



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO SERTÃO PERNAMBUCANO CAMPUS SALGUEIRO
CURSO DE TECNOLOGIA EM SISTEMAS PARA INTERNET**

DANIEL ANTONIO DA CONCEIÇÃO SILVA

**FERRAMENTAS DE SÍNTESE DE VOZ POR IA PARA COMUNICAÇÃO
AUMENTATIVA E ALTERNATIVA: ESTUDO COMPARATIVO NO CONTEXTO
DO PORTUGUÊS BRASILEIRO**

**SALGUEIRO - PE
2026**

DANIEL ANTONIO DA CONCEIÇÃO SILVA

**FERRAMENTAS DE SÍNTESE DE VOZ POR IA PARA COMUNICAÇÃO
AUMENTATIVA E ALTERNATIVA: ESTUDO COMPARATIVO NO CONTEXTO
DO PORTUGUÊS BRASILEIRO**

O Trabalho de Conclusão de Curso (TCC) foi apresentado na disciplina de TCC do curso de Tecnologia em Sistemas para Internet, ofertado pelo Campus Salgueiro do Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano (IFSertãoPE), como requisito final para obtenção do título de Tecnóloga em Sistemas para Internet.

Orientador: Prof. Marcelo Anderson Batista dos Santos.

S586 Silva, Daniel Antônio da Conceição.

SÍNTESE DE VOZ BASEADA EM INTELIGÊNCIA ARTIFICIAL PARA
COMUNICAÇÃO AUMENTATIVA E ALTERNATIVA: : UMA ANÁLISE
COMPARATIVA DE FERRAMENTAS PARA O PORTUGUÊS BRASILEIRO / Daniel
Antônio da Conceição Silva. - Salgueiro, 2026.
18 f.

Trabalho de Conclusão de Curso (Sistemas para Internet) -Instituto Federal de Educação,
Ciência e Tecnologia do Sertão Pernambucano, Campus Salgueiro, 2026.
Orientação: Prof. Dr. Marcelo Anderson Batista dos Santos.

1. Inteligência artificial. 2. Síntese de voz. 3. Comunicação aumentativa e alternativa. 4.
Paralisia cerebral. I. Título.

CDD 006.3

Ferramentas de Síntese de Voz por IA para Comunicação Aumentativa e Alternativa: Estudo Comparativo no Contexto do Português Brasileiro

Daniel Antônio, Marcelo Santos

Coordenação de Informática - Instituto Federal do Sertão Pernambucano (IFSertãoPE),
campus Salgueiro

BR-232, Km 508, s/n - Zona Rural, 56000-000 – Salgueiro - PE – Brasil

{marcelo.santos, daniel.antonio}@ifsertaope.edu.br

RESUMO

A comunicação constitui direito humano fundamental, e para pessoas com paralisia cerebral que apresentam comprometimento severo da fala, a Comunicação Aumentativa e Alternativa (CAA) representa o principal meio de expressão. Este estudo realizou uma análise comparativa de cinco sistemas de síntese de voz baseados em inteligência artificial com suporte ao português brasileiro: ElevenLabs, deAPI.ai, Google Cloud TTS, Coqui TTS e Piper TTS. A pesquisa adotou metodologia qualitativa com elementos quantitativos, combinando análise documental baseada em critérios técnicos predefinidos e testes práticos de inteligibilidade utilizando reconhecimento automático de fala. Como diferencial metodológico, o estudo incorporou a perspectiva vivencial do autor, usuário de CAA com paralisia cerebral que utiliza rastreamento ocular para comunicação. Os resultados indicaram que ferramentas comerciais como ElevenLabs oferecem maior naturalidade e recursos de personalização, enquanto alternativas de código aberto como Piper TTS e Coqui TTS apresentam vantagens significativas em privacidade e uso offline. A pesquisa conclui que não existe solução universalmente superior, sendo a escolha dependente do contexto específico de uso, recursos disponíveis e necessidades individuais do usuário.

Palavras-chave: Síntese de voz. Inteligência artificial. Comunicação Aumentativa e Alternativa. Paralisia cerebral. Tecnologia assistiva.

ABSTRACT

Communication is a fundamental human right, and for individuals with cerebral palsy who have severe speech impairments, Augmentative and Alternative Communication (AAC) represents the primary means of expression. This study conducted a comparative analysis of five artificial intelligence-based speech synthesis systems supporting Brazilian Portuguese: ElevenLabs, deAPI.ai, Google Cloud TTS, Coqui TTS, and Piper TTS. The research adopted

a qualitative methodology with quantitative elements, combining documentary analysis based on predefined technical criteria and practical intelligibility tests using automatic speech recognition. As a methodological differential, the study incorporated the lived experience perspective of the author, an AAC user with cerebral palsy who uses eye-tracking for communication. Results indicated that commercial tools such as ElevenLabs offer greater naturalness and customization features, while open-source alternatives like Piper TTS and Coqui TTS present significant advantages in privacy and offline use. The research concludes that no universally superior solution exists, with the choice depending on specific usage context, available resources, and individual user needs.

