

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA

RODRIGO OLIVEIRA SOUSA



Desafios, ferramentas e intervenções
na dublagem automática

Uberlândia/MG
2025

RODRIGO OLIVEIRA SOUSA

Desafios, ferramentas e intervenções
na dublagem automática

Monografia apresentada ao Curso de Graduação em Tradução do Instituto de Letras e Linguística da Universidade Federal de Uberlândia como requisito parcial para a obtenção do título de Bacharel em Tradução.

Orientadora: Profa. Dra. Silvana Maria De Jesus.

Uberlândia/MG

2025

RODRIGO OLIVEIRA SOUSA

Desafios, ferramentas e intervenções
na dublagem automática

Monografia apresentada ao Curso de
Tradução do Instituto de Letras e
Linguística da Universidade Federal de
Uberlândia como requisito parcial para a
obtenção do título de Bacharel em
Tradução

Banca de Avaliação:

Prof. Dra. Silvana Maria de Jesus – UFU
Orientador

Prof. Dra. Tânia Liparini Campos – UFPB
Membro

Prof. Dra. Francine de Assis Silveira – UFU
Membro

Uberlândia/MG, 15 de setembro de 2025

AGRADECIMENTOS

Primeiramente, gostaria de agradecer à Universidade Federal de Uberlândia (UFU), que foi minha casa durante os últimos anos. Estendo também meus agradecimentos ao Curso de Graduação em Tradução, incluso secretaria e corpo docente, pelo tempo dedicado à minha formação.

À minha orientadora Prof.^a Dr.^a Silvana Maria de Jesus, agradeço pela paciência, partilha de conhecimentos e dedicação durante essa trajetória.

À Associação Brasileira de Iyengar Yoga (ABIY), agradeço por autorizar o uso do material que foi pivô de análise neste estudo, permitindo uma investigação rica em detalhes e perspectiva.

Agradeço à minha família. Notadamente à minha mãe Celeste, pai Giovani e avó Flora pela imensurável dedicação, apoio e amor a mim e à minha jornada como ser humano e estudante.

Aos amigos de dentro e fora da universidade, especialmente a Caio, agradeço pelo apoio e conselhos.

À minha namorada Cristiane, agradeço pela paciência, amor e por me encorajar a sempre enxergar o próximo passo.

Finalmente, agradeço a Deus e a todos que, mesmo indiretamente, contribuíram para meu crescimento pessoal e profissional durante os últimos anos.

RESUMO

A presente pesquisa tem como objetivo analisar o campo da dublagem automática no âmbito da Tradução Audiovisual (TAV), impulsionado pelos avanços da Inteligência Artificial (IA), a fim de analisar algumas das principais ferramentas gratuitas disponíveis no mercado e discutir os desafios desse processo. A partir de um estudo exploratório, foi possível identificar as principais características e falhas tecnológicas de dez ferramentas online no que se refere ao processo de dublagem por IA no par linguístico inglês-português. A análise destaca inconsistências recorrentes em clonagem de voz, sincronização labial e erros de tradução, além de ressaltar o papel e a importância da intervenção humana para corrigir as limitações dos sistemas automatizados disponibilizados atualmente, visto como essencial para garantir a qualidade final na tradução de produtos audiovisuais.

Palavras-chave: Dublagem Automática. Tradução Audiovisual (TAV). Inteligência Artificial. Intervenção Humana.

ABSTRACT

This research aims to analyze the field of automatic dubbing within Audiovisual Translation (AVT), driven by advancements in Artificial Intelligence (AI), in order to evaluate some of the main free tools available on online platforms and discuss the challenges of this process. Through an exploratory study, it was possible to identify the main features and technological flaws of online tools, concerning the AI dubbing process for the English-Portuguese language pair. The analysis highlights recurrent inconsistencies in voice cloning, lip-syncing, and translation errors. Furthermore, it emphasizes the importance and role of human intervention in correcting the limitations of currently available automated systems, which is seen as essential to ensure the final quality of translated audiovisual products.

Keywords: Automatic Dubbing. Audiovisual Translation (AVT). Artificial Intelligence. Human Intervention.

LISTA DE ILUSTRAÇÕES

Figura 1 – Interface da plataforma CAMB.AI.....	22
Figura 2 – Recurso de velocidade e alongamento de dublagem.....	23
Figura 3 – Interface da plataforma Deepdub.ai.....	24
Figura 4 – Interface da plataforma Dubverse.ai.....	25
Figura 5 – Visualização de vídeo na plataforma ElevenLabs.....	26
Figura 6 – Painel de opções da plataforma HeyGen.....	27
Figura 7 – Painel de <i>download</i> de vídeo da plataforma HeyGen.....	28
Figura 8 – Interface da plataforma Kapwing.....	29
Figura 9 – Painel de opções da plataforma RASK.AI.....	30
Figura 10 – Interface de edição de dublagem da plataforma RASK.AI.....	30
Figura 11 – Painel de clonagem de voz na plataforma RASK.AI.....	31
Figura 12 – Interface da plataforma RASK.AI.....	31
Figura 13 – Painel de opções da plataforma Vidnoz.....	32
Figura 14 – Interface de edição de dublagem da plataforma Vidnoz.....	32
Figura 15 – Sincronização labial com animação facial da plataforma Vidnoz.....	33
Figura 16 – Recurso de identificação de multi-falantes da Vozo.....	34

Figura 17 – Interface da plataforma Vozo.....	34
Figura 18 – Recurso de chat bot da plataforma Vozo.....	35
Figura 19 – Exemplo de erro gerado na legenda e no áudio.....	38
Figura 20 – Excerto sobre sincronização labial da plataforma Vozo.....	40

SUMÁRIO

INTRODUÇÃO	10
CAPÍTULO 1 – FUNDAMENTAÇÃO TEÓRICA	14
CAPÍTULO 2 – METODOLOGIA.....	18
CAPÍTULO 3 – ANÁLISE DE DADOS.....	21
CONCLUSÕES	44
REFERÊNCIA	46

INTRODUÇÃO

A Tradução Audiovisual (TAV) constitui um campo de estudo ligado à evolução tecnológica dos meios de comunicação. Em um mundo cada vez mais globalizado e inovador no que tange a práticas de produção e distribuição de conteúdo, a TAV desempenha um papel fundamental na superação de barreiras linguísticas e culturais ao permitir acesso a uma variada gama de produtos culturais nas mais diversas línguas.

Nesse contexto de inovação contínua, Chaume e Dotel (2024) observam que a história da TAV se desenvolve paralelamente à história do cinema, abrangendo desde o tempo da televisão até a contemporaneidade, com as plataformas digitais. No Brasil, embora houvesse diferentes períodos de maior destaque entre a legendagem e dublagem, esses processos sempre foram tradicionalmente manuais, com especial atenção às sincronias labiais e produção de legendas (Araújo, 2000 apud Carvalho, 2005). É importante pontuar que isso demanda qualificação e tempo adequados dos profissionais envolvidos.

Atualmente, a era da IA (Inteligência Artificial) tem impulsionado a implementação de sistemas avançados, que aqui chamaremos de plataformas, disponíveis na Web, e que oferecem ferramentas para reconhecimento de fala (*speech to text* - STT), tecnologias de síntese de voz (*text to speech* - TTS) e a dublagem automática¹ (*AI dubbing*), popularizando o acesso à produção audiovisual traduzida. Contudo, apesar desses avanços, a legendagem e dublagem automáticas enfrentam desafios significativos em termos de qualidade² e adaptação cultural.

Isso sugere que a automação sem falhas não é uma realidade.

Nesse cenário de rápidas e dinâmicas transformações, a pesquisa acadêmica brasileira direcionada à TAV revela certo grau de insuficiência. Pesquisas em bases de dados como CAPES e Google Scholar frequentemente resultam em estudos de tradução automática geral ou TAV sem foco em automação. Essa lacuna aponta para a necessidade de investigar e descrever as tecnologias de automação no contexto da TAV no Brasil.

¹ O termo “automática” é consagrado na área, mas, do ponto de vista computacional, o termo mais adequado seria automação que remete ao uso de máquinas para controlar e executar tarefas quase sem interferência humana, empregando instruções previamente programadas e nelas registradas.

² Para uma discussão sobre “qualidade” nos Estudos da Tradução, remetemos o leitor ao verbete Quality da ENTI, disponível em https://www.aieti.eu/enti/quality_ENG/entry.html. Acesso em setembro 2025. Para esta pesquisa, considerou-se a adequação da dublagem ao sentido do áudio fonte, a partir da legenda feita por especialistas.

A motivação inicial para esta pesquisa surgiu de uma conversa sobre um projeto de dublagem automática de vídeos religiosos do português para o inglês, visado principalmente a um público anglófono. A inviabilidade de contato com esta instituição direcionou o estudo para os processos de produção de legendas e dublagens automáticas a partir de vídeos autorais produzidos por minha orientadora.

Para o direcionamento desejado na presente pesquisa, foi solicitada à Tânia Liparini, professora do bacharelado em Tradução da UFPB³ e membro da Associação Brasileira de Iyengar Yoga (ABIY), entidade que representa esse método no Brasil, a autorização para o uso de um vídeo de seu fundador, o mestre B.K.S. Iyengar, uma das mais influentes autoridades em ioga do século XX. O vídeo apresenta o mestre falando em inglês sobre a filosofia iogue, com áudio marcado por seu sotaque e oscilações típicas da fala, possivelmente um fator complicador para as ferramentas de dublagem automática e uma valiosa oportunidade de análise. O vídeo possui uma legenda, produzida pela própria ABIY, que foi utilizada como parâmetro de comparação na avaliação da adequação das dublagens automáticas geradas pelas plataformas.

A pesquisa trabalha com a premissa de que, apesar dos avanços, a intervenção humana permanece essencial em etapas como a pós-edição de legendas, adequação de sincronia labial e validação final de qualidade, para corrigir incongruências que as máquinas não identificam. Conforme Díaz-Cintas (2008) observa, a TAV tende a utilizar formas híbridas de tradução. Essa visão, assim como a presente pesquisa, sugere que a automação coexistirá com a curadoria humana, redefinindo a relação entre homem e máquina sem que esta última suprima integralmente o papel do profissional da dublagem.

As perguntas que norteiam esta pesquisa são:

- Quais recursos são utilizados para a dublagem automática?
- Qual a relevância da inteligência artificial (IA) nesse processo?
- É possível produzir dublagem automática adequada sem intervenção humana?
- Quais são as principais plataformas de dublagem automática disponíveis no mercado? E quais são as características e falhas dessas ferramentas?

³ Coordenadora do projeto de extensão Ytrad que traduz materiais sobre Iyengar Yoga (ver campsite.bio/ytrad.ufpb)

O objetivo geral desta pesquisa é contribuir para o estudo da legendagem e dublagem automáticas, com foco em abordar os tipos de ferramentas necessários para realização dessas tarefas. Os objetivos específicos são:

- Explorar as ferramentas de dublagem automáticas disponíveis e os recursos oferecidos;
- Entender quais são os atuais recursos utilizados para a dublagem automática;
- Investigar se a dublagem automática, ou seja, sem intervenção humana, pode ser considerada adequada nos dias de hoje;
- Analisar o papel da intervenção humana no processo de dublagem automática.

