



TRANSCRIÇÃO DE ÁUDIO EM TEMPO REAL PARA TEXTO: INTEGRAÇÃO DE PYTHON COM
SERVIÇO DE FALA DO MICROSOFT AZURE

REAL TIME AUDIO TO TEXT TRANSCRIPTION: PYTHON INTEGRATION WITH MICROSOFT
AZURE SPEECH SERVICE

Victor Eduardo Ramos de Almeida Costa ^{1, i}

Reinaldo de Nardi Sanches da Silva ^{2, ii}

Leonardo Aragão Ferreira Rovina ^{3, iii}

William Marini Toma ^{4, iv}

Levy Hideki Kayano ^{5, v}

Luan Morenghi Lopes ^{6, vi}

Daniel Otávio Tambasco Bruno ^{7, vii}

Sérgio Tadeu Bernatavicius ^{8, viii}

Data de submissão: (04/06/2024) Data de aprovação: (25/06/2024)

RESUMO

Este estudo descreve o desenvolvimento e a integração de um sistema de transcrição de áudio em tempo real com ênfase na acessibilidade para pessoas com deficiência auditiva. O sistema integra a biblioteca de reconhecimento de fala da Microsoft Azure com a linguagem de programação Python, através de APIs - *Application Programming Interface* (Interface de Programação de Aplicação). A metodologia empregada busca alcançar uma transcrição precisa e relevante, aproximando-se da linguagem natural utilizada na fala. Além de seu impacto direto na facilitação da comunicação para pessoas com deficiência auditiva, o sistema também possui aplicações potenciais em diversos contextos, incluindo transcrição de reuniões, legendas em vídeos e outras formas de comunicação mediada por áudio. Com a integração entre tecnologia e acessibilidade, este estudo destaca a importância da inovação na promoção da inclusão e igualdade de oportunidades em ambientes diversos.

¹ Graduando em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica. E-mail: victoreduardorac@gmail.com

² Graduando em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica. E-mail: reinaldodns@gmail.com

³ Graduando em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica. E-mail: leonardorovina@gmail.com

⁴ Graduando em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica. E-mail: william.m.toma@gmail.com

⁵ Graduando em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica. E-mail: levyh.kayano@gmail.com

⁶ Graduando em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica. Email: luanmorenghilopes@gmail.com

⁷ Mestre em Engenharia da Informação. Professor da Faculdade SENAI de Tecnologia Mecatrônica. E-mail: daniel.bruno@sp.senai.br

⁸ Doutor em Engenharia Mecânica. Professor da Faculdade SENAI de Tecnologia Mecatrônica. E-mail: sergio.tadeu@sp.senai.br

Palavras-chave: transcrição de áudio; acessibilidade; tecnologia assistiva; Microsoft Azure; Python

ABSTRACT

This paper describes the development and integration of a real-time audio transcription system with an emphasis on accessibility for people with hearing impairments. The system integrates the Microsoft Azure speech recognition library with the Python programming language, through APIs (Application Programming Interface). The methodology employed aims to achieve accurate and relevant transcription, to get closer to the natural language used in speech. In addition to its direct impact on facilitating communication for people with hearing impairments, the system also has potential applications in a variety of contexts, including meeting transcription, video captioning, and other forms of audio-mediated communication. By integrating technology and accessibility, this paper highlights the importance of innovation in promoting inclusion and equal opportunities in diverse environments.

Keywords: audio transcription; accessibility; assistive technology; Microsoft Azure; Python

1 INTRODUÇÃO

Segundo o Censo Escolar da Educação Básica 2023 (Brasil, 2023a), o número de matrículas da educação especial chegou a 1,8 milhão em 2023, um aumento de 41,6% em relação a 2019. O maior número está no ensino fundamental, que concentra 62,9% dessas matrículas. De acordo com Brasil (2023b), dos 47,3 milhões de alunos da educação básica, 61.594 possuem alguma deficiência relacionada à surdez. Muitos desses alunos possuem dificuldades para chegar até cursos técnicos profissionalizantes ou no ensino superior.

Diversos trabalhos encontrados na literatura relatam problemas de alunos surdos durante sua vida escolar, como citados por Ziliotto e Souza (2018).

A participação efetiva em aulas e atividades educacionais pode ser uma tarefa desafiadora para indivíduos surdos devido às barreiras de comunicação que enfrentam. A impossibilidade de ouvir o conteúdo apresentado pelo professor ou colegas de sala pode resultar em dificuldades significativas na compreensão e absorção de informações, bem como socialização. Essas dificuldades podem resultar em desvantagens acadêmicas e sociais, afetando negativamente o desenvolvimento educacional e profissional desses indivíduos.