Keywords: *Speech synthesis. Artificial intelligence. Augmentative and Alternative Communication. Cerebral palsy. Assistive technology.*

1 INTRODUÇÃO

A comunicação constitui um dos pilares fundamentais da existência humana, sendo reconhecida pela Declaração Universal dos Direitos Humanos (ONU, 1948) como direito inalienável de todo ser humano. Para pessoas com paralisia cerebral que apresentam comprometimento motor severo afetando a produção da fala, a Comunicação Aumentativa e Alternativa representa não apenas uma ferramenta tecnológica, mas o principal meio pelo qual podem expressar pensamentos, necessidades, emoções e participar ativamente da vida em sociedade.

O avanço das tecnologias assistivas tem proporcionado importantes contribuições para a CAA, ampliando as possibilidades de comunicação e interação para pessoas com paralisia cerebral, além disso os recursos, serviços e estratégias usadas podem colaborar com a acessibilidade, com o processo de aprendizagem e com o desenvolvimento das habilidades dos estudantes com deficiências. Às vezes essas tecnologias têm baixo custo e podem variar desde dispositivos de baixa tecnologia, até sistemas altamente sofisticados, que recorrem a computadores e dispositivos eletrônicos. A escolha da tecnologia mais adequada depende das necessidades individuais de cada estudante com deficiência, garantindo maior independência e participação ativa na sociedade.

Historicamente, os sistemas de CAA evoluíram de pranchas de comunicação com símbolos pictográficos até sofisticados dispositivos eletrônicos com síntese de voz. Contudo, durante décadas, as vozes sintéticas disponíveis caracterizavam-se por qualidade robótica e

artificial, criando uma barreira adicional na comunicação. O advento das tecnologias de inteligência artificial, especialmente as redes neurais profundas, revolucionou este cenário, possibilitando a geração de vozes sintéticas com naturalidade sem precedentes.

A voz, para além de sua função comunicativa instrumental, constitui elemento fundamental da identidade pessoal. Para usuários de CAA, a voz sintética torna-se literalmente sua voz, representando-os em todas as interações sociais (MILLS et al., 2014). Neste contexto, a qualidade, naturalidade e possibilidade de personalização da síntese de voz adquirem importância que transcende aspectos meramente técnicos, impactando diretamente a autoestima, a percepção social e a qualidade de vida dessas pessoas (WÜLFING et al., 2020; JUDGE; TOWNEND, 2013).

Apesar dos avanços tecnológicos significativos, observa-se uma lacuna na literatura científica quanto a análises comparativas sistemáticas de ferramentas de síntese de voz com foco específico no português brasileiro e nas necessidades de usuários de CAA. A maioria dos estudos concentra-se em idiomas de maior expressão internacional, especialmente o inglês, deixando falantes de português brasileiro com informações limitadas para embasar suas escolhas tecnológicas.

Este estudo diferencia-se das pesquisas existentes não apenas pelo foco no português brasileiro, mas principalmente pela incorporação da perspectiva vivencial do pesquisador. Como pessoa com paralisia cerebral que utiliza rastreamento ocular e sistemas de CAA para comunicação cotidiana, o autor traz uma compreensão única das necessidades, desafios e aspirações dos usuários dessas tecnologias, combinando rigor metodológico científico com a experiência de quem vive a realidade investigada.

O objetivo geral desta pesquisa foi realizar uma análise comparativa de ferramentas de síntese de voz baseadas em inteligência artificial com suporte ao português brasileiro, avaliando sua adequação para uso em contextos de Comunicação Aumentativa e Alternativa. Como objetivos específicos, buscou-se: mapear o estado atual das tecnologias de síntese de voz neural disponíveis para português brasileiro; estabelecer critérios de avaliação relevantes para o contexto de CAA; realizar análise sistemática das ferramentas selecionadas; conduzir testes práticos de inteligibilidade; e documentar a experiência vivencial de uso dessas tecnologias.

Diante desse cenário, coloca-se a seguinte questão de pesquisa: quais ferramentas de síntese de voz baseadas em inteligência artificial apresentam melhor adequação técnica e funcional para uso em Comunicação Aumentativa e Alternativa no contexto do português brasileiro? Este estudo busca responder a essa questão por meio de análise comparativa de soluções disponíveis, considerando aspectos técnicos, funcionais e a experiência prática de uso.

2 REFERENCIAL TEÓRICO

2.1 Paralisia Cerebral e o Direito à Comunicação

A paralisia cerebral constitui um grupo de desordens permanentes do desenvolvimento do movimento e da postura, causando limitações funcionais atribuídas a distúrbios não progressivos ocorridos no cérebro fetal ou infantil em desenvolvimento. Trata-se da causa mais comum de deficiência física na infância, com prevalência estimada entre 2 e 3 casos por 1.000 nascidos vivos em países desenvolvidos.

Entre as diversas manifestações da paralisia cerebral, o comprometimento da comunicação oral representa um dos aspectos mais impactantes na qualidade de vida. Estudos indicam que aproximadamente 50% das pessoas com paralisia cerebral apresentam algum grau de dificuldade na fala, variando de comprometimentos leves até a impossibilidade completa de produção verbal inteligível.

É fundamental destacar que a dificuldade ou impossibilidade de falar não implica, de forma alguma, déficit cognitivo ou incapacidade de compreensão. Este equívoco, ainda prevalente na sociedade, constitui uma das principais barreiras enfrentadas por pessoas com paralisia cerebral, que frequentemente têm suas capacidades intelectuais subestimadas devido à aparência física ou à dificuldade de expressão verbal.