O presente estudo se alinha às demandas contemporâneas por acessibilidade em plataformas digitais, áreas em que a tradução audiovisual, incluindo a audiodescrição e a localização, tem papel crucial. O avanço do uso de IA no mercado audiovisual provoca o setor acadêmico a investigar essas novas práticas alinhadas às novas necessidades desse mercado, sem perder de vista as questões éticas, sociais e humanas. Isso pode ser feito ao preparar tradutores para um cenário cada vez mais híbrido, onde o domínio de ferramentas automatizadas e a capacidade de corrigir suas limitações são competências complementares e indispensáveis.

A justificativa deste trabalho se encontra na necessidade de investigar os avanços das tecnologias utilizadas para legendas e dublagem de produtos audiovisuais no mundo contemporâneo. Passando pelo teatro, cinema, TV, até chegar às redes sociais, a legendagem e a dublagem são recursos de integração entre diferentes povos e culturas. Diante da globalização e difusão de produtos digitais, tornou-se cada vez mais acessível a possibilidade de pessoas leigas conseguirem, por conta própria e por razões pessoais, traduzir, legendar e dublar produtos audiovisuais de maneira automática. Entretanto, é necessário investigar que tipo de desafios são encontrados nesse caminho, inclusos processos e intervenções humanas necessárias durante esse processo automatizado.

As questões que nortearam este estudo foram a produção da dublagem automática, a relevância da IA nesse processo, o papel da intervenção humana, as plataformas e ferramentas disponíveis no mercado e, enfim, suas características e

falhas. Espera-se compreender o papel da intervenção humana para a produção de dublagens automáticas com padrões de qualidade adequados e profissionais.

CAPÍTULO 1 – FUNDAMENTAÇÃO TEÓRICA

Carvalho (2005, p. 19) define a dublagem⁴ como "a substituição dos canais de áudio com o texto oral na língua original por um canal de áudio com o texto traduzido". Por outro lado, a legendagem caracteriza-se pela manutenção do áudio na língua fonte e apresentação escrita da tradução dos enunciados para a língua alvo, colocados de forma escrita na parte inferior da tela e "exibidas em sincronia com as falas e eventuais textos escritos do material audiovisual" (Carvalho, 2005, p. 95). A dublagem e a legendagem constituem as principais modalidades de tradução audiovisual.

A tradução audiovisual (TAV) constitui um campo de estudos que acompanha a evolução tecnológica dos meios de comunicação. Como observam Chaume e Dotel (2024, p. 1), "a história da tradução audiovisual transcorre, não sendo de outra forma, de maneira paralela à história do cinema". Conforme descrito por Araújo (2000 apud Carvalho, 2005), não houve predominância entre produtos legendados ou dublados no Brasil, mesmo após o decreto de 1962, de Jânio Quadros, que tornou obrigatória a dublagem para produtos estrangeiros na televisão aberta. Nesse período, os processos eram manuais com especial atenção às sincronias labiais e produção de legendas.

Após o auge do período de transmissões de filmes e séries legendadas ou dubladas na televisão brasileira, houve a popularização do VHS e DVDs, o que implicou em avanços relacionados à TAV e nova dinâmica no consumo de produtos audiovisuais. Durante a primeira década do século XXI, ferramentas de legendagem tornaram-se mais acessíveis e utilizadas pelo público em geral. Um exemplo desse tipo de ferramenta é o Subtitle Edit, que, como outras, permite a automatização parcial de processos como *timecoding* e *spotting*⁵. Contudo, como ressalta Díaz-Cintas (2008), embora a TAV tenha existido como prática profissional por muitos anos, somente a partir dos anos 1990 ganhou maior reconhecimento acadêmico, acompanhando a proliferação e distribuição de materiais audiovisuais na sociedade.

⁴ Para uma visão mais detalhada sobre a dublagem, remeto o leitor ao livro *O Processo da Tradução para a Dublagem Brasileira: teoria e prática* de Dilma Machado (2016).

⁵ *Timecoding* é o processo de marcação precisa de horas, minutos e segundos (no formato HH:MM:SS:FFF) para sincronizar legendas ou áudio, enquanto o *Spotting* é o processo de identificar e marcar a entrada e saída de diálogos e sons diegéticos de um vídeo.

Embora a pesquisa brasileira apresente relevantes trabalhos no campo da TAV, como Franco e Araújo (2011) e Carvalho (2005), trabalhos acadêmicos brasileiros que abordem especificamente a tradução audiovisual automática são escassos. Buscas realizadas no Portal de Periódicos da CAPES, no Google Scholar e em repositórios institucionais (como os da USP e UFMG) frequentemente resultam em artigos sobre tradução automática geral, que não são aplicadas ao contexto audiovisual, como o artigo “Tradução automática, pós-edição e o pronome indefinido ‘man’”, de Bolacio Filho et al (2025) ou em estudos de TAV sem foco em automação, como o artigo “Legendagem tradicional e seus parâmetros técnicos: uma estratégia formal de tradução audiovisual”, de Peres (2023). Quando se adicionam termos como "IA", "TAV" ou "automática" ao filtro de busca dos sites anteriormente citados, os resultados são irrelevantes ou, novamente, tratam de áreas correlatas. Isto aponta que poucos são os artigos ou dissertações que analisam essas tecnologias no contexto brasileiro. Esta lacuna sugere que o tema da automação via inteligência artificial é uma vertente a ser investigada e descrita no campo da tradução audiovisual.

Atualmente, na fase da inteligência artificial, observa-se a criação de plataformas que vendem um sistema, que Federico et al. (2020) denominam tradução fala-a-fala (em inglês, *speech to speech translation* – STS), que funciona como um *pipeline* (linha de montagem) que combina reconhecimento de fala (em inglês, *automatic speech recognition* – ASR), tradução automática (em inglês, *machine translation* – MT) e síntese de fala (em inglês, *text to speech* – TTS), ou seja, a plataforma capta o áudio fonte, faz uma tradução automática e gera o novo áudio para o vídeo dublado na língua alvo. Ainda segundo Federico et al. (2020), o objetivo desta tecnologia é gerar um resultado que seja “temporal, prosódico, e acusticamente próximo ao áudio original”.

Ao mesmo tempo em que ferramentas de auxílio à TAV avançam e se tornam acessíveis, a nova dinâmica da comunicação de massa nas redes sociais tem demandado soluções cada vez mais ágeis para a acessibilidade de conteúdos. Plataformas como YouTube, TikTok e Instagram já incorporam dispositivos que geram legendas automáticas (a partir da tecnologia *speech to text* – STT), enquanto serviços de *streaming* como a *Amazon Prime Video*, anunciaram em 2025 (TUDOCELULAR) investimentos em dublagens feitas por IA para o inglês e o espanhol (latino-americano).

Ainda segundo a empresa TUDOCELULAR (2025), o sistema adotado pelas empresas desse ramo “combina IA e revisão humana para garantir a qualidade das vozes e a naturalidade das interpretações”. É possível supor que a nova dinâmica das redes sociais, que é marcada por vídeos curtos, rápidos e uso de linguagem coloquial, possa expor os limites da automação dos processos da TAV que conhecemos hoje, como más traduções de termos idiomáticos, falha de sincrônica e vozes mecanizadas.

Apesar dos avanços tecnológicos, a legendagem e dublagem automáticas parecem enfrentar desafios significativos, especialmente em relação à qualidade da tradução e à adaptação cultural. Ferramentas como ElevenLabs prometem oferecer tecnologia eficiente para síntese de voz e incluem orientações para intervenção humana visando à correção de ambiguidades, gírias e expressões idiomáticas, por exemplo. A precisão da transcrição inicial feita pelo reconhecimento de fala, por exemplo, é frequentemente afetada por “variações dialetais” (Azevedo, Melo e Côrtes, 2024, p. 30), que é, em outras palavras, a maneira na qual a máquina interpreta as diferenças de sotaques e formas como uma língua é falada em diferentes regiões. Além disso, a sincronia temporal depende de técnicas como controle de comprimento (em inglês, *length control*) para ajustar o tamanho do texto traduzido (Federico et al., 2020). Isso implica diretamente em como as ferramentas alteram a velocidade de falas ou até a reformulação de textos orais na dublagem, para a aplicação da sincronização labial de forma mais acurada.

Supõe-se que, mesmo diante da utilização de processos automatizados, a intervenção humana continue crucial para garantir padrões adequados e profissionais, especialmente em pares linguísticos como inglês-português, onde diferenças sintáticas e fonéticas podem gerar erros de dublagem e legendagem. Assim, embora a inteligência artificial possa otimizar tempo e custos, sua eficácia está diretamente ligada à intervenção humana pós-automação, evidenciando o aspecto complementar entre tecnologia e técnica tradutória, seja na edição de legendas, sincronização entre vídeo e falas, ajuste de *timing* ou até mesmo validação cultural.

É nesse contexto que a intervenção humana converge com a qualidade de traduções automáticas e com a velocidade da produção digital nos dias de hoje, evidenciando-se a necessidade de melhor investigar os processos e novos desafios

que a TAV, em conjunto à dublagem e legendagem automáticas vão enfrentar nos próximos anos.

Essa pesquisa pretende investigar a premissa de que, apesar dos avanços no campo da dublagem e legendagem, a intervenção humana permanece essencial em etapas como a edição pós-tradução automática de legendas, ou até mesmo a adequação de sincronia labial para dublagens, e, sobretudo, em relação ao processo de validação final de qualidade, para corrigir incongruências que as máquinas não identificam com facilidade.

Mesmo antes da IA, Díaz-Cintas (2008) já apontava para a tendência na tradução audiovisual de utilização de formas híbridas em processos de tradução. Podemos realocar essa visão para o cenário atual, onde a automação terá mais peso, mas continuará a coexistir com a curadoria humana. Ou seja, acredita-se que a evolução tecnológica não substitui, mas redefine as relações entre homem e máquina, estabelecendo uma nova lógica de colaboração, mas ainda assim, as máquinas não dispõem de meios para suprimir o papel do profissional da dublagem e legendagem.

Diante disso, a pesquisa busca contribuir para o estudo da legendagem e dublagem automáticas, investigando plataformas disponíveis, recursos utilizados e a viabilidade de processos automatizados, com o objetivo de compreender quais são os desafios e intervenções necessárias para conseguir produtos audiovisuais de qualidade com as tecnologias que disponibilizamos hoje. Como foi apontado, o baixo número de pesquisas brasileiras sobre o tema, em contraste com o rápido avanço e alcance de produtos para mídias sociais e plataformas de *streaming* reforçam a importância de mapear esse cenário. Esta investigação permitirá analisar as possibilidades da IA e o potencial de sua integração ao trabalho humano, preenchendo lacunas e contribuindo para uma melhor compreensão do futuro da TAV.

