Xaveco sai correndo desesperado, pois ficou careca. Logo em seguida, Xabéu também aparece correndo sem seu

[5] Com intuito de fornecer mais credibilidade ao trabalho e para mais certeza a respeito da AD do canal da Turma da Mônica no YouTube, a assessoria de imprensa da Maurício de Sousa Produções foi contactada para responder se a narração de AD dos vídeos do canal são vozes sintéticas ou de ser humano. Na resposta, a assessoria explicou que não usa método automático e que a voz é sempre de um profissional da área.

cabelo. Então, ao sair da casa de Xaveco, Cebolinha e Cascão começam a ver que outros personagens também perderam o cabelo, como Denise, Magali e Mônica.

Uma personagem chamada Madame Capilar aparece perto deles e confessa que tem roubado o cabelo das pessoas para fabricar perucas em sua máquina sugadora de cabelos e assim todos os personagens carecas teriam que recorrer a ela para comprar perucas. Madame Capilar rouba o cabelo de Mônica e quando coloca Cascão na máquina para sugar seu cabelo, ela acaba explodindo, pois com tanta sujeira o cabelo dele entope a máquina. Após a explosão, Madame Capilar aparece careca e Cebolinha, Mônica e Cascão descobrem que ela usava peruca. O episódio termina quando os três amigos percebem que Madame Capilar deixou para trás uma caixa cheia de perucas ao sair correndo depois de ter seu segredo descoberto, assim, eles vão até ela e recuperam seus cabelos. Cebolinha, que não teve seu cabelo roubado por ter poucos fios, também pega uma peruca de cabelos longos e castanhos para usar.

7.2 Roteiro de AD do episódio

A narração de AD deste episódio foi feita com voz feminina, já que a narração do título do desenho está em voz masculina. Esclareço que a tabela abaixo foi criada por mim, com base na transcrição da AD, já que o roteiro de audiodescrição dos episódios de Turma da Mônica não está em domínio público. Entretanto, nenhuma alteração foi feita, o *time-code*, as deixas[6] para a narração e as unidades descritivas[7] da narração foram mantidas as mesmas. Já as rubricas, ou seja, as instruções para a narração, não foram inseridas, pois se trata de um elemento que costuma ser utilizado para auxiliar os narradores humanos no momento da gravação. Toda a informação extraída do episódio foi analisada cuidadosamente. Seguem abaixo, dispostas na Tabela 1, as unidades descritivas da narração de AD extraídas do episódio:

TABELA 1: UNIDADES DESCRITIVAS EXTRAÍDAS DO EPISÓDIO

Time-Code	Audiodescrição
00:00:06 - 00:00:07	Assim que a primeira cena aparece.

[6] Deixas: "[...] diálogos finais antes do início da descrição" (Campos, 2019, p. 10).
[7] Unidade descritiva: "composição com informações para o narrador da audiodescrição. Contém o texto a ser narrado e o seu ponto de inserção indicado com marcação, seja time code in (tc in), time code out (tc out), ou equivalente" (ABNT, 2016, p. 3).

	Na rua, carros passam.
00:00:12 - 00:00:14	Quando o terceiro carro passa. Na calçada, Cebolinha e Cascão caminham.
00:00:26 - 00:00:27	- Eu tive uma ideia genial. Hahaha. É o seguinte: → DEIXA Casa de Xaveco. Na janela…
00:01:16 - 00:01:16	- Aah, seus, seus traíras. → DEIXA Eles entram.
00:01:44 - 00:01:44	- Grrr, grandes amigos. → DEIXA Xaveco sai.
00:01:54 00:01: 57	- Xabéu… → DEIXA Atrás da parede, Xaveco aparece em cima de um banquinho com uma toalha na cabeça.
00:02:39 - 00:02:42	- O público vai amar e… Ahhh → DEIXA Uma sombra de duas mãos aparece atrás de Xaveco.
00:03:10 - 00:03:10	- É. Tem que ter pelo menos cinco fios de cabelo, né? → DEIXA Eles saem da casa.
00:03:12 - 00:03:14	- Aah, meu cabelo, meu cabelo! → DEIXA Uma menina de vestido rosa passa correndo careca.
00:03:27 - 00:03:29	- Ahhh, meu cabelo, meu cabelo. → DEIXA Magali passa correndo careca comendo uma melancia.
00:04:05 - 00:04:06	- Ora, seu... → DEIXA Mônica careca dá uma coelhada nele.
00:04:54 - 00:04:56	- Ei, ei, ei, ei, ei. Eu ainda estou aqui. Silêncio! → DEIXA Madame Capilar enrola os meninos com o seu cabelo.
00:04:58 - 00:04:59	- Só faltava vocês dois nesse bairro. Hahaha. → DEIXA