O direito à comunicação está consagrado em diversos instrumentos legais internacionais e nacionais. No Brasil, a Lei Brasileira de Inclusão (BRASIL, 2015) estabelece que a pessoa com deficiência tem direito à comunicação por meio de recursos de tecnologia assistiva que ampliem suas habilidades funcionais (BRASIL, 2015; COMITÊ DE AJUDAS TÉCNICAS, 2007). Este direito fundamenta a importância do acesso a tecnologias de CAA de qualidade.

2.2 Comunicação Aumentativa e Alternativa

A Comunicação Aumentativa e Alternativa compreende um conjunto de estratégias, técnicas e recursos destinados a complementar, suplementar ou substituir a comunicação oral de pessoas que apresentam limitações nesta modalidade (ASHA, 2024; NUNES; WALTER, 2016). O termo "aumentativa" refere-se a sistemas que complementam a fala natural, enquanto "alternativa" designa sistemas que substituem integralmente a comunicação oral.

Os recursos de CAA podem ser classificados em sistemas de baixa tecnologia, como pranchas de comunicação impressas com símbolos pictográficos, e sistemas de alta tecnologia, que incluem dispositivos eletrônicos dedicados e softwares especializados com síntese de voz. A evolução tecnológica tem permitido o desenvolvimento de sistemas cada vez mais sofisticados, incluindo dispositivos com rastreamento ocular que possibilitam a comunicação mesmo para pessoas com mobilidade extremamente reduzida. Tanto um recurso como o outro, devem ser cuidadosamente avaliados com o propósito de atender a necessidade do estudante com deficiência, as vezes é necessário haver adaptações, pois a eficácia da CAA não depende apenas dos recursos utilizados, mas sim da sua adaptação às necessidades individuais de cada utilizador, considerando as suas capacidades cognitivas, motoras e sensoriais.

No contexto brasileiro, destaca-se o trabalho pioneiro de pesquisadores como Nunes e Walter, que sistematizaram o uso de sistemas pictográficos e desenvolveram metodologias de implementação da CAA em contextos educacionais e clínicos. Entretanto, o acesso a tecnologias de CAA de alta qualidade ainda constitui desafio significativo, especialmente devido aos custos elevados de equipamentos importados e à limitada oferta de vozes sintéticas em português brasileiro com qualidade adequada.

2.3 Evolução da Síntese de Voz

A síntese de voz, também conhecida como Text-to-Speech, passou por três grandes fases evolutivas (TAN et al., 2021). A primeira geração baseava-se em síntese concatenativa, que consistia na junção de segmentos de áudio pré-gravados para formar palavras e frases. Embora funcional, este método produzia resultados com qualidade limitada e transições frequentemente artificiais entre segmentos.

A segunda geração introduziu a síntese paramétrica, utilizando modelos estatísticos como Hidden Markov Models para gerar parâmetros acústicos a partir do texto. Esta

abordagem ofereceu maior flexibilidade, permitindo ajustes de prosódia e velocidade, mas ainda produzia vozes com características robóticas perceptíveis.

A revolução atual corresponde à terceira geração, baseada em redes neurais profundas. Marcos importantes incluem o WaveNet, desenvolvido pela DeepMind (VAN DEN OORD et al., 2016), que introduziu redes convolucionais para geração de áudio amostra por amostra, alcançando qualidade sem precedentes. Posteriormente, arquiteturas como Tacotron e seus sucessores possibilitaram o aprendizado automático da relação entre texto e fala, enquanto modelos baseados em Transformers, como FastSpeech, permitiram síntese em tempo real com controle prosódico refinado.

Os avanços mais recentes incluem modelos de difusão e sistemas de clonagem de voz que permitem criar vozes personalizadas a partir de poucos minutos de áudio de referência (PATEL et al., 2018). Esta capacidade abre possibilidades revolucionárias para usuários de CAA, que podem potencialmente ter uma voz única que os represente, em vez de compartilhar vozes genéricas com milhares de outros usuários.

2.4 A Voz como Elemento de Identidade

A voz humana constitui um dos principais marcadores de identidade pessoal, comunicando não apenas conteúdo verbal, mas também características como gênero, idade aproximada, origem regional, estado emocional e traços de personalidade. Para pessoas que utilizam síntese de voz como meio primário de comunicação, a voz sintética torna-se, de fato, sua voz, representando-os em todas as interações sociais (MILLS et al., 2014).

Pesquisas na área de CAA têm demonstrado que a qualidade e adequação da voz sintética impactam significativamente a receptividade dos interlocutores e a autoestima dos usuários. Vozes excessivamente robóticas ou inadequadas ao perfil do usuário podem criar barreiras adicionais na comunicação, enquanto vozes naturais e personalizadas facilitam interações mais fluidas e autênticas.

Iniciativas como o projeto VocaliD, desenvolvido por Rupal Patel, pioneiramente abordaram a questão da personalização vocal para usuários de CAA, desenvolvendo técnicas para criar vozes únicas combinando características prosódicas do usuário com propriedades acústicas de doadores de voz. Mais recentemente, serviços comerciais como ElevenLabs têm democratizado o acesso à clonagem de voz, oferecendo a possibilidade de criar vozes personalizadas a partir de pequenas amostras de áudio.