CAPÍTULO 2 – METODOLOGIA

Para a realização dessa pesquisa, foi desenvolvido um estudo de caso exploratório e descritivo da produção de vídeos com dublagem automática em plataformas de IA. Para isso, foram utilizados dois vídeos do YouTube cedidos para a pesquisa, a fim de testar as ferramentas que podem ser utilizadas no processo.

O primeiro vídeo⁶, utilizado para a análise piloto, narra algumas funcionalidades do *software* AntConc e foi produzido pela orientadora dessa pesquisa; o áudio de 3m40s minutos foi traduzido do português para o inglês com a ferramenta ElevenLabs.

O segundo vídeo⁷, utilizado para a análise de várias plataformas, é da Associação Brasileira de Iyengar Yoga (ABIY). Nele, o mestre indiano B.K.S. Iyengar, fundador do método que leva seu nome e uma das mais influentes autoridades em ioga no século XX, fala, em inglês, sobre a filosofia iogue em um áudio de 14 minutos rico em oscilações e sotaque. O vídeo foi dublado para o português brasileiro. A ABIY fez a legenda do vídeo, do inglês para o português, que está anexada ao vídeo, e serviu para compreender a terminologia do áudio fonte e seus correspondentes em português.

A metodologia da pesquisa abarcou as seguintes etapas: i) análise piloto, ii) análise e descrição das plataformas a partir das informações dos sites, iii) produção da dublagem automática em cada ferramenta, gerando 9 vídeos com duração entre 45 segundos e 3 minutos e iv) análise de cada vídeo produzido, visando identificar as características adequadas e as falhas; v) resumo dos dados encontrados.

Análise piloto

Após análise inicial de plataformas de dublagem automáticas disponíveis, foram selecionadas a HeyGen, ElevenLabs, BitFractal, Voiser e Amazon Polly⁸ a fim de verificar a metodologia de pesquisa e especificar inicialmente o processo de dublagem automática.

⁶ Disponível em <https://www.youtube.com/watch?v=Um8vZ5fp00I>. Acesso em março de 2025.

⁷ Disponível em https://www.youtube.com/watch?v=T_d3R0WVSs4. Acesso em junho de 2025.

⁸ As três últimas plataformas não foram analisadas posteriormente, pois, a Voiser não liberou o uso da ferramenta gratuita em algumas tentativas, a BitFractal é paga e não disponibiliza versão gratuita e a Amazon Polly não oferece recurso para dublagem, apenas text to speech (TTS).

Para a análise piloto de nosso estudo de caso, foi analisado um vídeo que consiste em um breve tutorial de uso de um software de linguística de corpus, produzido pela profa. Silvana Maria de Jesus, orientadora dessa pesquisa, denominado “Uso corpus paralelo alinhado, parte 1” A linguagem empregada no vídeo é a linguagem natural, com exceção de alguns termos da área de corpus e noções básicas de informática. O vídeo é um passo a passo para uso do software AntConc utilizado em análises textuais na Linguística de Corpus e foi produzido em português brasileiro. As explicações são acompanhadas pela demonstração simultânea das tarefas executadas ao longo do vídeo.

Nessa análise piloto foram utilizadas as ferramentas Downsub, Amazon Polly e ElevenLabs. Nessa exploração inicial, observou-se duas possibilidades de produção de vídeos dublados automaticamente: i) texto para áudio (do inglês, *text to speech* – TTS) e ii) dublagem feita por IA (do inglês, *AI dubbing*). No primeiro caso (TTS), extrai-se a legenda do vídeo fonte, que então é traduzida para a língua alvo e, posteriormente, converte-se a legenda na língua alvo para áudio e então, sincroniza-se a legenda finalizando o vídeo alvo. No segundo caso, a ferramenta vale-se da tecnologia da IA para traduzir diretamente o áudio do vídeo fonte em áudio na língua alvo. Segue uma breve descrição de cada um destes métodos.

Método 1 – texto para áudio (TTS)

Para a produção do vídeo em inglês via TTS, foram utilizados os seguintes procedimentos:

- Extração da legenda automática do vídeo, originalmente feito em língua portuguesa e postado no Youtube, utilizando-se a ferramenta Downsub;
- Tradução da legenda em português para a língua inglesa, utilizando-se a ferramenta Google Tradutor;
- Conversão da legenda em inglês para o áudio em inglês, utilizando-se a ferramenta Amazon Polly;
- Revisão e finalização do áudio em inglês, feita manualmente pelo pesquisador;
- Substituição e sincronização do novo áudio no vídeo original, feita com a ferramenta CapCut;
- Resultado final do procedimento TTS: vídeo dublado em inglês.

Método 2 – dublagem feita por IA (AI dubbing)

Para a produção do vídeo em inglês via AI dubbing, foram utilizados os seguintes procedimentos:

- Baixar o vídeo fonte do Youtube;
- Inserir o vídeo na ferramenta ElevenLabs;
- Resultado final do procedimento AI dubbing: vídeo dublado em inglês.

Para se chegar a esses dois métodos foram necessários várias leituras e investigação de ferramentas disponíveis, bem como vídeos instrucionais do YouTube e das ferramentas sobre os procedimentos. A análise piloto mostrou-se positiva, especialmente no que diz respeito à análise da qualidade dos vídeos produzidos na língua alvo, nos dois métodos, e do papel da intervenção humana no processo.

CAPÍTULO 3 – ANÁLISE DE DADOS

Como foi dito, foram desenvolvidas quatro etapas para o estudo da dublagem automática: i) análise e descrição das plataformas a partir das informações dos sites; ii) produção da dublagem automática em cada ferramenta, gerando 9 vídeos com duração entre 45 segundos e 3 minutos; os vídeos serão disponibilizados para a banca da defesa de monografia via link do Drive; iii) análise de cada vídeo produzido, visando identificar as características adequadas e as falhas; e iv) resumo dos dados encontrados.

Descrição das plataformas de dublagem automática

Fez-se um levantamento inicial das plataformas e das ferramentas de IA oferecidas online. As plataformas foram escolhidas a partir dos principais resultados classificados no *website* Google para buscas como “ferramenta de dublagem IA”, “plataforma de dublagem automática” e semelhantes, em consultas realizadas no início deste estudo.

Os principais recursos encontrados foram: utilização da ferramenta online ou download do programa para PC ou aplicativo para celular; disponibilidade em português e no Brasil; tratamento de vídeos curtos; resolução do vídeo final; legendagem; clonagem de voz com timbre, pausas, ritmo e emoção semelhantes ao falante fonte; variedade de vozes conforme número de falantes (multi-falante); sincronização labial; edição do áudio final (via texto), download do vídeo final; colocação de marca d'água ou “marca auditiva”.

A proposta foi testar a dublagem automática nessas plataformas e descrever suas possibilidades e limitações. As dez plataformas analisadas foram CAMB.AI, CapCut, Deepdub.ai, Dubverse.ai, ElevenLabs, HeyGen, Kapwing, RASK.AI, Vidnoz, Vozo. O Quadro 1⁹ resume e compara as principais características das plataformas analisadas.

⁹ A título de curiosidade, planos da plataforma HeyGen custam a partir de U\$24 mensais (cerca de R\$130 em agosto de 2025).

Quadro 1 – Resumo das características das 10 plataformas

<u>Plataforma</u>	<u>Principais Recursos de Dublagem (STS) (Plano Gratuito)</u>	<u>Idiomas Suportados</u>	<u>Limitações do Plano Gratuito (informações retiradas dos próprios sites e fontes similares)</u>	<u>Disponível no Brasil?</u>
CAMB.AI	Dublagem, Sincronização labial, Legendagem, Edição de legenda e dublagem, Multi-falante, Clonagem de voz original	+140 idiomas	3 minutos de dublagem gratuita por mês; por 5 min é pago. Plano gratuito insere marca d'água.	Sim
CapCut	Dublagem, Sincronização labial	+20 idiomas	TTS limitado em alguns idiomas, qualidade de saída inferior (bites inferiores). Plano gratuito insere marca d'água.	Não
Deepdub.ai	Dublagem, Sincronização labial, Edição de dublagem, Multi-falante, Clonagem de voz original	+130 idiomas	Combina IA com revisão humana para alta qualidade na versão paga. Plano gratuito insere marca d'água.	Sim
Dubverse.ai	Dublagem	5 idiomas (plano gratuito) +60 idiomas	50 créditos iniciais (<i>speech minutes</i>), qualidade de dublagem variável. Plano gratuito insere marca d'água.	Sim
ElevenLabs	Dublagem, Sincronização labial, Clonagem de voz original (em partes)	+29 idiomas (dublagem), +32 idiomas (outras ferramentas)	Foco em qualidade de voz. Plano gratuito insere marca d'água.	Sim
HeyGen	Dublagem, Sincronização labial, Reedição de dublagem, Clonagem de voz original	+30 idiomas	3 vídeos/mês, 3 min/vídeo, 720p, processamento padrão, 1 avatar/voz clonada, limite de 2 falantes, requer internet. Plano gratuito insere marca d'água.	Sim

Kapwing	Dublagem, Sincronização labial, Legendagem, Clonagem de voz original	+60 idiomas	Vídeos até 4 min, 720p, 3 projetos/3 dias armazenamento. Plano gratuito insere marca d'água.	Sim
RASK.AI	Dublagem, Sincronização labial, Edição de legenda e dublagem, Clonagem de voz original	+130 idiomas	Sincronização labial é um recurso pago. Plano gratuito insere marca d'água.	Sim
Vidnoz	Dublagem, Sincronização labial, Legendagem Edição de legenda e dublagem, Clonagem de voz original	+140 idiomas	1 min/dia de tradução. Plano gratuito insere marca d'água.	Sim
Vozo	Dublagem, Sincronização labial, Legendagem, Edição de legenda e dublagem, Clonagem de voz original	+31 idiomas	30 pontos (<i>speech minutes</i> , que expiram em 7 dias), 3 a 6 min de dublagem. Plano gratuito insere marca d'água	Sim

Em geral, as plataformas pedem um breve cadastro para uso e limitam o tempo de dublagem do vídeo, variando entre cerca de 40 segundos e 3 minutos. Algumas plataformas oferecem pontos, os *speech minutes*, que podem ser trocados por serviços; cada ponto corresponde a um minuto dublado; para obter os *speech minutes* é necessário fazer o cadastro, voltar em outro dia, etc. A seguir, destaca-se brevemente alguns pontos de cada uma das dez plataformas analisadas, bem como aspectos gerais das ferramentas disponíveis na versão gratuita.