	Ergue-os com o cabelo.
00:05:13 - 00:05:17	- Sobrou você. Hahahaha. → DEIXA Ela coloca Cascão na máquina que prende o sugador de cabelo em sua cabeça e arranca todo o seu cabelo.
00:05:30 - 00:05:32	- Mais uma peruca saindo… Hahaha. → DEIXA As perucas caem dentro de uma caixa.
00:05:45 - 00:05:47	- É melhor a gente sair daqui, essa coisa vai explodiiiiiir! → DEIXA Uma fumaça preta sai da máquina.
00:06:14 - 00:06:16	- Estou arruinada! Ahhh. → DEIXA Madame Capilar sai correndo.
00:06:27 - 00:06:28	- Vejam! Ficou uma caixa cheia de perucas. → DEIXA Eles correm até a caixa.
00:06:38 - 00:06:39	- Achei um legal. Que tal? → DEIXA Põe de volta seu cabelo.
00:06:42 - 00:06:43	- Ah, que linda cabeleira. Agora *tô* gato. → DEIXA E Cascão sua moitinha.
00:06:48 - 00:06:50	- Eh, nem tudo… → DEIXA Cebolinha está com um cabelão castanho e liso.

Fonte: *Turma da Mônica*, canal oficial no Youtube

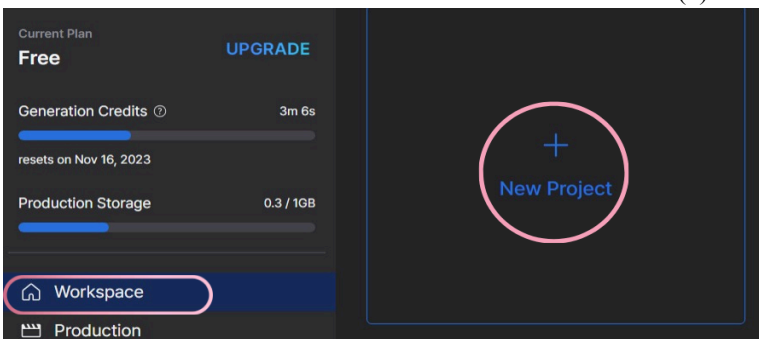
Apesar de não ser possível o acesso às rubricas do roteiro de AD original, o que se pode perceber ao longo do episódio é que em alguns casos a narração é sobreposta rapidamente ao final de alguma fala, risada, grito ou trilha sonora, também houve momentos em que a narração foi mais rápida para caber no tempo disponível. As deixas destacadas em vermelho marcam quando as unidades descritivas foram sobrepostas no desenho. Nesse caso, quando utilizado o programa de IA como sintetizador de voz, a velocidade das falas será mantida a mesma, assim como os casos em que houve sobreposição.

7.3 *Genny* (LOVO.AI)

A segunda etapa deste experimento foi inserir este roteiro no programa *Genny*, que utiliza a IA para gerar vozes sintéticas, ou seja, funciona como um sintetizador de voz. O programa promete oferecer 400 vozes naturais e humanas em 100 idiomas diferentes e vários sotaques. Além disso, a plataforma promete o ajuste de emoções, tom, velocidade, pausas e ênfase na leitura das frases.

O primeiro passo para criar a narração de AD na plataforma é criar um novo projeto. Para isso, é necessário clicar em "Workspace" e então em "New Project".