Para pessoas que nunca tiveram a capacidade de falar, como é o caso de muitos indivíduos com paralisia cerebral congênita, a questão da identidade vocal assume contornos particulares. Não se trata de recuperar uma voz perdida, mas de criar uma identidade vocal que represente quem a pessoa é ou deseja ser. Esta perspectiva ressalta a importância de ferramentas que permitam não apenas vozes de alta qualidade, mas também personalização e expressividade adequadas.

3 METODOLOGIA

3.1 Caracterização da Pesquisa

Esta pesquisa caracteriza-se como aplicada quanto à sua natureza, exploratória quanto aos objetivos e de abordagem qualitativa, com apoio de indicadores quantitativos básicos. O estudo apresenta delineamento comparativo de caráter técnico, combinando análise documental, testes práticos de inteligibilidade e relato de experiência do autor, usuário de Comunicação Aumentativa e Alternativa.

3.2 Seleção das Ferramentas

A seleção das ferramentas para análise baseou-se nos seguintes critérios: disponibilidade de suporte ao português brasileiro; relevância no mercado ou comunidade de código aberto; potencial de aplicação em contextos de CAA; e disponibilidade de documentação técnica adequada. Foram selecionadas cinco ferramentas que representam o espectro atual de opções disponíveis.

As ferramentas comerciais incluem o ElevenLabs, reconhecido pela alta qualidade de síntese e recursos avançados de clonagem de voz (ELEVENLABS, 2025); o deAPI.ai, plataforma que disponibiliza uma API unificada para acesso a modelos de síntese de voz baseados em inteligência artificial, permitindo integração simplificada em aplicações e sistemas de acessibilidade (DEAPI.AI, 2026); e o Google Cloud TTS, que oferece diversas vozes neurais e integração com o ecossistema Google (GOOGLE CLOUD, 2025). As ferramentas de código aberto compreendem o Coqui TTS, que disponibiliza modelos treináveis e clonagem de voz via biblioteca Python (COQUI TTS, 2025), e o Piper TTS (PIPER TTS, 2025), otimizado para execução local em dispositivos com recursos limitados.

3.3 Critérios de Avaliação

Os critérios de avaliação foram definidos considerando aspectos técnicos e práticos relevantes para o contexto de CAA. O primeiro critério, Suporte a Português Brasileiro, avalia a disponibilidade e qualidade das vozes em português do Brasil. O segundo critério, Recursos de Personalização, examina as possibilidades de customização da voz, incluindo clonagem, ajuste de estilos e controle prosódico. O terceiro critério, Integração e Compatibilidade, analisa a facilidade de integração com softwares de CAA e sistemas operacionais. O quarto critério, Uso Offline, verifica a possibilidade de operação sem conexão à internet, aspecto crítico para dispositivos de comunicação que devem funcionar em qualquer contexto. O quinto critério, Documentação e Suporte, avalia a qualidade da documentação disponível e canais de suporte. O sexto critério, Planos Gratuitos, examina a disponibilidade e limitações de versões gratuitas. O sétimo critério, Modelo de Custo, analisa a estrutura de preços para uso além do tier gratuito. O oitavo critério, Uso Comercial, verifica as permissões de licenciamento para uso em produtos. O nono critério, Privacidade dos Dados, avalia as políticas de tratamento dos dados enviados para síntese.

3.4 Análise Documental

A análise documental foi conduzida através de pesquisa sistemática na documentação oficial de cada ferramenta, incluindo sites oficiais, documentação técnica de APIs, termos de serviço e políticas de privacidade. Complementarmente, foram consultados repositórios GitHub para as ferramentas de código aberto, avaliações em plataformas especializadas e publicações técnicas relevantes.

As informações coletadas foram organizadas em matriz comparativa, permitindo análise sistemática de cada critério para todas as ferramentas. Esta análise foi realizada entre janeiro e fevereiro de 2026, refletindo o estado das ferramentas neste período.

3.5 Testes Práticos de Inteligibilidade

Os testes práticos de inteligibilidade foram conduzidos utilizando um corpus de dez frases representativas de situações comunicativas típicas em CAA. As frases foram selecionadas para abranger diferentes contextos de uso, incluindo necessidades básicas, interações sociais e expressões emocionais.

O procedimento de teste consistiu em sintetizar cada frase utilizando as ferramentas disponíveis para teste prático, gravar os áudios resultantes e submetê-los a reconhecimento automático de fala utilizando modelos de linguagem como Gemini ou ChatGPT. A taxa de inteligibilidade foi calculada como a proporção de frases corretamente reconhecidas pelo sistema de transcrição automática.

Este método, embora não substitua avaliações com ouvintes humanos, oferece uma medida objetiva e replicável da clareza da síntese de voz, sendo particularmente útil para comparações sistemáticas entre ferramentas.

Ressalta-se que o procedimento adotado possui caráter exploratório e não substitui avaliações perceptivas com ouvintes humanos. Entretanto, o uso de sistemas de reconhecimento automático de fala permite uma medida objetiva e replicável da clareza da síntese, adequada ao escopo comparativo deste estudo.