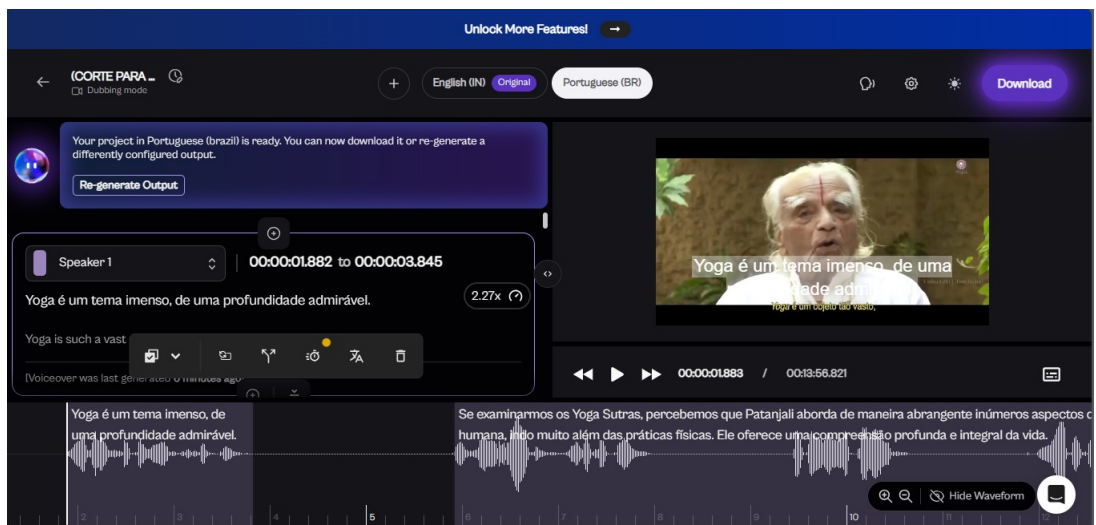
CAMB.AI

Após solicitação de cadastro, a plataforma CAMB.AI prontamente abre a página de edição. Apenas os 3 primeiros minutos do vídeo podem ser dublados. Diferentemente das outras plataformas gratuitas, a CAMB.AI disponibiliza

configurações avançadas, em sua maior parte gratuitamente. A sincronização labial está habilitada, com opção de habilitar o “silenciamento” dos sons de fundo.

Também há a opção de realizar a dublagem para português “por cima” do áudio original em inglês, como um *voice-over*; essa opção não foi testada. A opção de transcrição do áudio requer *speech minutes* para ser utilizada. Há também, na versão gratuita, a opção de identificar os falantes no vídeo e inserir música de fundo.

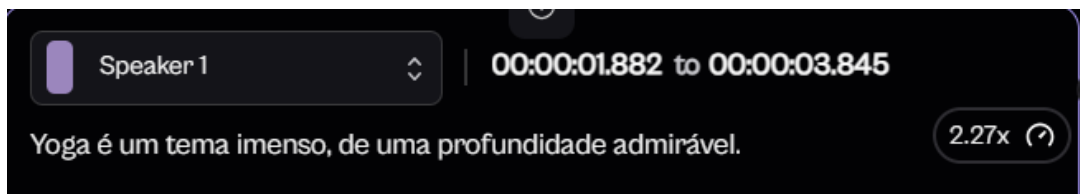
Figura 1 – Interface da plataforma CAMB.AI



Fonte: O autor (2025).

Destacou-se na dublagem feita pela CAMB.AI a aceleração no primeiro trecho do vídeo. A ferramenta informa que é possível editar essa velocidade ao alongar o tempo ou o número de caracteres. Isso pode ser feito sem gastar *speech points*. Nota-se que a dublagem foi acelerada para acomodar a sincronização labial automática.

Figura 2 – Recurso de velocidade e alongamento de dublagem



Fonte: O autor (2025).

Quando a faixa de texto acompanha adequadamente o intervalo de fala, a sincronização labial é feita de forma satisfatória. A voz utilizada na dublagem parece natural e coerente com a voz do vídeo original. Uma vantagem da plataforma CAMB.AI é o sistema de edição robusto e gratuito, apesar do limite de 3 minutos gratuitos (com utilização de *speech minutes*).

CapCut

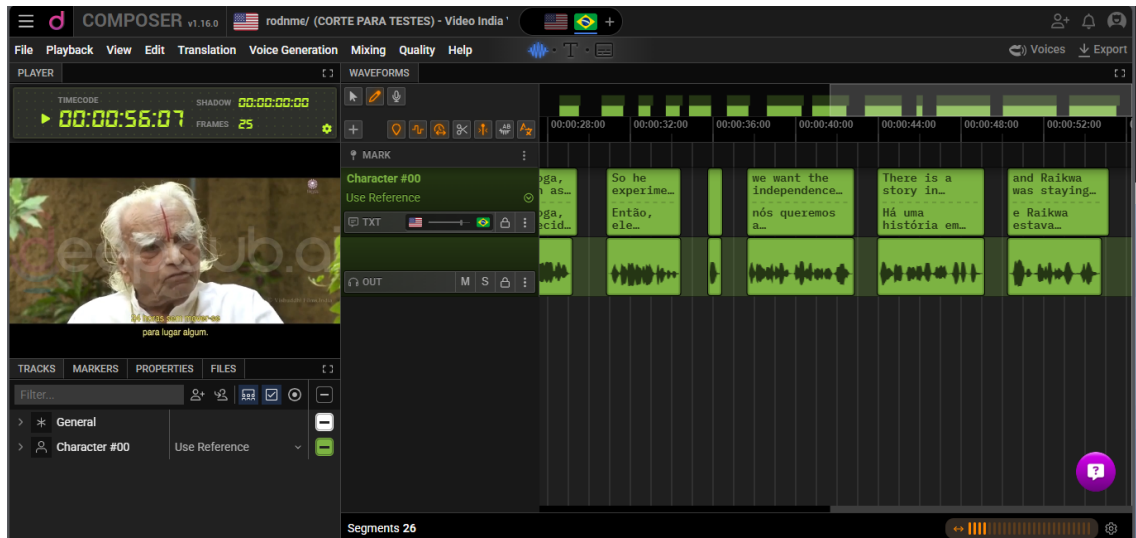
A plataforma oferece diversas ferramentas de edição de texto. No entanto, a ferramenta de dublagem automática ainda não está disponível no Brasil, apesar de já haver propaganda disponível (em inglês) para a nossa região. O site informa que a ferramenta de dublagem oferece ferramenta de edição online, opção de baixar o programa da ferramenta no computador ou em aplicativo de celular. Entretanto, não foi possível testar a dublagem do vídeo, pois a plataforma informa que não está disponível para nossa região.

Deepdub.ai

A plataforma Deepdub.ai disponibiliza a mesma quantidade de dublagens para plano gratuito e Premium, entretanto, só é possível fazer download do vídeo dublado com a marca d'água com o nome da plataforma. Antes de acessar os recursos, a plataforma demanda registro de usuário.

Observou-se que a qualidade e naturalidade da dublagem variam bastante ao longo do vídeo. Também apresenta marca d'água para o vídeo gratuito. O ponto que se destaca é como o editor da plataforma é robusto, como várias opções de edição do resultado. É, realmente, a união entre IA e revisão humana, a fim de alcançar o melhor resultado possível, mesmo na versão gratuita.

Figura 3 – Interface da plataforma Deepdub.ai



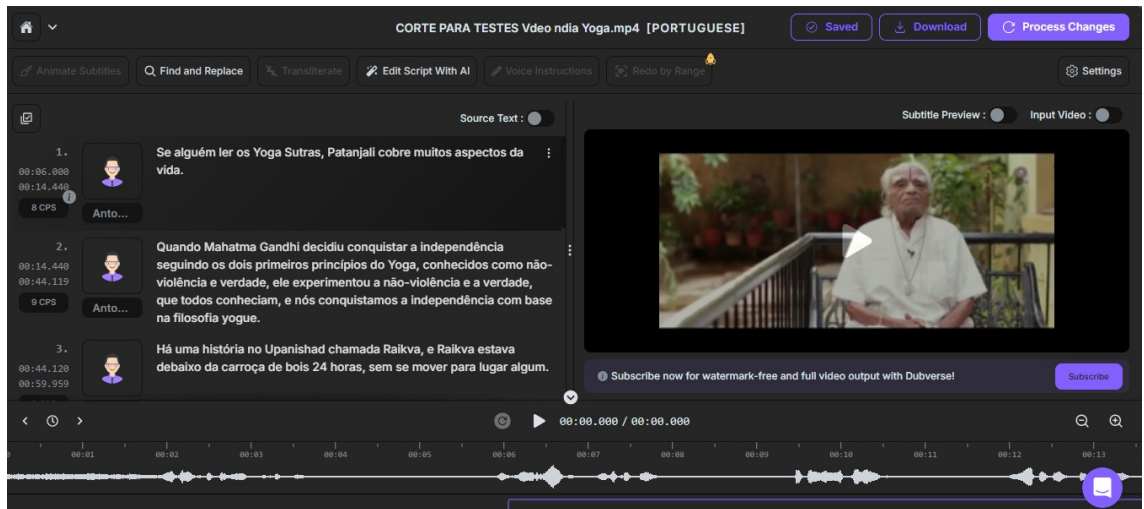
Fonte: O autor (2025).

A ferramenta da Deepdub.ai permite, de forma gratuita, alongar o período de falas do locutor do vídeo e oferece o recurso de redublar o vídeo automaticamente, além de possibilitar a edição manual de legendas.

Dubverse.ai

O processo começa com o cadastro inicial, depois, permite a inserção do vídeo que é cortado automaticamente. Também utiliza o sistema de pontos e solicita a língua alvo. Essa plataforma, além de apresentar a marca d'água visual no início do vídeo, traz ainda uma “marca auditiva”, ou seja, um som que pronuncia o nome da plataforma. Essa “marca auditiva” pôde ser facilmente removida, conforme teste realizado na ferramenta, sem necessidade de utilizar os pontos. Não é possível editar os áudios finais. A conta gratuita tem uma limitação de tempo, bloqueando-se automaticamente, entretanto, pode ser reativada mediante pagamento. A adição de áudios como música de fundo ou similares, também pode ser feita.

Figura 4 – Interface da plataforma Dubverse.ai



Fonte: O autor (2025).

Essa plataforma apresenta algumas limitações como voz robótica. Não há sincronização labial. Suas pausas são confusas e não acompanham o áudio fonte. Não é possível alterar manualmente a minutagem da dublagem feita; parece possível escrever um script para que a IA da plataforma faça correções.