A convergência da tecnologia e acessibilidade desempenha um papel crucial na promoção da inclusão e na superação das barreiras enfrentadas por pessoas com deficiência auditiva. Este artigo propõe-se a explorar as possibilidades oferecidas pela tecnologia para atender às necessidades de comunicação por meio da transcrição de áudio para texto.

Este trabalho tem como objetivo geral desenvolver um sistema voltado para auxiliar pessoas com deficiência auditiva e outros usuários em ambientes acadêmicos e profissionais, permitindo a conversão precisa e em tempo real de áudio para texto. Para isso, os objetivos específicos resumem-se a implementar um programa que faça uso dos Serviços de Inteligência Artificial – Serviço de Fala do Microsoft Azure em conjunto com Python e biblioteca gráfica, visando atender às necessidades de transcrição de áudio para texto em diversos contextos.



2 REVISÃO DE LITERATURA

Sob a perspectiva da realização do projeto, foi necessária a compreensão dos recursos e processos disponíveis para o seu desenvolvimento. Antes de chegar-se ao uso das ferramentas de reconhecimento e conversão de fala para texto, também conhecidas como STT (*Speech to Text* – Fala em Texto), é de suma importância o entendimento de como essas ferramentas funcionam. Segundo Andrade et al (2020), a fala, que é captada através de um microfone, é convertida de um som físico para um sinal elétrico, e em seguida, para dados digitais com um conversor analógico-digital. Depois de digitalizados, é implementado um modelo, como o HMM (Modelo Oculto de Markov), para transcrever o áudio para texto. Ainda diz o autor, que esses processos e algoritmos passam por complexas estruturas de programação, além da implementação de outros artifícios.

As ferramentas de reconhecimento e conversão de fala para texto atualmente estão disponíveis em diversos serviços, como as APIs (*Application Programming Interface* – Interface de Programação de Aplicação) para programação em Python, que foi empregado nesse projeto.

Ao utilizar bibliotecas e estruturas Python, pode-se criar um sistema robusto e preciso que é capaz de reconhecer e interpretar a linguagem falada e gerar a saída apropriada. Com o uso dessas ferramentas e o conhecimento aplicado, foi possível implementar o sistema desejado, onde, ao utilizar um microfone, as palavras proferidas são transcritas automaticamente para texto na tela do computador. A implementação da plataforma Microsoft Azure, que é definida por Collier e Shahan (2016) como uma plataforma de computação em nuvem da Microsoft, fornece uma ampla variedade de serviços que podem ser utilizados, como a aplicação da IA (Inteligência Artificial) e seu *Speech Service* (Serviço de Fala), para ampliar os recursos da aplicação.

A combinação da programação em Python com a plataforma Microsoft Azure viabiliza o projeto com serviço em nuvem, além da compatibilidade com os aplicativos da Microsoft.

3 METODOLOGIA

Para o desenvolvimento do sistema proposto, foi utilizado um microcomputador com sistema operacional Windows 11, processador Intel Core i5-1035 G1, com 16 GB de Memória RAM, fone de ouvido HAVIT H2002d com Microfone *HD Gaming*, acesso a uma conta educacional na Microsoft Azure, Ambiente de desenvolvimento Integrado (IDE – *Integrated Development Environment*) Visual Studio Code (versão: x64-1.90.0), em conjunto com o Kit de desenvolvimento da linguagem de Programação Python (versão: 3.12.4) e a biblioteca Tkinter para construção da interface gráfica da aplicação.

4 RESULTADOS E DISCUSSÕES

Os resultados alcançados até o momento demonstram um progresso notável no desenvolvimento do sistema de transcrição de áudio para texto, com a integração com o Microsoft Azure. A eficácia do sistema foi validada através da métrica Taxa de Erro de Palavra (WER – *Word Error Rate*). Observou-se uma WER de 4%, indicando uma alta precisão na

captura e conversão de fala para texto. A Figura 1 ilustra o fluxo de trabalho do sistema, mostrando o processo de captura do áudio até a sua transcrição em texto.

Figura 1: Representação do sistema desenvolvido



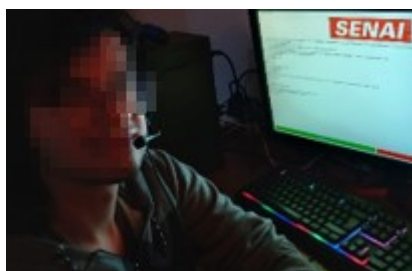
Fonte: Elaborada pelo autor.

A eficiência do processo é evidenciada pelo tempo de resposta rápido, com o sistema levando em média 2 segundos para transcrever uma fala em tempo real, demonstrando a robustez do uso de tecnologias em nuvem para processamento de dados. Ressalta-se que é necessário conexão a Internet com boa velocidade para manter essa medida.