3.6 Experiência Vivencial

Esta pesquisa adota abordagem autoetnográfica, considerando que o autor é pessoa com paralisia cerebral e usuário cotidiano de sistemas de Comunicação Aumentativa e Alternativa. A experiência pessoal foi utilizada como fonte complementar de dados qualitativos, permitindo análise contextualizada do uso das tecnologias em situações reais de comunicação.

Como diferencial metodológico, esta pesquisa incorporou a perspectiva vivencial do autor, pessoa com paralisia cerebral que utiliza sistema de rastreamento ocular Tobii PCEye 5 integrado ao software Communicator 5 (TOBII DYNAVOX, 2024) para comunicação diária. Esta posição privilegiada de pesquisador-usuário permitiu avaliar aspectos que frequentemente escapam a análises puramente técnicas, incluindo a experiência subjetiva de uso, a adequação contextual das vozes e o impacto na interação com interlocutores.

A documentação da experiência vivencial seguiu abordagem sistemática, registrando observações sobre facilidade de configuração e uso, adequação das vozes para diferentes contextos comunicativos, reações de interlocutores às diferentes vozes, fadiga auditiva em uso prolongado e aspectos emocionais relacionados à identidade vocal.

4 RESULTADOS E ANÁLISE

4.1 Análise Comparativa das Ferramentas

A análise documental das cinco ferramentas selecionadas revelou um panorama diversificado de opções, cada uma com características distintivas que as tornam mais adequadas para contextos específicos de uso. O Quadro 1 apresenta a síntese comparativa completa baseada nos nove critérios estabelecidos.

Quadro 1 – Análise comparativa das ferramentas de síntese de voz

Critério	ElevenLabs	deAPLai	Google TTS	Coqui TTS	Piper
Suporte a pt-BR	Sim – Multilingual v2/v3, vozes pré-treinadas na biblioteca	Sim – depende do modelo utilizado via API	Sim – ~5 WaveNet, 3 Neural2, 24 Chirp premium	Sim – Modelos comunitários (YourTTS, CSS10)	Sim – Modelos pt-BR disponíveis (M e F)
Personalização	Clonagem instantânea (5-10s) e profissional (30min+), multilíngue, ajustes de emoção	Variável conforme modelo TTS selecionado	Custom Voice (Beta) com horas de áudio, controle SSML de rate/pitch	Total – treinar com qualquer voz, YourTTS zero-shot, parâmetros editáveis	Limitado – requer treinar novo modelo, sem interface de clonagem
Integração	API REST, bibliotecas Python, interface web Studio, sem SAPI nativo	API REST simples, integração direta em aplicações	API REST/gRPC, bibliotecas várias, sem SAPI nativo	CLI/Python, requer desenvolvimento para integrar CAA	CLI/C++/Python, módulo SAPI5 comunitário, usado em Home Assistant
Uso Offline	Não – requer nuvem, áudio pode ser baixado depois	Não – funcionamento em nuvem	Não – inteiramente cloud	Sim – totalmente offline após download	Sim – totalmente offline, otimizado para Raspberry Pi
Documentação	Boa – guias, API docs, comunidade Discord/Reddit	Boa – documentação para desenvolvedores	Boa – documentação Cloud, \$300 créditos iniciais	Média – README GitHub, comunidade em chat	Média – iniciativas comunitárias, uso simples
Plano Gratuito	10.000 chars/mês (~10min), clonagem básica	Créditos gratuitos iniciais para testes	4M chars Standard + 1M WaveNet/mês, Neural2 só com crédito	100% gratuito, sem limite (hardware local)	100% gratuito, sem custos
Custo Pago	US\$5-99/mês, clonagem profissional requer ≥US\$22	Modelo pay-as-you-go baseado em uso	US\$4-16/1M chars, Chirp US\$30/1M	N/A – apenas custos de hardware	N/A – software livre
Uso Comercial	Sim – planos pagos incluem licença comercial	Permitido conforme termos da plataforma	Sim – seguindo políticas Google Cloud	Sim – licença MIT permite redistribuição	Sim – open source, verificar licença de cada modelo
Privacidade	Servidores ElevenLabs, dados para melhorar modelo (vozes clonadas privadas)	Processamento em servidores remotos	Servidores Google, não usa para outros fins sem permissão	Local – máxima privacidade, dados não saem do dispositivo	Local – total confidencialidade

Fonte: Elaborado pelo autor (2026)

A análise do Quadro 1 revela padrões distintos entre ferramentas comerciais e de código aberto. O ElevenLabs destaca-se pela qualidade de síntese e recursos de personalização, oferecendo clonagem de voz instantânea com apenas 5-10 segundos de áudio de referência, recurso particularmente valioso para usuários de CAA que desejam uma voz única. Sua principal limitação reside na dependência de conexão com a internet e nos custos de assinatura para uso além do tier gratuito.

O deAPI.ai apresenta como principal diferencial a oferta de uma API unificada que permite acesso a diferentes modelos de síntese de voz baseados em inteligência artificial através de uma única interface de integração. Essa abordagem reduz a complexidade técnica para desenvolvedores e pesquisadores interessados em implementar síntese de voz em aplicações de acessibilidade e sistemas de Comunicação Aumentativa e Alternativa. Entretanto, por operar inteiramente em nuvem, a ferramenta depende de conexão constante com a internet e não possui integração nativa com padrões tradicionais de voz do sistema operacional, como o SAPI5 do Windows.