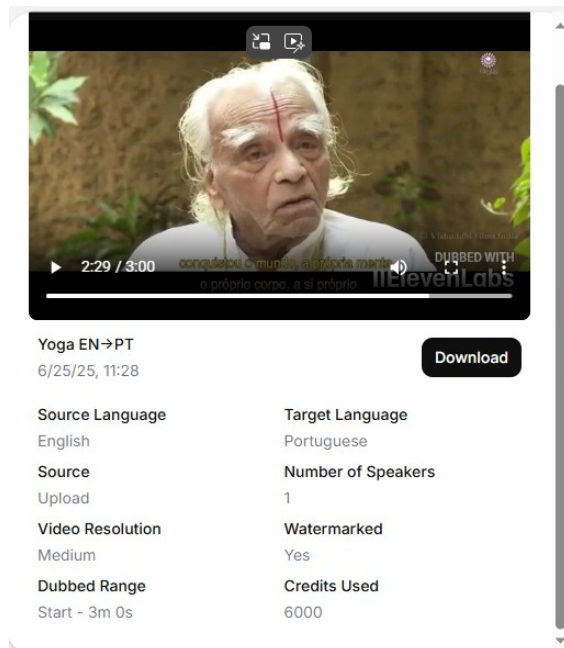
ElevenLabs

A ElevenLabs é uma das plataformas que foram testadas ainda na fase piloto da pesquisa. Naquela oportunidade, foi utilizado um vídeo que não exigiu a realização de dublagem com sincronização labial, pois o vídeo não mostrava o rosto do falante.

A ferramenta tem *layout* intuitivo, leve e fácil de usar. Após registro, é solicitada a inserção de vídeo-teste e as línguas fonte e alvo. A plataforma também solicita o número de falantes no vídeo para garantir melhor qualidade de dublagem. Além disso, a plataforma adverte que é necessário optar pela quantidade de tempo de vídeo a ser dublado, com limite máximo de 5 minutos (quanto mais minutos, mais *speech minutes* são utilizados). Há também um aviso informando que, na versão gratuita, o vídeo gerado terá qualidade padrão e marca d'água.

O vídeo produzido pela ElevenLabs tem boa sincronização labial. Mas não é possível editar o vídeo, seja para alterar o texto ou espaçamento de falas. Também não fica claro se a versão paga permite qualquer forma de edição.

Figura 5 – Visualização de vídeo na plataforma ElevenLabs



Fonte: O autor (2025).

É curioso observar que outras plataformas mencionam utilizar a tecnologia da ElevenLabs. Entretanto, apesar de fornecer tecnologia de IA para outras plataformas, não é fácil localizar (ou saber se existe) um editor do vídeo final na plataforma ElevenLabs, característica comum nas outras analisadas.

HeyGen

Após cadastro, a plataforma apresenta em sua página principal diversas ferramentas de IA. Dentre elas, a ferramenta de dublagem, descrita como “*Translate Video*”, que disponibiliza opções para dublagem simples e avançada. A opção avançada apresenta recursos como remoção do som de fundo original do vídeo, glossário, e opções *premium*, como disponibilização de legendas e sotaque desejado para a dublagem. Um recurso exclusivo da plataforma HeyGen, o qual já estava previamente selecionado e não foi encontrado em nenhuma das outras ferramentas analisadas, é a edição dinâmica (*dynamic duration*). Essa função edita automaticamente trechos do vídeo para adequar a duração da imagem ao áudio dublado pela ferramenta, resultando em um vídeo final mais curto. O plano gratuito alega oferecer dublagens de até 3 minutos.

Figura 6 – Painel de opções da plataforma HeyGen

Video Translation ⓘ

Your Free plan supports videos up to 3 minutes in length.

Source language

🗣️ Auto Detect ▼

📎

🔄 Add Transcription (SRT/ASS) 🟡

Number of speakers

Auto Detect ▼

Target language *

🗣️ Choose language ▼

🟡 Upgrade to access additional accents

Brand glossary

None ▼

Advanced options

☒ Allow dynamic duration ⓘ

☐ Translate audio only (no lip sync)

☐ Remove background music

☐ Output to original video specs ⓘ 🟡

☐ Enable captions 🟡

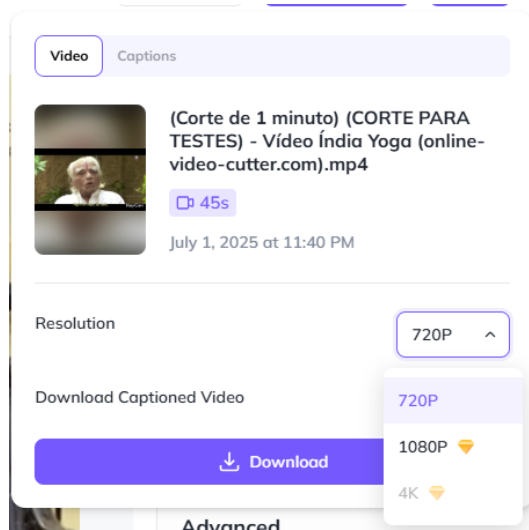
☐ Enhance voice ⓘ

Fonte: O autor (2025).

A seção de “opções avançadas” apresentada na figura anterior demonstra como a ferramenta oferece opções para refinar o processo de dublagem. Conforme dito, a opção de edição dinâmica já vem pré-selecionada. Logo abaixo, o usuário pode escolher traduzir apenas o áudio, sem sincronização labial (*Translate audio only (no lip sync)*), ou remover a música de fundo (*Remove background music*), com função de isolar melhor o diálogo do vídeo. A outra funcionalidade apresentada como gratuita é a de melhoria na qualidade da voz (*Enhance voice*). As duas outras finalidades, indicadas como Premium, permitem fazer download do vídeo com suas especificações de qualidade originais (*Output to original video specs*) e habilitar legendas (*Enable captions*).

Na dublagem dessa ferramenta, a marca d'água aparece apenas nos primeiros segundos de vídeo. A dublagem tem voz natural e similar à do áudio original, além de sincronização labial. A ferramenta de “edição dinâmica” sugerida pela plataforma acabou reduzindo a duração do vídeo original em 10 segundos. Esse corte propõe deixar o vídeo mais curto, visto que o original tem a fala mais cadenciada. A ferramenta oferece download gratuito em 720p e em qualidades superiores nos planos pagos.

Figura 7 – Painel de *download* de vídeo da plataforma HeyGen



Fonte: O autor (2025).

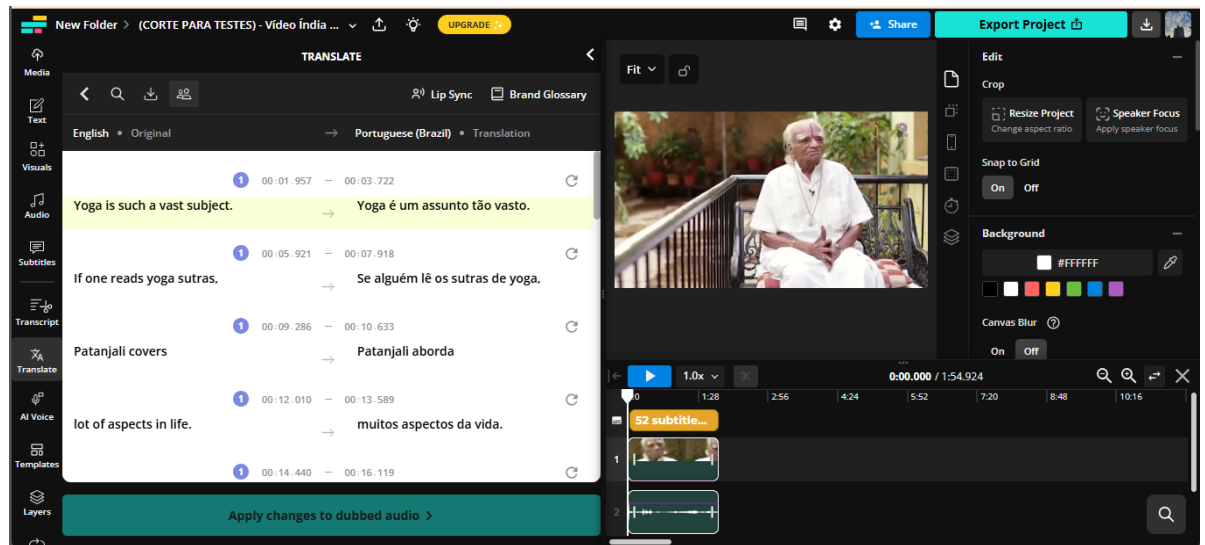
Kapwing

A plataforma Kapwing se mostra bem acessível desde o primeiro instante; insere-se o vídeo a ser dublado e em poucos instantes o processo é inicializado e então um cadastro é solicitado. Há uma opção para revisar a transcrição do vídeo fonte antes da dublagem ser feita, mas essa opção é paga.

Ao solicitar a dublagem automática, o site informa sobre o limite de 4 minutos e recomenda o corte do vídeo para realizar o processo gratuitamente. Entretanto, a plataforma deu acesso a pouco menos de 2 minutos de dublagem gratuita. A plataforma pede para informar a língua fonte e a língua alvo da dublagem automática.

A Figura 8 mostra a tela com o texto fonte e o texto alvo alinhados ao lado esquerdo, e com a marcação dos minutos; nesse alinhamento, basta clicar no botão refazer (na tela branca, seta circular a direita do segmento de texto) para que a ferramenta refaça um segmento da dublagem, o que é feito automaticamente, sem a intervenção do usuário; pode-se assistir ao vídeo dublado no canto direito.

Figura 8 – Interface da plataforma Kapwing



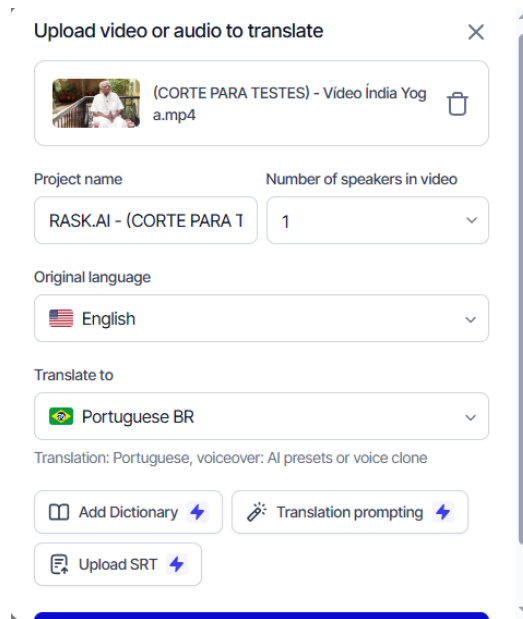
Fonte: O autor (2025).

O resultado da dublagem mostrou-se positivo em termos de estilização e naturalidade da voz, além de versatilidade durante as pausas do vídeo original. Além disso, ao baixar o vídeo dublado, percebe-se que para utilizar o recurso de sincronização labial é preciso fazer uso dos *speech minutes* e que o vídeo final leva uma marca d'água. Para fazer quaisquer alterações na dublagem, é necessário também usar os *speech minutes*. O uso da plataforma Kapwing é simples e o resultado adequado.

RASK.AI

A RASK.AI pede as mesmas informações que a maior parte das plataformas anteriores, mas fornece algumas opções (pagas) diferentes, como adicionar glossário e adicionar sugestões para a máquina realizar a dublagem automática.