A plataforma Microsoft Azure mostra-se uma ferramenta robusta, proporcionando recursos avançados de armazenamento e análise de dados que ampliaram o potencial do sistema para aplicativos futuros. A capacidade de escalabilidade da Azure permitiu experimentar com diferentes configurações de infraestrutura, otimizando os custos e a eficiência operacional.

No entanto, enfrentamos desafios relacionados ao acesso limitado e ao conhecimento restrito sobre a plataforma entre o público geral, que dificultam a adoção mais ampla da solução proposta. A falta de familiaridade com as funcionalidades do Azure e suas aplicações práticas sugere uma necessidade de programas de treinamento e divulgação mais efetivos, para capacitar estudantes e profissionais a explorarem plenamente as possibilidades oferecidas pela computação em nuvem.

A Figura 2, demonstra a interface do usuário do sistema de transcrição (b), destacando sua usabilidade e como os usuários interagem com a tecnologia (a) para obter transcrições em tempo real.

Figura 2: (a) Aplicação do Sistema de Transcrição de Fala para Texto

Fonte: Elaborada pelo autor

Figura 3: (b) Interface gráfica do sistema

Fonte: Elaborada pelo autor

CONCLUSÃO

O desenvolvimento do programa de conversão de áudio para texto resultou em uma solução eficiente, capaz de realizar transcrições em tempo real com precisão, sem erros gramaticais ou de pontuação. Este trabalho contribuiu significativamente para o processo de aprendizagem de pessoas com deficiência auditiva, proporcionando um acesso mais rápido e eficiente a informações no meio virtual. Além disso, o projeto promoveu a inclusão de pessoas com deficiência auditiva em diversos ambientes, como instituições de ensino superior e locais de trabalho.

Para trabalhos futuros, pretende-se avaliar de forma sistemática através da aplicação de questionários e casos de uso o sistema, em diferentes níveis de ensino e de deficiência. Além disso, pretende-se implantar funcionalidades que permitam a transcrição do texto para a Língua Brasileira de Sinais (Libras), ampliando ainda mais o alcance e a inclusão de pessoas com deficiência auditiva. Essa expansão do projeto visa proporcionar um suporte ainda mais abrangente, atendendo a um número maior de indivíduos.

REFERÊNCIAS

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Diretoria de Estatísticas Educacionais. **Censo escolar da educação básica 2023: resumo técnico versão preliminar**. Brasília, DF: INEP, 2023.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Confira o panorama dos surdos na educação brasileira**. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/censo-escolar/confira-o-panorama-dos-surdos-na-educacao-brasileira>. Acesso em: 9 jun. 2024.

COLLIER, Michael; SHAHAN, Robin. **Fundamentals of Azure**. [S. l.]: MS Press, 2015. Disponível em: <https://scadahacker.com/library/Documents/eBooks/MS%20Press%20-%20Fundamentals%20of%20Azure.pdf>. Acesso em: 12 maio 2024.

ANDRADE, Acácio de; MOURA, Shayenne; GOLDMAN, Alfredo. Processamento e Transcrição de Voz em Língua Portuguesa voltado para Assistente Inteligente. *In: ESCOLA REGIONAL DE APRENDIZADO DE MÁQUINA E INTELIGÊNCIA ARTIFICIAL DE SÃO PAULO, 1., 2020, São Paulo. [Anais...]* Porto Alegre: Sociedade Brasileira de Computação, 2020.

MICROSOFT. **Speech to text transcription and analytics architecture**. Disponível em: <https://learn.microsoft.com/pt-br/azure/architecture/ai-ml/architecture/speech-to-text-transcription-analytics>. Acesso em: 9 jun. 2024.

AGRADECIMENTOS

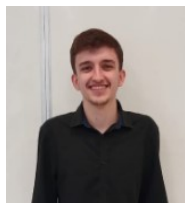
Agradecimentos iniciais ao nosso orientador Prof. Daniel Otávio Tambasco Bruno, pela orientação, paciência e apoio ao longo de todo o processo de elaboração deste trabalho. A todos os participantes do grupo pelo empenho, dedicação e suporte diretos e indiretos em todos os campos de pesquisa do trabalho. A integração e comunicação de todos foi fundamental para que fosse obtido um funcionamento pleno e organização adequada. À instituição de ensino e demais professores que disponibilizaram recursos e materiais necessários para realização desta pesquisa, além do tempo e atenção dedicados para atender as requisições providas durante o período de elaboração do trabalho. Por fim, agradecimentos a todos aqueles que, direta ou indiretamente, contribuíram para o sucesso deste projeto acadêmico. Todas as contribuições foram inestimáveis e serão lembradas com extrema gratidão.