O Google Cloud TTS oferece ampla variedade de vozes e generoso tier gratuito para vozes Standard e WaveNet, adequado para experimentação e uso pessoal moderado. Entretanto, suas vozes mais avançadas (Neural2 e Chirp) requerem pagamento, e a ausência de suporte offline limita sua aplicabilidade em dispositivos de CAA.

Entre as opções de código aberto, o Coqui TTS destaca-se pela flexibilidade, permitindo o treinamento de modelos personalizados e oferecendo clonagem de voz zero-shot através do modelo YourTTS. Esta ferramenta é particularmente interessante para desenvolvedores e pesquisadores que desejam customização profunda, embora sua utilização requeira conhecimentos técnicos significativos.

O Piper TTS representa a opção mais acessível para uso offline, tendo sido especificamente otimizado para execução em dispositivos com recursos limitados como Raspberry Pi. Sua integração com projetos de automação residencial como Home Assistant demonstra sua maturidade e facilidade de uso. A principal limitação reside nas opções limitadas de personalização sem treinamento de novos modelos.

4.2 Testes Práticos de Inteligibilidade

Os testes práticos de inteligibilidade foram conduzidos utilizando um corpus de dez frases representativas, apresentadas no Quadro 2. As frases foram selecionadas para

representar diferentes contextos comunicativos típicos de CAA, incluindo necessidades básicas, interações sociais e expressões emocionais.

Quadro 2 – Corpus de frases para teste de inteligibilidade

Nº	Frase	Categoria
1	Estou com sede, pode me dar água?	Necessidade básica
2	Preciso ir ao banheiro agora.	Necessidade básica
3	Bom dia, como você está?	Interação social
4	Muito prazer em conhecer você.	Interação social
5	Tenho uma pergunta sobre a aula.	Contexto educacional
6	Pode repetir, por favor? Não entendi.	Interação social
7	Estou muito feliz hoje!	Expressão emocional
8	Isso me deixa triste.	Expressão emocional
9	Obrigado pela sua ajuda.	Interação social
10	O rato roeu a roupa do rei de Roma.	Desafio fonético

Fonte: Elaborado pelo autor (2026)

O Quadro 3 apresenta o modelo para registro dos resultados dos testes de inteligibilidade. A metodologia consiste em sintetizar cada frase nas ferramentas disponíveis e verificar se sistemas de reconhecimento automático de fala conseguem transcrever corretamente o áudio gerado.

Quadro 3 – Registro de testes de inteligibilidade

Ferramenta	IA	Voz	% acerto	Observações
Elevenlabs	Gemini	Rachel	100	Retornou as 10 frases em texto corrido
deAPI.ai	Gemini	Alex	100	Retornou as 10 frases como dito no áudio
Piper TTS	Gemini	Edresson	100	As frases foram transcritas corretamente de acordo com o áudio gerado
Coqui TTS	Gemini	Leonor	100	Retornou as 10 frases corretamente, nota-se que a depender da voz usada pela ferramenta o áudio sai errado
Google TTS	_____	_____	_____	Devido ao requisito de pré-pagamento para poder ativar a conta de faturamento não foi possível realizar os testes

Fonte: Elaborado pelo autor (2026)

Os resultados obtidos encontram-se apresentados no Quadro 3, indicando, para cada ferramenta testada, a taxa de acerto obtida no processo de reconhecimento automático de fala.

Observou-se taxa de inteligibilidade de 100% para as ferramentas testadas. Esse resultado indica que, no conjunto de frases analisado, todas as soluções apresentaram nível adequado de clareza para reconhecimento automático. No entanto, diferenças qualitativas foram observadas quanto à naturalidade, fluidez e expressividade das vozes, aspectos discutidos na seção seguinte.

4.3 Experiência Vivencial

O relato a seguir apresenta a trajetória do autor no uso de sistemas de CAA, com foco nos aspectos funcionais e nas implicações práticas para a comunicação cotidiana.

A experiência vivencial como usuário de CAA com paralisia cerebral proporcionou insights que complementam a análise técnica das ferramentas. Como usuário do sistema Tobii PCEye 5 integrado ao software Communicator 5 (TOBII DYNAVOX, 2024), utilizo cotidianamente as vozes SAPI5 nativas do Windows para comunicação em diversos contextos.

A seguir, apresenta-se o relato da experiência pessoal do autor com a CAA, abordando: descrição do uso diário de CAA e suas necessidades específicas; comparação subjetiva das ferramentas testadas em situações reais; reflexões sobre identidade vocal e o significado de ter uma voz própria; relatos de reações de interlocutores às diferentes vozes; e considerações sobre aspectos emocionais da comunicação mediada por síntese de voz.

Fiz uso pela primeira vez de CAA quando ingressei no Ensino Fundamental I. Eu usava uma pasta silábica que continha o alfabeto e os números impressos. Com a ajuda de um Profissional de Apoio Escolar que apontava letra por letra e eu balançava a cabeça na letra desejada. Assim era escrito a palavra, a frase ou texto. Contudo foi usado essa estratégia de comunicação até ser concluído o Ensino Fundamental II.