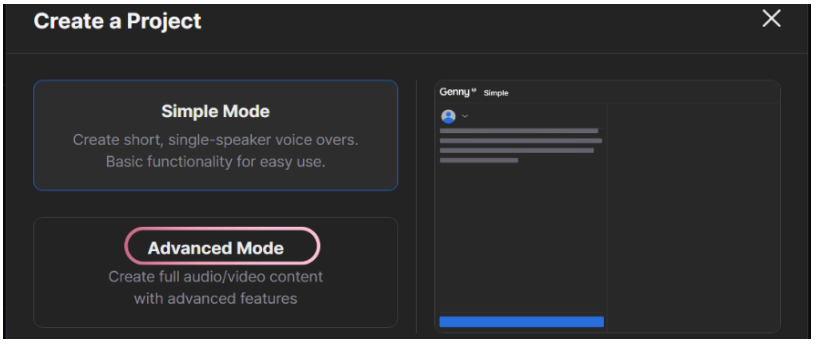
FIGURA 4 - PASSO A PASSO NO PROGRAMA GENNY (1)



Fonte: Plataforma *Genny, LOVO.AI*.

Em seguida, duas opções de modelo de projeto aparecerão. O "Simple Mode" possibilita que o usuário transforme textos em fala sem mídia visual, com a alternativa de inserir textos manualmente ou importando um documento para a plataforma. Já o "Advanced Mode" proporciona ao usuário a edição das vozes sintéticas em vídeos, contudo, as unidades descritivas são inseridas manualmente. Este último foi o modelo escolhido para o experimento, já que proporciona a mixagem do vídeo com a narração de AD.

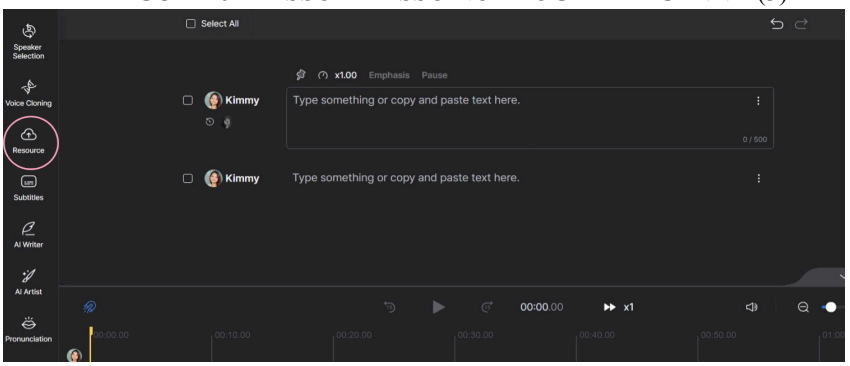
FIGURA 5 - PASSO A PASSO NO PROGRAMA GENNY (2)



Fonte: Plataforma *Genny, LOVO.AI*.

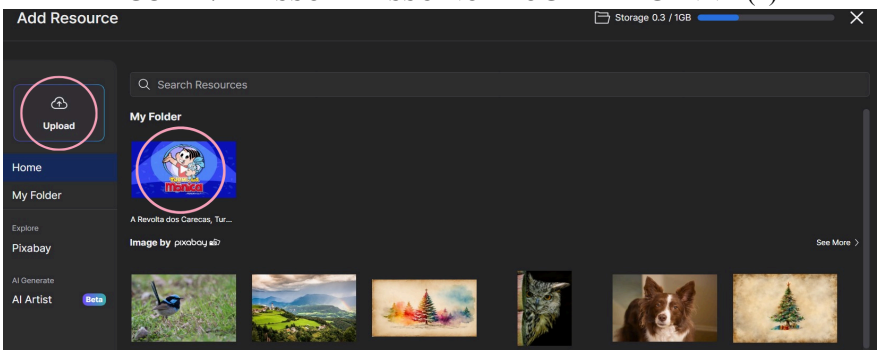
Quando a tela de edição aparecer, é necessário clicar em “Resource” para inserir na plataforma o vídeo em que a AD será mixada, o qual já deve estar salvo no computador. Em seguida deve-se clicar em “Upload” e selecionar o vídeo; vale ressaltar que este deve estar em formato MP4.

FIGURA 6 - PASSO A PASSO NO PROGRAMA GENNY (3)



Fonte: Plataforma *Genny, LOVO.AI*.