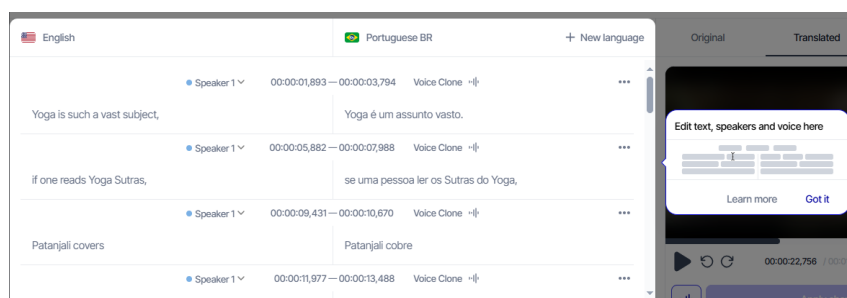
Figura 9 – Painel de opções da plataforma RASK.AI



Fonte: O autor (2025).

No primeiro momento, a plataforma não limitou o vídeo enviado, que tem duração de 14:29 minutos. Entretanto, apenas 1 minuto do vídeo foi dublado. A opção de editar o vídeo é ilustrada na Figura 10, com a transcrição do áudio fonte, em inglês, à esquerda, e a transcrição do áudio alvo, em português, à direita.

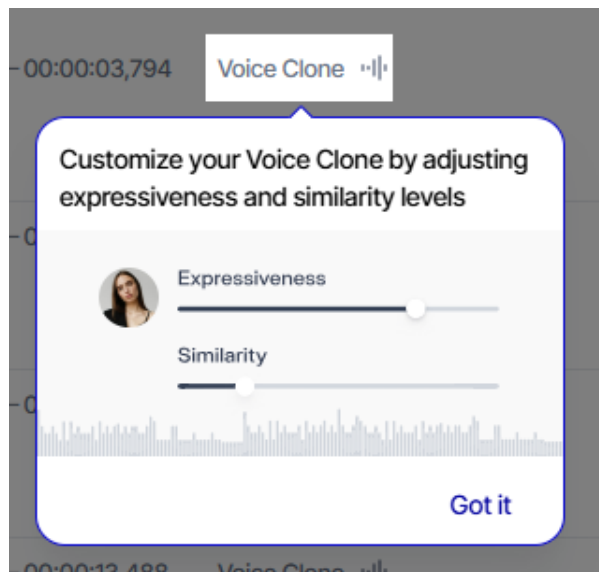
Figura 10 – Interface de edição de dublagem da plataforma RASK.AI



Fonte: O autor (2025).

Um recurso pouco explorado por outras plataformas é o de clonagem de voz. Na RASK.AI, mesmo na versão gratuita, é possível ajustar a expressividade e até mesmo a similaridade da voz dublada com a original.

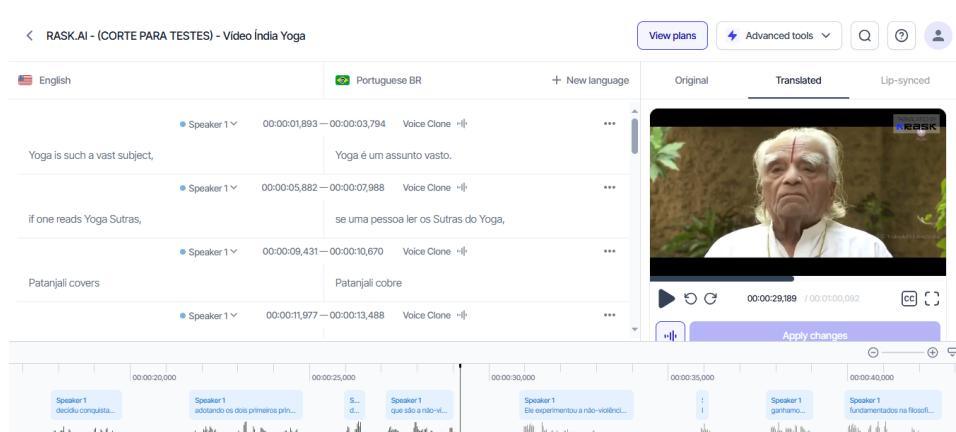
Figura 11 – Painel de clonagem de voz na plataforma RASK.AI



Fonte: O autor (2025).

Apesar de não apresentar uma sincronização labial tão natural, oferece várias opções de edição. A versão gratuita insere marca d'água no vídeo.

Figura 12 – Interface da plataforma RASK.AI



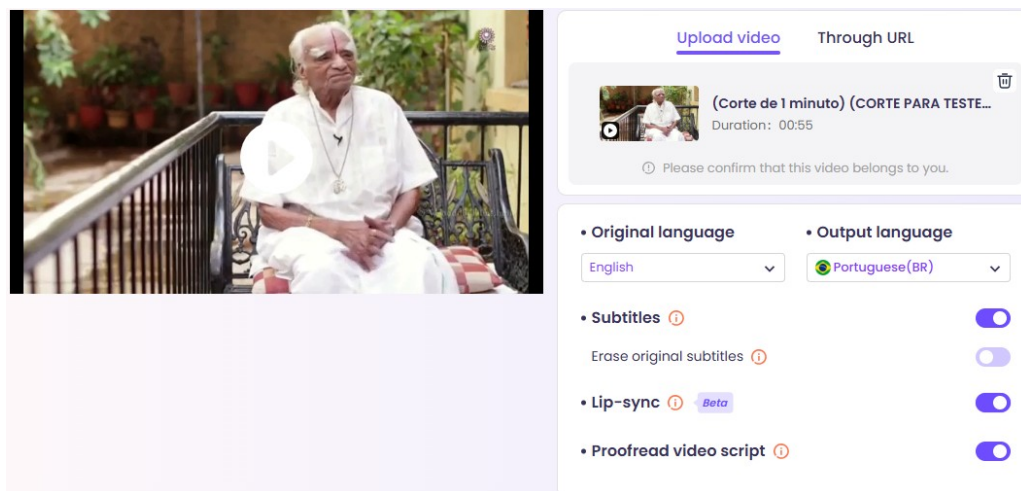
Fonte: O autor (2025).

O recurso de sincronização labial é pago. A ferramenta oferece resultados e recursos interessantes, como edição manual do texto dublado.

Vidnoz

Diferentemente de todas as outras plataformas, a Vidnoz não solicita registro antes de permitir o acesso à ferramenta de dublagem; oferece recursos como legendagem e sincronização labial, em versão *beta*, que será comentado a seguir.

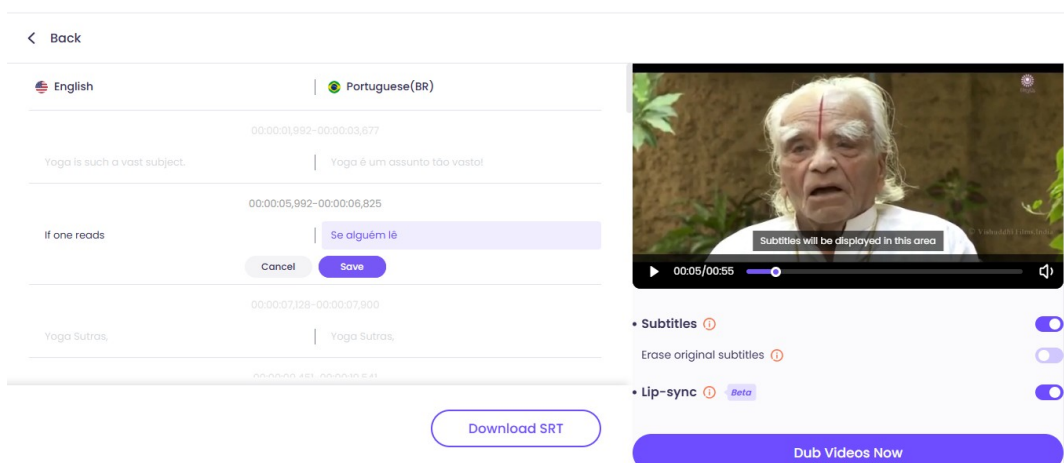
Figura 13 – Painel de opções da plataforma Vidnoz



Fonte: O autor (2025).

Além disso, também há o recurso de revisão da transcrição da dublagem. É possível editar gratuitamente a transcrição da dublagem antes da finalização do vídeo.

Figura 14 – Interface de edição de dublagem da plataforma Vidnoz



Fonte: O autor (2025).

O diferencial dessa ferramenta é a sincronização labial em versão beta, com o qual a Vidnoz ensaia uma sincronização labial com animação facial (Jayachandran, 2025). Isso é uma tentativa de fazer com que o falante do vídeo faça movimentos labiais coerentes com a dublagem. Entretanto, como a ferramenta alerta, o recurso é inovador e evidentemente necessita melhorias, já que a edição do vídeo é perceptível.

Figura 15 – Sincronização labial com animação facial da plataforma Vidnoz



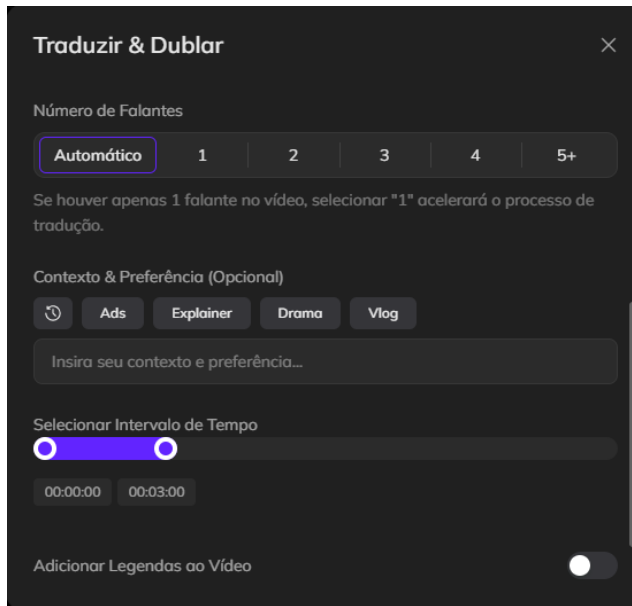
Fonte: O autor (2025).

Para baixar o vídeo dublado, é necessário então realizar o cadastro na plataforma.

Vozo

Após realizado o registro de usuário, a plataforma oferece uma quantidade de *speech minutes* para uso. Com os *speech minutes* é possível dublar um vídeo de até 3 minutos. Além de disponibilizar recursos de dublagem e legendagem, a ferramenta oferece recursos como identificar multi-falantes.