Com o avanço da tecnologia, no Ensino Médio passei a usar os softwares de comunicação com voz sintetizada como: Telepatix (que converte texto em fala) e o teclado Tix que pode ser usado de várias formas, além de funcionar com o toque da mão funciona com todos os tipos de acionadores externos como o acionador de piscadelas. Como tenho impedimento motor grave com ausência de fala, usei o acionador de piscadelas. O TiX e o Telepatix facilitaram muito a minha vida porque através desses recursos passei a usar todos os recursos do computador e isso me proporcionou autonomia na comunicação. Essas ferramentas são cada vez mais acessíveis.

O Toby facilitou a minha comunicação e me deu a oportunidade de acessar a internet, navegar no Google para fazer pesquisas e ter acesso ao YouTube, Instagram, Facebook e WhatsApp com total autonomia, apenas com o movimento dos olhos. Realmente o Toby é uma tecnologia que me proporcionou a superação de diversos

desafios que enfrento no dia a dia. Hoje tenho a comunicação autônoma sem depender de uma pessoa para interpretar meus gestos.

Atualmente uso a voz gerada de nome Daniel. Essa voz é caracterizada por ter ritmo. Algumas palavras são pronunciadas de forma desarticulada, por isso que prefiro escrever frases curtas. Até o presente momento essa voz não me causou nenhum transtorno, pelo contrário tem sido o meio que eu uso para apresentação de trabalhos acadêmicos, para me comunicar com amigos e familiares e fazer palestras públicas. Mesmo sem a linguagem verbal me sinto incluído na sociedade de modo geral.

A experiência relatada evidencia que, para além da inteligibilidade técnica, fatores como naturalidade, ritmo, conforto auditivo e identidade vocal desempenham papel central na escolha da tecnologia mais adequada.

5 DISCUSSÃO

A análise comparativa das cinco ferramentas de síntese de voz revelou um cenário tecnológico diversificado, onde não existe uma solução universalmente superior, mas sim opções mais adequadas para contextos específicos de uso. Esta constatação alinha-se com a literatura sobre tecnologia assistiva, que enfatiza a importância da adequação individual das soluções tecnológicas.

A dicotomia entre ferramentas comerciais e de código aberto reflete tensões fundamentais no campo da tecnologia assistiva. Por um lado, serviços como ElevenLabs oferecem qualidade de síntese e facilidade de uso superiores, mas impõem dependência de conexão à internet e custos recorrentes que podem ser proibitivos para muitos usuários brasileiros. Por outro lado, ferramentas como Piper TTS e Coqui TTS garantem autonomia, privacidade e custo zero, mas exigem maior conhecimento técnico para configuração e apresentam limitações em personalização.

A questão da integração com softwares de CAA existentes merece destaque especial. O deAPI.ai destaca-se por sua proposta de simplificação da integração de múltiplos modelos de síntese de voz através de uma única API, característica relevante para pesquisadores e desenvolvedores que desejam experimentar diferentes tecnologias sem necessidade de múltiplas implementações. Contudo, diferentemente de soluções tradicionalmente integradas ao sistema operacional, sua utilização em softwares de CAA pode exigir desenvolvimento adicional para conexão via API.

O aspecto da privacidade assume importância crítica no contexto de CAA, considerando que as mensagens sintetizadas frequentemente incluem informações pessoais e sensíveis. Neste aspecto, as ferramentas de código aberto com execução local oferecem

garantias que serviços em nuvem não podem igualar, independentemente de suas políticas de privacidade.

A personalização vocal emerge como tema central nas discussões sobre síntese de voz para CAA. A possibilidade de ter uma voz única, que represente a identidade do usuário, constitui avanço qualitativo em relação às vozes genéricas tradicionalmente disponíveis. O ElevenLabs democratizou o acesso à clonagem de voz com sua opção instantânea, enquanto o Coqui TTS oferece possibilidades similares através de seus modelos de código aberto.

As limitações deste estudo incluem o escopo restrito de ferramentas analisadas, a impossibilidade de testes longitudinais extensivos com todas as opções comerciais devido a restrições de custo, e a perspectiva individual do pesquisador que, embora valiosa, não representa a totalidade das experiências de usuários de CAA. Pesquisas futuras poderiam expandir a análise para incluir mais usuários, avaliar o impacto da qualidade vocal na participação social e investigar aspectos de longo prazo da relação entre identidade vocal e bem-estar.

Entre as limitações do estudo, destacam-se o número reduzido de frases utilizadas nos testes de inteligibilidade, a utilização de um único sistema de reconhecimento automático de fala e a análise baseada na experiência de um único usuário. Tais limitações não comprometem o caráter exploratório da pesquisa, mas indicam a necessidade de investigações futuras com amostras mais amplas e avaliações perceptivas com múltiplos participantes.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa realizou uma análise comparativa exploratória de ferramentas de síntese de voz baseadas em inteligência artificial com suporte ao português brasileiro, avaliando sua adequação para uso em contextos de Comunicação Aumentativa e Alternativa. A combinação de análise documental sistemática, testes práticos de inteligibilidade e perspectiva vivencial do pesquisador-usuário proporcionou compreensão multifacetada do tema.