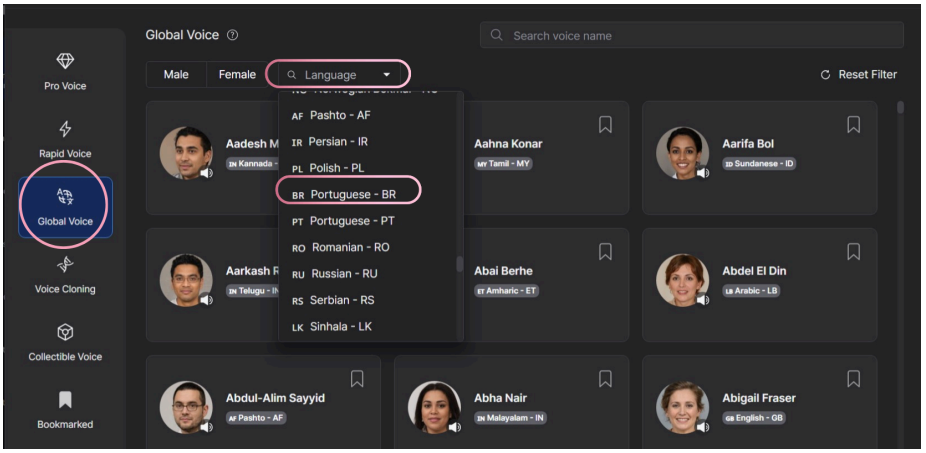
FIGURA 7 - PASSO A PASSO NO PROGRAMA GENNY (4)



Fonte: Plataforma *Genny, LOVO.AI*.

O próximo passo, após selecionar o vídeo, é escolher uma voz para o projeto. Para isso, basta clicar em “Speak Selection” no canto superior direito da tela. Em seguida, pode-se clicar em “Global Voice” para ter acesso a outras vozes estrangeiras além do inglês. Ao clicar em “Language”, uma gama de idiomas surgirá, basta descer o cursor até encontrar o idioma desejado, que, no caso deste projeto, é o português. A voz escolhida foi a voz de nome “Marcia”. É importante observar que primeiro se escolhe o idioma e depois a voz que irá ser usada, pois as vozes da plataforma não falam em vários idiomas, mas apenas em um.

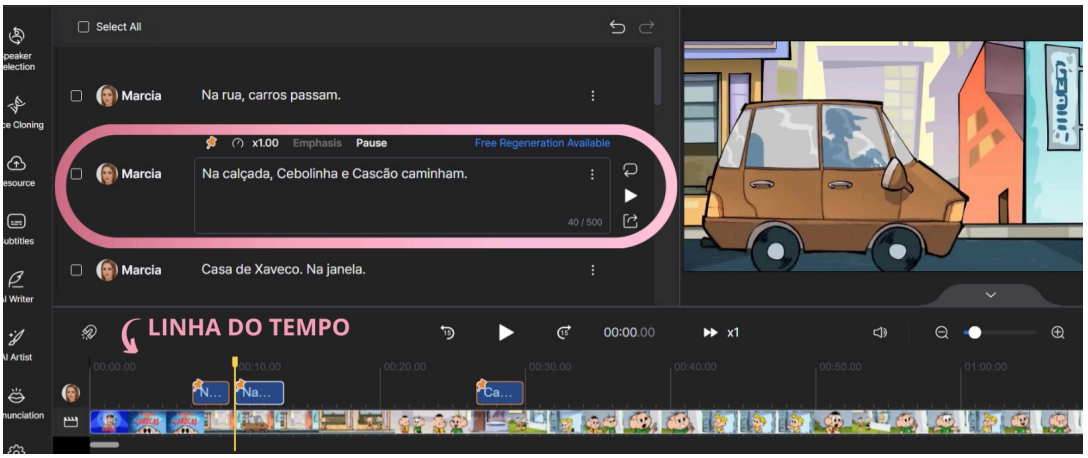
FIGURA 8 - PASSO A PASSO NO PROGRAMA GENNY (5)



Fonte: Plataforma *Genny, LOVO.AI*.

Finalmente, a edição pode começar. Para gerar a narração é necessário apenas inserir as unidades descritivas. Estas podem ser sincronizadas com o vídeo ao movê-las na linha do tempo na parte inferior da tela.

FIGURA 9 - PASSO A PASSO NO PROGRAMA GENNY (6)



Fonte: Plataforma *Genny, LOVO.AI*.