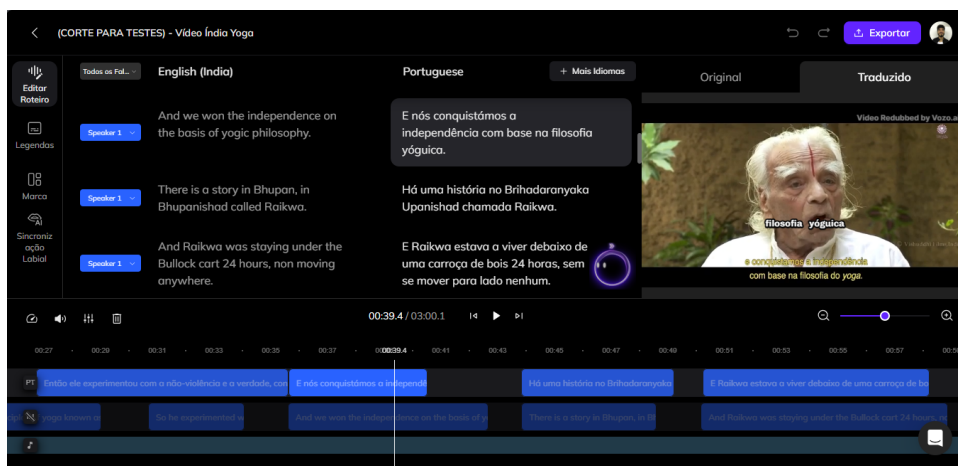
Figura 16 – Recurso de identificação de multi-falantes da Voz



Fonte: O autor (2025).

A plataforma oferece mais recursos na versão gratuita, como edição da dublagem e mudança de fonte da legenda.

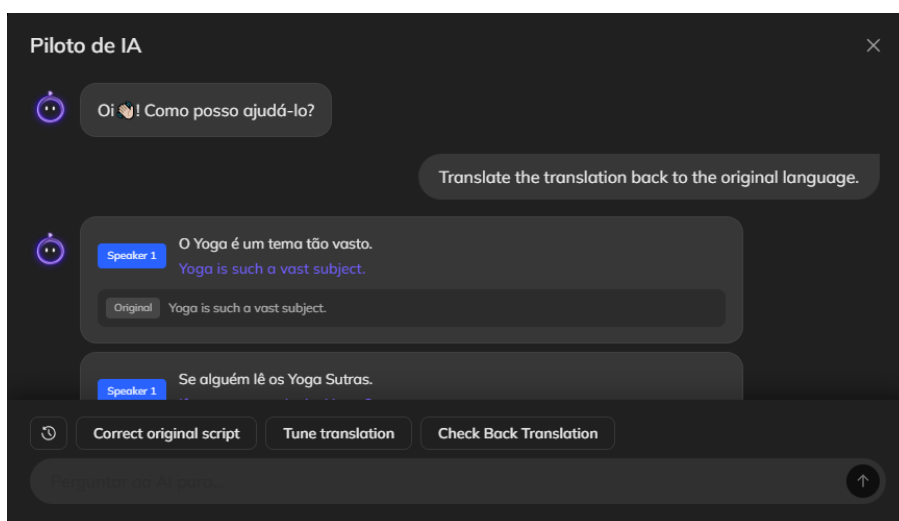
Figura 17 – Interface da plataforma Voz



Fonte: O autor (2025).

Outro recurso é um *chat bot* dentro da própria ferramenta, que pode auxiliar na edição ou correção da dublagem. Isso evidencia como o trabalho automatizado com curadoria humana pode ser uma opção acessível, eficiente e necessária.

Figura 18 – Recurso de *chat bot* da plataforma Vozo



Fonte: O autor (2025).

Após a análise das plataformas listadas, observa-se que os serviços gratuitos podem apresentar boa qualidade, mas esbarram em limitações dos recursos fornecidos, principalmente em relação ao tempo de vídeo que disponibilizam gratuitamente. Os serviços pagos prometem recursos superiores, mas seria necessário assinar a plataforma para testar esses recursos, o que foge do escopo desta pesquisa. Por exemplo, no serviço pago, o usuário tem maior liberdade para editar o resultado, se necessário.

Produção e análise dos 9 vídeos produzidos pelas plataformas

Após essa primeira observação das dez plataformas, realizou-se a segunda etapa, ou seja, produzir o vídeo dublado automaticamente em cada plataforma (exceto a que não está disponível para o Brasil) e analisar a qualidade do vídeo (características e falhas). Importa ressaltar as características do áudio original, que tem a voz de um indiano falando inglês, um vídeo de 14 minutos sobre a filosofia da yoga; a fala é lenta, com sotaque e oscilações na fala (há um trecho em que ele gagueja um pouco, por exemplo, o que afetou a leitura da ferramenta, em alguns casos). O estudo focou nas características e falhas de cada plataforma, com ênfase em clonagem de voz, sincronização labial e qualidade da tradução do sentido do áudio fonte. Utilizou-se a legenda produzida pela Associação Brasileira de Iyengar Yoga (ABIY) como parâmetro de análise dos áudios, ou seja, considerou-se a

legenda como a tradução adequada do sentido do áudio fonte, visto que a legenda foi produzida e publicada por especialistas da área.

1-Camb.AI

Características – clonagem da voz; sentido adequado, gerou 3m de vídeo; marca d'água com o nome da plataforma no cantor inferior direito.

Falhas – velocidade acelerada em alguns trechos gerou uma dessincronização labial com a legenda em alguns momentos (ouve-se o áudio, mas o falante não mexe a boca); manteve o vídeo completo, mas depois do minuto 3 fica mudo; parece uma tradução mais livre-interpretativa do que a das outras plataformas.

2- CapCut (não disponível no Brasil)

3- Deepdub.ai

Características – clonagem da voz; gerou 56s¹⁰; marca d'água com o nome da plataforma centralizada no vídeo.

Falhas – erros de sentido - gerou erro em um trecho em que o som do original está ruim (*we won* é traduzido como *we want*). O sotaque (às vezes parece português europeu) e o timbre da voz, principalmente, variam bastante ao longo do vídeo, que inclui cortes abruptos no áudio; sincronização labial pouco natural.

4- Dubverse.ai

Características – gerou 2m de vídeo; marca d'água com o nome da plataforma no cantor superior esquerdo.

Falhas – voz mais jovem e mais rápida gerando longos silêncios; não há sincronização labial; erros de sentido (ex. então vamos me aproximar dele);

5- ElevenLabs

Características – sentido adequado; sincronização labial; gerou 3m de vídeo, marca d'água com o nome da plataforma no cantor inferior direito.

Falhas – a voz é aproximada, mas não é uma clonagem e ocorrem algumas variações e oscilações; em alguns trechos há um *delay* e a voz fica diferente (ex.

¹⁰ Nessa ferramenta não foi possível subir o vídeo completo; portanto, o pesquisador fez um recorte de 1 minuto.

todo o meu... reino – 1m22s); no tempo 1m30 tem um som estranho no trecho "por que não asanas" parece uma voz diferente (tempo 2m40); diz "filosofia yogica".

6- HeyGen

Características – gerou 45s de vídeo; clonagem da voz; sentido adequado; sincronização labial com animação facial.

Falhas – há edição dinâmica automática, que gera cortes e pausas inadequadas (entre 11s e 16s), mas é possível desabilitar o recurso.

7- Kapwing

Características – clonagem da voz; sincronização labial; gera legenda automaticamente; gerou 1m54s de vídeo.

Falhas – erros de sentido (ex. o meu --- / seria reino, mas é omitido, 1m16s); pausas, cortes e desaceleração de dublagem causam dessincronização em alguns momentos.

8- Rask.AI

Características – gerou 1m de vídeo; clonagem da voz; sentido adequado; sincronização labial.

Falhas – a velocidade da voz varia um pouco.

9- Vidnoz

Características – gerou 55s de vídeo; há opção para gerar legenda, sincronização labial com animação facial.

Falhas – erros de sentido; voz mais jovem e com várias pausas inadequadas.

10- Vozo

Características – gerou 3m; clonagem da voz; gera legenda automaticamente, que pode ser retirada em edição posteriormente.

Falhas – erros de sentido (ex. *which everybody knew* foi traduzido como “conceitos que envolvem todos em Nova Iorque”, em 34s; *half my Kingdom* foi traduzido como “metade do meu forno”, em 1m19s); português europeu; uma pausa longa de silêncio; sincronização labial inconstante.

Figura 19 – Exemplo de erro gerado na legenda e no áudio



Fonte: O autor (2025).

Antes da análise dos vídeos, cabe ressaltar uma explicação sobre o processo de sincronização labial. Um dos pilares da dublagem tradicional é a busca pela sincronia, baseada em um conjunto de técnicas que buscam a “harmonia” entre o movimento labial e o som (áudio traduzido) de um produto audiovisual. Chaume (2006) classifica a sincronização em três tipos:

Sincronia fonética ou labial: consiste em adaptar a fala (do texto traduzido) aos movimentos de articulação labial de personagens na tela, especialmente em cenas conhecidas como “*closes*”. Segundo Chaume (p. 3), “A tradução deve respeitar particularmente as vogais abertas e as consoantes bilabiais e labiodentais pronunciadas na tela”, para que a tradução simule encaixar-se naturalmente na boca de quem fala.

Isocronia: se refere à sincronização de tempo. O diálogo traduzido deve ter exatamente a mesma duração do diálogo original, respeitando o tempo exato em que a personagem abre sua boca, do início ao fim de sua fala, incluindo pausas. Chaume afirma que falhas de isocronia são frequentemente percebidas pelo público de um produto audiovisual como um sinal de “dublagem mal feita”.

Sincronia Cinética: é a adequação da tradução e da dublagem aos movimentos corporais e gestos da personagem na tela. Por exemplo, “um aceno de cabeça afirmativo não pode ser acompanhado por uma fala negativa” (p.3).

Dessa forma, a sincronização tradicional busca ajustar o texto oral traduzido à imagem facial e corporal do vídeo original. Entretanto, as plataformas e ferramentas de dublagem por IA introduziram novos elementos no processo de sincronização. Plataformas como Vozo, Rask.ai e HeyGen utilizam o termo para descrever uma tecnologia que altera e manipula os movimentos labiais contidos em um vídeo, para que esses correspondam à dublagem gerada na língua-alvo, ou seja, há uma edição de imagem. Logo, há uma ressignificação do termo para incluir novas possibilidades técnicas.

Figura 20 – Excerto sobre sincronização labial da plataforma Vozo



Fonte: O autor (2025).

A plataforma Vidnoz, por exemplo, é uma das únicas plataformas entre as analisadas que oferece essa funcionalidade em sua versão gratuita. Conforme descrito, a ferramenta tenta fazer com que o falante execute movimentos labiais coerentes com a dublagem produzida, o que evidencia uma edição de imagem realizada durante o processo automático. Ou seja, ao invés de adaptar a tradução à imagem, a ferramenta de IA adapta a imagem à dublagem produzida.

Essa opção por ampliar terminologias pode ser também uma estratégia comercial, visto que os termos *sincronização labial* e *lip-sync* já são consagrados e facilmente reconhecidos pelo público em geral. Ao mesmo tempo, as plataformas pouco buscam diferenciar esse processo, que na TAV tradicional, seria mais

próximo de uma “sincronização de áudio dublado”, possivelmente para tornar seu produto mais facilmente identificável. Essa redefinição terminológica é algo a ser explorado no campo de estudos da Tradução Audiovisual na era da IA.