Os resultados demonstram que o campo da síntese de voz neural oferece atualmente opções que representam avanço qualitativo significativo em relação às gerações anteriores de tecnologia. Ferramentas comerciais como ElevenLabs lideram em termos de naturalidade e recursos de personalização, enquanto alternativas de código aberto como Piper TTS e Coqui

TTS democratizam o acesso a síntese de qualidade com garantias de privacidade e independência de conectividade.

Para usuários de CAA no contexto brasileiro, recomenda-se considerar prioritariamente: para aqueles com recursos financeiros disponíveis e conexão estável à internet, o ElevenLabs oferece a melhor combinação de qualidade e facilidade de uso, com destaque para seu programa de impacto social em parceria com a Smartbox; para usuários que necessitam de funcionamento offline confiável ou que priorizam privacidade, o Piper TTS representa opção robusta e totalmente gratuita; para usuários do ecossistema Windows que desejam integração direta com softwares de CAA, o deAPI.ai apresenta-se como alternativa flexível baseada em API, permitindo experimentação com diferentes modelos de síntese de voz em aplicações modernas baseadas em serviços web.

A questão da identidade vocal revelou-se aspecto central que transcende considerações puramente técnicas. Para pessoas que utilizam síntese de voz como meio primário de comunicação, a voz sintética não é mero instrumento, mas elemento constitutivo de sua presença comunicativa no mundo. A crescente acessibilidade de tecnologias de personalização vocal representa, portanto, não apenas avanço tecnológico, mas conquista em termos de direitos e dignidade.

Esta pesquisa contribui para o campo ao oferecer análise sistemática orientada especificamente para o contexto de CAA em português brasileiro, combinando rigor metodológico com a perspectiva privilegiada de quem vive a realidade investigada. O princípio de 'nada sobre nós sem nós', fundamental aos movimentos de direitos das pessoas com deficiência, materializa-se aqui na figura do pesquisador que é simultaneamente sujeito e objeto de sua investigação.

Por fim, reafirma-se que a comunicação constitui direito humano fundamental, e que as tecnologias de síntese de voz baseadas em inteligência artificial representam ferramentas poderosas para garantir este direito a pessoas com comprometimentos da fala. Espera-se que este estudo contribua para orientar escolhas tecnológicas que impactem positivamente a qualidade de vida de usuários de CAA no Brasil.

REFERÊNCIAS

- AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION. Augmentative and Alternative Communication. 2024. Disponível em: <https://www.asha.org/public/speech/disorders/aac/>. Acesso em: jan. 2026.
- BRASIL. Lei nº 13.146, de 6 de julho de 2015. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). Diário Oficial da União, Brasília, DF, 7 jul. 2015.
- COMITÊ DE AJUDAS TÉCNICAS. Ata da Reunião VII. Brasília: CORDE/SEDH, 2007.
- COQUI TTS. Open Source Text-to-Speech. 2025. Disponível em: <https://github.com/coqui-ai/TTS>. Acesso em: jan. 2026.
- ELEVENLABS. AI Voice Generator & Text to Speech. 2025. Disponível em: <https://elevenlabs.io/>. Acesso em: jan. 2026.
- GIL, A. C. Como elaborar projetos de pesquisa. 6. ed. São Paulo: Atlas, 2019.
- GOOGLE CLOUD. Text-to-Speech AI: Lifelike Speech Synthesis. 2025. Disponível em: <https://cloud.google.com/text-to-speech>. Acesso em: jan. 2026.
- JUDGE, S.; TOWNEND, G. Perceptions of the design of voice output communication aids. *International Journal of Language & Communication Disorders*, v. 48, n. 4, p. 366-381, 2013.
- DEAPI.AI. Text-to-Speech API Platform. 2026. Disponível em: <https://deapi.ai/>. Acesso em: fev. 2026.
- MILLS, T. et al. Effects of AAC output mode on perceptions of speaking children. *Augmentative and Alternative Communication*, v. 30, n. 1, p. 84-97, 2014.
- NUNES, L. R.; WALTER, C. C. F. A Comunicação Alternativa para além das tecnologias assistivas. *Arquivos Analíticos de Políticas Educativas*, v. 24, n. 79, p. 1-18, 2016.
- ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Declaração Universal dos Direitos Humanos. Paris: ONU, 1948.
- PATEL, R. et al. Personalized synthetic voices for individuals with ALS. *Journal of Speech, Language, and Hearing Research*, v. 61, n. 12, p. 2877-2894, 2018.
- PIPER TTS. A fast, local neural text-to-speech system. 2025. Disponível em: <https://github.com/rhasspy/piper>. Acesso em: jan. 2026.
- TAN, X. et al. A Survey on Neural Speech Synthesis. *arXiv preprint arXiv:2106.15561*, 2021.
- TOBII DYNAVOX. Communicator 5 Product Support. 2024. Disponível em: <https://www.tobiidynavox.com/>. Acesso em: jan. 2026.
- VAN DEN OORD, A. et al. WaveNet: A Generative Model for Raw Audio. *arXiv preprint arXiv:1609.03499*, 2016.

WÜLFING, P. et al. The importance of voice quality in AAC: User perspectives.
Augmentative and Alternative Communication, v. 36, n. 3, p. 168-180, 2020.