A plataforma também permite que a pronúncia de palavras seja corrigida manualmente, assim como proporciona aos usuários o ajuste da velocidade da leitura, a ênfase e as pausas.

É importante ressaltar que a exposição do passo a passo de como a narração foi feita no programa *Genny* não inclui uma análise completa da qualidade da plataforma e das ferramentas que ela disponibiliza aos usuários.

7.4 Resultados

Atendendo ao objetivo proposto nesta dissertação, a saber, analisar o desempenho da narração de AD feita por uma IA levando em consideração os parâmetros para a narração citados anteriormente e julgando questões como entonação, ritmo/fluidez, velocidade e voz natural, na tentativa de constatar o comportamento e eficiência da IA e averiguar se a narração com voz sintética se aproxima da narração original feita por um ser humano, segue a seguir a análise realizada.

- **Entonação/emoção:**

A plataforma promete a edição da entonação, emoção e ênfase. A ênfase pode ser alterada em cada entrada de unidade descritiva. Entretanto, não foi necessário editá-la, pois o desempenho da IA foi satisfatório, ou seja, sua leitura foi correta em todas as unidades descritivas. A entonação/emoção da leitura, por sua vez, só pode ser alterada no idioma inglês em vozes específicas de narração de *audiobook*, todas as opções de vozes em português não são próprias para narração. Sendo assim, a narração de AD ficou com tom neutro, e, apesar de não soar monótono, destoa das referências apontadas anteriormente e escolhidas como parâmetro para a análise.

- **Ritmo/fluidez:**

A voz sintética da Genny confere um ritmo mais natural à leitura, levando em consideração vozes sintéticas mais antigas. Além de não errar a pronúncia das palavras, a voz sintética consegue ler respeitando a pontuação, proporcionando assim mais fluidez na leitura, sem as pausas indesejáveis e fora de hora, comum das vozes sintéticas mais antigas. Outro ponto forte da plataforma é poder ensinar à IA como se pronunciam as palavras na opção "Pronunciation", à esquerda na parte inferior da tela.

- **Velocidade:**

A velocidade pode ser modificada em cada entrada de unidade descritiva, sendo assim, é um ponto positivo para a plataforma, já que se pode ajustar a fala ao *time-code* do roteiro. Dessa forma, pode-se adaptar a qualquer parâmetro escolhido para a AD.

- **Voz natural:**

Infelizmente, as vozes em português não apresentam um desempenho tão bom quanto as vozes Nível Pro do idioma inglês na plataforma. As vozes em português ainda apresentam resquícios de robotização, o que não ocorre com as vozes Nível Pro do idioma inglês. Contudo, isto não é um problema exclusivo da Genny; outras plataformas como Animaker, Murf.ai e NaturalReader evidenciam o mesmo problema: o nível de desenvolvimento de

vozes em inglês é consideravelmente superior a outras línguas. Todavia, apesar de ainda não estarem perfeitas, as vozes em português da plataforma trazem um desempenho melhor do que as vozes sintéticas do passado, a título de exemplo, as usadas no Google Tradutor e Waze; ou seja, houve significativa melhora nelas. Quanto à escolha da voz, foi levado em consideração que o vídeo já possuía um narrador homem na leitura do título do episódio, então foi escolhida uma voz feminina para a narração de AD, em conformidade com os parâmetros discutidos anteriormente.

8 Conclusão

Diante disso, conclui-se que a narração com voz sintética se aproxima da narração original feita por um ser humano, em partes na língua portuguesa, mas totalmente na língua inglesa na plataforma escolhida. Embora a plataforma *Genny*, e muitas outras, invistam mais no desempenho da língua inglesa, deixando a desejar em outros idiomas, as vozes e os sotaques Nível Pro em inglês são realistas e naturais, sem resquícios de robotização. O fato de a narração das vozes em português ainda não estarem completamente iguais à narração feita por um ser humano não invalida a qualidade das vozes em inglês e o seu desempenho na narração. Isto indica que o Brasil precisa desenvolver plataformas que utilizem a IA como sintetizador de voz, com foco em melhorar o desempenho das vozes no português, uma vez que as vozes em inglês se mostram melhores por serem programas desenvolvidos por norte-americanos. No entanto, o nível das vozes em inglês mostra a capacidade de avanço das vozes sintéticas modernas na língua portuguesa, sendo isto um sinal promissor da eficiência da IA na narração de AD.