Assim, observa-se que, embora a maioria das plataformas seja capaz de clonar a voz do locutor fonte e preservar o sentido da fala, até imitando o “gaguejo” do locutor, uma série de problemas foram identificados.

Há a presença de marca d’água na maioria dos vídeos, variando a disposição e podendo haver uma marca auditiva. A imitação de sotaque e do timbre da voz fonte apresentou inconsistências, resultando em variações perceptíveis nos vídeos. A sincronização labial, embora presente em algumas das ferramentas, demonstrou ser inconsistente e pouco precisa. Também houve problemas com a tradução, tanto por erros de sentido quanto por traduções literais.

Um recurso de ajuste automático da velocidade da fala, a fim de sincronizá-la com o vídeo original, frequentemente resulta em aceleração ou desaceleração abrupta da voz, levando a dessincronização entre áudio e movimentos labiais. Em certos casos, esse processo gerou pausas longas e silêncios indevidos. Além disso, algumas ferramentas incluíram automaticamente a edição dinâmica (no original *dynamic duration*, na plataforma HeyGen), que editou partes do vídeo para “adequar” a dublagem e o tempo, resultando em um vídeo de menor duração. Por fim, algumas ferramentas podem gerar legendas automaticamente, enquanto em outras há opção para incluir ou não a legenda. Entretanto, é necessário avaliar a qualidade das legendas, o que demandaria uma outra pesquisa.

Resumo da análise dos dados

As plataformas de dublagem automática analisadas demonstram características adequadas, como a tradução do sentido da fala, e, em alguns casos, até mesmo a capacidade de replicar nuances como o gaguejo. A clonagem de voz, quando bem-sucedida, e a tentativa de sincronização labial são pontos fortes, mas frequentemente acompanhadas de falhas.

Por outro lado, as principais falhas encontradas foram recorrentes. Entre elas, destaca-se a inconsistência na clonagem de sotaque e timbre, resultando em vozes que variam ao longo do vídeo ou que não são clonadas, distanciando-se da voz fonte. A sincronização labial, mesmo quando executada, foi pouco precisa em

diversas ferramentas. Ainda, a qualidade da dublagem foi comprometida por erros de sentido e traduções literais, em alguns casos. Um ponto interessante é a edição dinâmica automática e/ou o recurso manual de alteração da velocidade da fala. Entretanto, quando esses recursos foram utilizados, apresentaram resultados inadequados.

Do ponto de vista do usuário, é fácil manusear as ferramentas, pois apesar de serem bem diferentes entre si, são baseadas em editores de vídeo tradicionais. O processo de cadastro foi sempre bem semelhante, ainda que cada plataforma tenha suas próprias métricas em relação a utilização de recursos gratuitos e de *speech minutes*. Foi possível observar que algumas ferramentas, como a Vidnoz, excluem os vídeos dublados por contar com espaço de nuvem limitado a certa quantidade de tempo. Já a Dubverse.ai “tranca” o vídeo, e só permite acessá-lo novamente mediante assinatura. A ElevenLabs, por sua vez, não permite qualquer tipo de edição após a conclusão da dublagem.

Por fim, parece que a limitação mais significativa foi a de que o usuário da plataforma possui pouca ou nenhuma autonomia para editar a maneira na qual as características são aplicadas, o que acaba por comprometer o resultado dos vídeos. Essa falta de controle sobre fatores essenciais ao sucesso do processo demonstra como a intervenção humana faz falta durante a produção de dublagens automáticas. Considerando-se que a tecnologia de dublagem automática com IA é nova, que as plataformas estão em constantes aprimorações e que os testes foram com as versões gratuitas, optou-se por não classificar qual foi o melhor resultado, o que será melhor investigado em pesquisas futuras.

CONCLUSÕES

Este estudo buscou investigar o campo da dublagem automática, identificando plataformas, ferramentas e seus recursos, e, finalmente os desafios que protagonizam esse processo na era da IA. As perguntas que nortearam a investigação visam analisar quais recursos são utilizados, a relevância da IA durante a realização das dublagens, a possibilidade de uma produção adequada sem interferência humana e quais as principais características e falhas das plataformas disponíveis no mercado.

A análise das ferramentas gratuitas revelou que, embora existam avanços significativos na tecnologia de clonagem de voz e na tentativa de sincronização labial, a qualidade final dos produtos ainda é inconsistente. Falhas na clonagem de voz, erros de tradução e problemas de sincronização de dublagem foram recorrentes, demonstrando que a automação, por si só, não garante um resultado adequado.

A limitação na capacidade de intervenção humana no processo de dublagem automática dessas ferramentas se mostra como um fator importante para os resultados apresentados. Limitações em relação ao controle do estilo da clonagem de voz e a impossibilidade de edição de sincronia labial, a fim de promover maior precisão, demarcam as barreiras que o processo promove. A falta de recursos para corrigir erros de dublagem resulta em um vídeo que, na maior parte das vezes, se mostra aquém da qualidade que as ferramentas descrevem dispor.

A natureza automatizada das ferramentas dessas plataformas gratuitas, embora garanta agilidade no processo, também impõe restrições que acabam comprometendo a qualidade da dublagem, o que evidencia como a intervenção humana é essencial para um resultado adequado. Dessa forma, conclui-se que a intervenção humana permanece indispensável para corrigir as limitações dos sistemas automatizados e garantir a qualidade final do produto audiovisual.

É importante reforçar que se trata de um estudo de caráter exploratório, focado em versões gratuitas das tecnologias de dublagem automática. Diante das limitações encontradas, sugere-se que, em futuras pesquisas, uma investigação aprofundada sobre os recursos oferecidos pelas versões pagas dessas plataformas. Tal análise permitirá verificar se as ferramentas *premium* superam os desafios identificados e oferecem maior controle e autonomia ao usuário, contribuindo para

uma compreensão mais completa do futuro da TAV e, portanto, da dublagem automática.

REFERÊNCIA

ARAÚJO, V. L. S. **Ser ou não ser natural, eis a questão dos clichês de emoção na tradução audiovisual**. Tese (Doutorado em Língua Inglesa e Literaturas Inglesa e Norte-Americana) – Faculdade de Filosofia Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2000. Disponível em: <https://doi.org/10.11606/T.8.2000.tde-16012023-171916>. Acesso em: mar. 2025.

ASSOCIAÇÃO BRASILEIRA DE IYENGAR YOGA. **Yogacharya BKS Iyengar's Unique Revelations LEGENDADO**. Youtube, São Paulo, 2023. Disponível em: https://www.youtube.com/watch?v=T_d3R0WVSs4. Acesso em: jun. 2025.

AZEVEDO, C. A.; MELLO, H. R.; CÔRTEZ, P. O. **Texto ou fala?**. In: CASELI, H. M.; NUNES, M. G. V. (Org.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap 2. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao>. (Grupo Brasileiras em PLN <https://brasileiraspln.com>). Acesso em: jun. 2025.

BOLACIO FILHO, E. et al. **Tradução automática, pós-edição e o pronome indefinido "man"**. *PandaemoniumGer.*, São Paulo, v. 28, 2025, e240020. Disponível em: <https://revistas.usp.br/pg/article/view/230552/211543>. Acesso em: abr. 2025.

CARVALHO, C. A. **A tradução para legendas: dos polissistemas à singularidade do tradutor**. Dissertação (Programa de PósGraduação em Letras do Departamento de Letras do Centro de Teologia e Ciências Humanas da PUC-Rio). *Cadernos Maxwell*, Rio de Janeiro, 2005, e6613. Disponível em: <https://doi.org/10.17771/PUCRio.acad.6613>. Acesso em: mar. 2025.

CHAUME, F.; DOTEI, M. D. **A história da tradução audiovisual**. *Cadernos de Tradução*, Florianópolis, v. 44, n.1, 2024. p. 2. Disponível em: <https://doi.org/10.5007/2175-7968.2024.e99043>. Acesso em: mar. 2025.

CHAUME, F. Screen Translation: Dubbing. In **Encyclopedia of Language & Linguistics**, 2nd ed. Disponível em: [DOI:10.1016/B0-08-044854-2/00471-5](https://doi.org/10.1016/B0-08-044854-2/00471-5). Acesso em: ago. 2025.

DÍAZ-CINTAS, J. Introduction: Audiovisual translation comes of age. In BUCARIA, C.; HEISS, C.; CHIARO, D. (Ed.) **Between Text and Image: Updating research in**

screen translation. Josh Benjamins Publishing Company, Reino Unido, 2008. p. 1-7. Disponível em: DOI:[10.1075/btl.78.02dia](https://doi.org/10.1075/btl.78.02dia). Acesso em: mar. 2025.

FEDERICO, M. et al. From Speech-to-Speech Translation to Automatic Dubbing. In **Proceedings of the 17th International Conference on Spoken Language Translation**, pages 257–264, Online. 2020. Association for Computational Linguistics. Disponível em: DOI:[10.18653/v1/2020.iwslt-1.31](https://doi.org/10.18653/v1/2020.iwslt-1.31). Acesso em: jun. 2025.

FRANCO, E. P. C.; ARAÚJO, V. L. S. **Questões terminológico-conceituais no campo da Tradução Audiovisual (TAV).** Artigo. Tradução em Revista, Rio de Janeiro, 2011. Disponível em: DOI:[10.17771/PUCRio.TradRev.18884](https://doi.org/10.17771/PUCRio.TradRev.18884). Acesso em: mar. 2025.

JAYACHANDRAN, A. et. al. **Lip Dubbing Tools Suite: Enhancing Video Dubbing with AI-Driven Lip Syncing and Facial Restoration.** International Journal of Science, Architecture, Technology and Environment, 2025. Disponível em: <https://doi.org/10.63680/ijstate0525161.32>. Acesso em: ago. 2025.

JESUS, S. M. **Uso corpus paralelo alinhado, parte 1.** YouTube, Uberlândia, 2020. Disponível em: <https://www.youtube.com/watch?v=Um8vZ5fp00I>. Acesso em fev. 2025.

MACHADO, Dilma. **O Processo da Tradução para a Dublagem Brasileira: teoria e prática.** Rio de Janeiro: Altiya Gráfica, 2016.

PERES, T. **Legendagem tradicional e seus parâmetros técnicos: uma estratégia formal de tradução audiovisual.** Dissertação (Programa de Pós-graduação em Estudos da Tradução. Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2023. Disponível em: DOI:[10.13140/RG.2.2.36155.39202](https://doi.org/10.13140/RG.2.2.36155.39202). Acesso em abr. 2025.

TUDOCELULAR. **Amazon Prime Video aposta em dublagem por IA para inglês e espanhol em 2025.** TudoCelular, São Paulo, 2025. Disponível em: <https://www.tudocelular.com/curiosidade/noticias/n232450/amazon-prime-video-dublagem-por-ia.html>. Acesso em: abr. 2025.