Sobre os Autores:

ⁱ Victor Eduardo Ramos de Almeida Costa



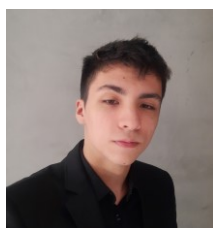
Possui formação Técnica em Mecatrônica (2020) e Tecnóloga em Mecatrônica Industrial (2023). Graduando em Engenharia de Controle e Automação Faculdade SENAI de Tecnologia Mecatrônica. Atualmente trabalha como Analista PLM pela TECMES. Tem experiência na área de projetos para o setor automotivo e aeroespacial utilizando softwares CAD/CAE. Atua como especialista nas soluções Dassault Systèmes. Possui conhecimentos sólidos na área de programação. <https://orcid.org/0009-0006-2476-4656>

ii Reinaldo de Nardi Sanches da Silva

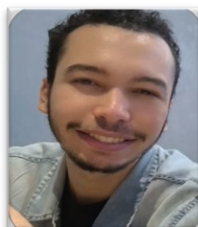
Formação Técnica em Mecatrônica (2020) e Tecnóloga em Mecatrônica (2023), na qual foi escolhido como melhor aluno da turma pela equipe de docentes. Está graduando Engenharia de Controle e Automação. Todos na Faculdade SENAI de Tecnologia Mecatrônica. Tem experiência em Automação Industrial, com ênfase em projetos com CLP e IHM. Campeão mundial de robótica em 2019 pela equipe Robonáticos. Atualmente é Técnico de Laboratório Didático na rede SESI-SP, responsável por um ambiente maker e pelo desenvolvimento de projetos visando o incentivo ao ensino e protagonismo dos alunos. <https://orcid.org/0009-0001-3274-4843>

iii Leonardo Aragão Ferreira Rovina

Possui formação Técnica em Mecatrônica (2020) e Tecnólogo em Mecatrônica (2023). Atualmente cursa graduação em Engenharia de Controle e Automação na Faculdade SENAI de Tecnologia Mecatrônica e trabalha como Estagiário em Automação Industrial pela empresa BCG. <https://orcid.org/0009-0003-6023-4036>

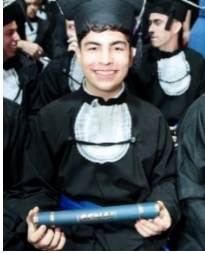
iv William Marini Toma

Possui formação técnica em Mecatrônica (2019), técnica em Eletroeletrônica (2020) e Superior de Tecnologia em Mecatrônica Industrial (2023). Graduando em Engenharia de Controle e Automação pela Faculdade SENAI-SP Campus Armando de Arruda Pereira. Atualmente trabalha como Técnico em Eletrônica na NELMAR. Tem experiência na área de engenharia clínica e aparelhos médicos. Conhecimentos sólidos em manutenção, gestão, prototipagem e eletrônica. <https://orcid.org/0009-0008-6692-0279>

v Levy Hideki Kayano

Possui formação Técnica em Superior de Tecnologia em Automação Industrial (2023). Está fazendo graduação em Engenharia de Controle e Automação Faculdade SENAI de Tecnologia Mecatrônica. Atualmente trabalha como Assistente de Processos Junior pela The PAC Group. Experiências de trabalho em elaboração e interpretação de diagramas; correlação técnica de desenho técnico; e manuseio de ferramentas da parte mecânica e elétrica. Possui conhecimentos sólidos em esquemas-elétricos. <https://orcid.org/0009-0000-4633-8398>

vi Luan Morengi Lopes



Possui formação em Tecnologia em Mecatrônica Industrial (2023). Está fazendo graduação em Engenharia de Controle e Automação Faculdade SENAI de Tecnologia Mecatrônica. Atualmente trabalha como Analista de Dados e Processos na SCANIA. Tem experiência sólida na área de compras e projetos técnicos. Possui conhecimentos sólidos na área de dados e ferramentas da Power Platform. <https://orcid.org/0009-0009-8222-2988>

vi Daniel Otávio Tambasco Bruno



Doutorando e Mestre em Engenharia da Informação pela Universidade Federal do ABC. Especialista em Banco de Dados pela Universidade de Ribeirão Preto (2007) e Especialista em Educação à Distância pela Universidade Paulista (2012). Possui graduação em Análise de Sistemas (2003). Atualmente é Professor da Faculdade SENAI de Tecnologia Mecatrônica industrial. <https://orcid.org/0000-0002-9007-338X>

vii Sérgio Tadeu Bernatavicius



Pós-Doutorado na PUC-SP, possui Mestrado e Doutorado em Engenharia Mecânica pela UNICAMP. Graduado em Engenharia Ambiental pela UFSCar (2013), Engenharia Mecânica pela UNESP (1988) e Engenharia de Computação pela UNIVESP, (2021). É professor na Pontifícia Universidade Católica de São Paulo (PUC-SP) e no MBA em Gestão de Projetos e Produção do SENAI-SP e FIAP. <https://orcid.org/0000-0002-9308-4981>