Sem embargo, o custo-benefício da automação tanto da elaboração de roteiro quanto da narração de AD é algo a se considerar, pois, além de ser mais barato, as etapas da AD são realizadas de maneira mais rápida que a convencional, promovendo assim uma pós-edição mais breve das produções. Tudo isso se soma ao fato de que, automatizando o processo de criação e execução da AD, mais produções terão esta ferramenta assistiva. Portanto, este trabalho de conclusão de curso se propôs a apontar o futuro da automação da AD ao mostrar a possibilidade de se utilizar a IA na narração como sintetizador de voz. Ainda que as plataformas de elaboração automática de roteiros de AD não estejam disponíveis para uso público e que a narração com voz sintética precise de ajustes, o futuro da AD completamente automatizada e de qualidade aceitável se mostra mais palpável a cada dia.

Quanto ao futuro dos profissionais da área, a discussão que se pode prever é a respeito de como eles poderão se adaptar às novas tecnologias. Como observado no trabalho, a narração de AD foi feita por um programa de IA, contudo, seu manuseio foi feito por um ser humano. Neste caso, ao se fazer um paralelo com o futuro dos tradutores, que se tornam cada vez mais revisores de textos traduzidos por tradutores automáticos, talvez os audiodescritores se tornem corretores de roteiros de AD criados de forma automática, pessoas que manusearão os programas de IA para ajustar a narração com voz sintética dentro dos parâmetros da AD e os revisores de qualidade do produto final. Assim, por mais que a criação e realização da AD sejam automatizadas, será necessário seres humanos por trás dos processos, dessa vez, não criando roteiros do zero ou narrando a AD, mas revisando e, se necessário, melhorando, o desempenho da máquina nas tarefas delegadas.

9 Referências bibliográficas

AENOR. *Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías.* **Bibbase.org.** Disponível em: <https://bibbase.org/network/publication/aenor-une153020audiodescripcinparapersonascondiscapacidadvisualrequisitosrequisitosparalaaudiodescripcinyelaboracindeaudioguas-2005> Acessado em 08 de ago. de 2023.

ALI, I; MNASRI, Z; LACHIRI, Z. *DNN-based grapheme-to-phoneme conversion for Arabic text-to-speech synthesis*. **Researchgate.net.** Disponível em: <https://www.researchgate.net/publication/343861915_DNN-based_grapheme-to-phoneme_conversion_for_Arabic_text-to-speech_synthesis#fullTextFileContent> Acessado em 10 de ago. de 2023.

American Council of the Blind. *Audio Description Standards*. **Abc.org.** Disponível em : <http://www.acb.org/adp/docs/ADP_Standards.doc> Acessado em 10 de ago. de 2023.

Audio description coalition. *Margaret Rockwell: a mãe da audiodescrição*. **Vercompalavras.com.br.** Disponível em: <http://vercompalavras.com.br/blog/margaret-rockwell-a-mae-da-audiodescricao/> Acessado em 05 de jul. de 2023.

BERNSTEIN, Adam. *A local life: Margaret Pfanstiehl, 76, blind activist*. **Washingtonpost.com.** Disponível em: <https://www.washingtonpost.com/wp-dyn/content/article/2009/10/03/AR2009100302661.html> Acessado em 05 de jul. de 2023

B. J. Copeland. *Alan Turing*. **Britannica.com.** Disponível em: <https://www.britannica.com/biography/Alan-Turing> Acessado em 07 de ago. de 2023.

CABEZA-CÁCERES, Cristóbal. *Efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió fílmica.* **Tdx.cat.** Disponível em: <https://www.tdx.cat/bitstream/handle/10803/113556/ccc1de1.pdf?sequence=1&isAllowed=> Acessado em 08 de ago. de 2023.

CAMPO, Virginia Pinto. *UM SISTEMA DE GERAÇÃO AUTOMÁTICA DE ROTEIROS DE AUDIODESCRIÇÃO*. 2015. **Repositorio.ufpb.br.** Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/tede/7860/2/arquivototal.pdf> Acessado em 14 de ago. de 2023.

CAMPO, Virginia Pinto. *Sistema de Geração Automática de Audiodescrição a Partir de Análise de Conteúdo de Vídeo.* 2019. **Repositorio.ufpb.br.** Disponível em: <https://repositorio.ufrn.br/bitstream/123456789/28616/1/Sistemageracaoautomatica_Campos_2019.pdf> Acessado em 19 de set. de 2023.

CASTELLS, Manuel. *Sociedade em Rede*. **Wordpress.com.** Disponível em: <https://perguntasaopo.files.wordpress.com/2011/02/castells_1999_parte1_cap1.pdf> Acessado em 02 de out. de 2023.

DAMACENO, Siuari; VASCONCELOS, Rafael. *Inteligência Artificial: uma breve abordagem sobre seu conceito real e o conhecimento popular.* **Periodicos.grupotiradentes.com.** Disponível em: <https://periodicos.grupotiradentes.com/cadernoexatas/article/view/5729/2966> Acessado em 06 de jul. de 2023.

Enap. *Introdução à Audiodescrição.* **Repositorio.enap.gov.br.** Disponível em: <https://repositorio.enap.gov.br/bitstream/1/5299/1/Mod_1_Introdu%C3%A7%C3%A3o%20%C3%A0%20Audiodescri%C3%A7%C3%A3o.pdf> Acessado em 08 de set. de 2023.

Fundação Dorina Nowill Para Cegos. *Apostila Audiodescrição Claudia Scheer*. **Trocandosaberes.com.br**. Disponível em: <https://trocandosaberes.com.br/wp-content/uploads/2022/03/02-Apostila-de-Audiodescricao.pdf> Acessado em 19 de set. de 2023.

IBGE. *Censo demográfico 2010*. **Ibge.gov.br**. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html?edicao=9749&t=destaques> Acessado em 05 de jul. de 2023.

Independent Television Commission. *ITC Guidance On Standards for Audio Description.* **Ofcom.org.** Disponível em: <https://www.ofcom.org.uk/about-ofcom/website/regulator-archives> Acessado em 10 de ago. de 2023.

JAKOBSON, Roman. *Linguística e comunicação*. **Edisciplina.usp.br**. Disponível em: <https://edisciplinas.usp.br/pluginfile.php/2799405/mod_resource/content/1/Aspectos%20lingu%C3%ADsticos%20da%20tradu%C3%A7%C3%A3o%20-%20Roman%20Jakobson.pdf>Acessado em 06 de jul. de 2023.

JORGE, Ana Cristina. *Prosódia afetiva na esquizofrenia.* **Teses.usp.br.** Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8142/tde-24072019-150020/publico/2018_AnaCristinaAparecidaJorge_VOrig.pdf> Acessado em 09 de ago. de 2023.

KUMAR, Yogesh; KOUL, Apeksha; SINGH, Chamkaur. *A deep learning approaches in text-to-speech system: a systematic review and recent research perspective*. **Link.springer.com**. Disponível em: <https://link.springer.com/article/10.1007/s11042-022-13943-4> Acessado em 06 de jul.

LOVO. *AI Voice Generator with Online Video Editor.* **Lovo.ai.** Disponível em: <https://lovo.ai/> Acessado em 22 de ago. de 2023.

Ministério da Educação. *Data reafirma os direitos das pessoas com deficiência visual*. **Potal.mec.gov.br.** Disponível em: <http://portal.mec.gov.br/component/tags/tag/deficiencia-visual> Acessado em 05 de jul. de 2023.

Ministério da Educação. *Deficiência visual*. **Potal.mec.gov.br.** Disponível em: <http://portal.mec.gov.br/seed/arquivos/pdf/deficienciavisual.pdf> Acessado em 05 de jul. de 2023.

Ministério da Educação. *Guia para produções audiovisuais acessíveis.* **Inclusao.enap.gov.br**. Disponível em: <https://inclusao.enap.gov.br/wp-content/uploads/2018/05/Guia-para-Producoes-Audiovisuais-Acessiveis-com-audiodescricao-das-imagens-1.pdf> Acessado em 06 de jul. de

MORISSET, L; GONANT, F. *Charte de l'audiodescription*. **Csa.fr.** Disponível em: <https://www.csa.fr/Media/Files/Espace-Juridique/Chartes/Charte-de-l-audiodescription> Acessado em 08 de ago. de 2023.

MOTTA, Lívia; FILHO, Paulo. *Audiodescrição: transformando imagens em palavras*. **Vercompalavras.com.br.** Disponível em: <http://www.vercompalavras.com.br/download/audiodescricao-transformando-imagens-em-palavras.pdf> Acessado em 06 de jul. de 2023.

NUNES, Elton Vergara *et. al. Mídias do conhecimento: um retrato da audiodescrição no Brasil*. **Guaiaca.ufpel.edu.br**. Disponível em: <http://guaiaca.ufpel.edu.br/handle/123456789/712> Acessado em 06 de jul. de 2023.

LIANG, Wei *et. al. Toward Automatic Audio Description Generation for Accessible Videos.* **Bitwangyuia.github.io.** Disponível em: <https://bitwangyujia.github.io/research/paper/SIGCHI2021-ad.pdf> Acessado em 14 de ago. de 2023.

PEREIRA, Bianca Nathália da Silva. Audiodescrição de desenhos infantis para crianças cegas: uma análise da "Turma da Mônica - o corpo fala". 2020. 35 f., il. Trabalho de Conclusão de Curso (Bacharelado em Línguas Estrangeiras Aplicadas)—Universidade de Brasília, Brasília, 2020. Disponível em: <https://bdm.unb.br/bitstream/10483/27162/1/2020_BiancaNathaliaDaSilvaPereira_tcc.pdf> Acessado em 19 de dez. de 2023.

PIETY, Philip J. *The language system of audio description: a investigation as a discursive process*. **Files.eric.ed.gov.** Disponível em: <https://files.eric.ed.gov/fulltext/EJ683817.pdf> Acessado em 05 de jul. de 2023.

Planalto. *Lei Nº 13.146/2015*. **Planalto.gov.br.** Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm> Acessado em 05 de jul. de 2023.

SALWAY, Andrew. *A corpus-based analysis of audio description*. **Citeseerx.ist.psu.edu**. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.139.1116&rep=rep1&type=pdf> Acessado em 06 de jul. de 2023.

Secretaria especial dos direitos humanos e Coordenadoria nacional para integração da pessoa portadora de deficiência. *Ata VII Reunião do Comitê de Ajudas Técnicas- CAT*. **Assistiva.com.br**. Disponível em: <https://www.assistiva.com.br/Ata_VII_Reuni%C3%A3o_do_Comite_de_Ajudas_T%C3%A9cnicas.pdf> Acessado em 03 de out. de 2023.

SILVA, Adriano. *A tradução intersemiótica de Jakobson revisitada e uma pequena análise dos quadrinhos Asterix*. **Revistadogel.emnuvens.com.br.** Disponível em: <https://revistadogel.emnuvens.com.br/estudos-linguisticos/article/view/1953/1389> Acessado em 06 de jul. de 2023.

TEIXEIRA, C.R.; FIORE, S. F. A.; CARVALHO, B. *Filmes infantis audiodescritos no Brasil: Uma Análise dos Filmes A Turma da Mônica 2 e Hotel Transilvânia*. Traduções & Comunicações – Revista Brasileira de Tradutores, Brasília, Nº 27, 2013.

TEIXEIRA, Pedro. *IA não é inteligência e sim marketing para explorar trabalho humano, diz Nicolelis.* Folha de São Paulo, 08 de julho de 2023. Disponível em: <https://www1.folha.uol.com.br/tec/2023/07/ia-nao-e-inteligencia-e-sim-marketing-para-explorar-trabalho-humano-diz-nicolelis.shtml> acessado em 19 de dez. de 2023.

TURMA DA MÔNICA. *[AUDIODESCRIÇÃO] A revolta dos carecas*. **Youtube.com.br**. Disponível em: <https://www.youtube.com/watch?v=9CdX-xpxraA> Acessado em 21 de ago. de 2023.

TURNING, A. M. *Computing Machinery and Intelligence.* **Redirect.cs.umbc.edu.** Disponível em: <https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf> Acessado em 07 de ago. de 2023.