CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA
Faculdade de Tecnologia de Santana de Parnaíba
Curso Superior de Tecnologia em Ciência de Dados

Joyce Leal Bezerra da Silva
Luzia Gabriela Abreu da Silva Santos
Vinicius Augusto Alves da Silva

SPEECH SNAKE: APLICAÇÃO DE ANÁLISES GRÁFICAS SOBRE A VOZ E AS CLASSIFICAÇÕES DE FLUÊNCIA NA FALA

Santana de Parnaíba
2023

Joyce Leal Bezerra da Silva
Luzia Gabriela Abreu da Silva Santos
Vinicius Augusto Alves da Silva

SPEECH SNAKE: APLICAÇÃO DE ANÁLISES GRÁFICAS SOBRE A VOZ E AS CLASSIFICAÇÕES DE FLUÊNCIA NA FALA

Trabalho de Graduação apresentado à Faculdade de Tecnologia de Santana de Parnaíba como requisito parcial para a obtenção do título de Tecnólogo em Ciência de Dados, sob a orientação do Professor Sérgio Luciano de Oliveira Soares e do Professor Dr. Celio Aparecido Garcia.

Santana de Parnaíba
2023

ATA DE APROVAÇÃO

Este trabalho é
dedicado aos professores e
alunos da Fatec Santana de Parnaíba.

AGRADECIMENTOS

Agradecemos a todos que de alguma forma contribuíram e incentivaram nossa formação.

Aos nossos familiares que nos motivaram em nossa jornada e não permitiram a desistência em meio às dificuldades.

Aos nossos orientadores, Professor Sergio Luciano de Oliveira Soares e Professor Dr. Celio Aparecido Garcia, que dedicaram tempo para nos acompanhar e auxiliar na elaboração do Trabalho de Graduação, e aos demais professores do curso de Ciência de Dados que através dos seus ensinamentos, viabilizaram a conclusão deste projeto.

Aos nossos amigos que compartilharam as mesmas expectativas e algumas vezes angústias, vivenciaram e vibraram a cada etapa vencida nesta fase de graduação.

Se falares a um homem numa linguagem que ele compreenda, a tua mensagem entra na cabeça. Se falares na sua própria linguagem, a tua mensagem entra-lhe diretamente no coração.

Nelson Mandela

SILVA, Joyce L. B.; SANTOS, Luzia Gabriela A. S.; SILVA, Vinicius Augusto A. **SPEECH SNAKE: APLICAÇÃO DE ANÁLISES GRÁFICAS SOBRE A VOZ E AS CLASSIFICAÇÕES DE FLUÊNCIA NA FALA.** 41 f. Trabalho de Conclusão de Curso de Tecnólogo em Ciência de Dados. Faculdade de Tecnologia de Santana de Parnaíba. Centro Estadual de Educação Tecnológica Paula Souza. Santana de Parnaíba. 2023.

RESUMO

Este trabalho de graduação descreve o desenvolvimento da aplicação web *open-source* nomeada de Speech Snake, que é capaz de gerar e analisar de forma gráfica os dados de áudio com foco na voz e na fluência, a partir da extração automatizada de métricas em áudios reais. A metodologia adotada envolveu a utilização da linguagem de programação Python em conjunto com Javascript, ferramentas, técnicas e bibliotecas auxiliares relacionadas ao processamento de dados de áudio, acompanhadas por pesquisas qualitativas especializadas na descrição de características e métricas fonéticas importantes dedicadas a voz e fluência. Por meio das bibliotecas de processamento de áudio e com o conhecimento adquirido, foram selecionados métodos para a visualização gráfica das análises, como WaveForm, Frequência Fundamental (F0), Espectrograma e, para a classificação da clareza da voz, foram utilizadas as Frequências Formantes (F1 e F2). Os resultados obtidos foram diferentes dos planejados inicialmente devido à alta complexidade do tema, no qual a aplicação demonstrou eficiência na extração e manipulação dos parâmetros e métricas dos dados de áudio, assim como, nas análises gráficas correspondentes. Como conclusão, apesar do campo de estudo emergente, o Speech Snake atingiu o objetivo de possibilitar e facilitar a análise dos dados de áudios de voz, com um grande potencial de utilidade nos campos de Ciência de Dados, processamento de áudios e sistemas especialistas relacionados as áreas de estudo.

Palavras-chave: dados; fluência; processamento de áudios; Python; voz.

SILVA, Joyce L. B.; SANTOS, Luzia Gabriela A. S.; SILVA, Vinicius Augusto A. **SPEECH SNAKE: APPLICATION OF GRAPHIC ANALYSIS ON VOICE AND SPEECH FLUENCY CLASSIFICATIONS.** 41 p. End-of-course paper in Technologist Degree in Data Science. Faculdade de Tecnologia de Santana de Parnaíba. Centro Estadual de Educação Tecnológica Paula Souza. Santana de Parnaíba. 2023.

ABSTRACT

This graduation work describes the development of the *open-source* web application named Speech Snake, which is capable of graphically generating and analyzing audio data with a focus on voice and fluency, from the automated extraction of metrics in real audios. The methodology adopted involved the use of the Python programming language together with Javascript, tools, techniques and auxiliary libraries related to the processing of audio data, accompanied by qualitative research specialized in the description of important phonetic characteristics and metrics dedicated to voice and fluency. Through the audio processing libraries and with the acquired knowledge, methods were selected for the graphic visualization of the analyzes, such as WaveForm, Fundamental Frequency (F0), Spectrogram and, for the classification of the voice clarity, the Formant Frequencies were used (F1 and F2). The results obtained were different from those initially planned due to the high complexity of the theme, in which the application demonstrated efficiency in the extraction and manipulation of the parameters and metrics of the audio data, as well as in the corresponding graphic analyses. In conclusion, despite the emerging field of study, Speech Snake achieved the objective of enabling and facilitating the analysis of voice audio data, with great potential for use in the fields of Data Science, audio processing and expert systems related to Study areas.

Keywords: data; fluency; audio processing; Python; voice.

LISTA DE ILUSTRAÇÕES

SUMÁRIO

1 INTRODUÇÃO

Os distúrbios de linguagem afetam a comunicação de forma negativa e impactam a qualidade de vida emocional e social do indivíduo, pois comprometem a estabilidade na produção vocal, esforço de fonação, além de desconfortos para manter uma conversa fluida (LÓSS et al., 2020). A melhoria da articulação da fala deve ser orientada por profissional habilitado. Apesar do incômodo e do comprometimento da comunicação, é comum que a demanda por ajuda seja tardia, mas a identificação precoce do problema pode reforçar a busca por auxílio profissional.

O foco deste projeto é a criação de uma aplicação web que possa analisar dados em arquivos de áudio a partir da extração e coleta de métricas importantes no processamento de voz, que podem ser utilizadas como base para o aprofundamento de campos especialistas no estudo da voz humana.

Inicialmente a motivação do projeto era desenvolver uma Inteligência Artificial para auxiliar o profissional de Fonoaudiologia no tratamento de voz e fluência, a fim de diagnosticar e acompanhar pacientes com transtornos relacionados, porém devido à complexidade da proposta e tempo limite para o desenvolvimento, o grupo optou por alterar o escopo com a exclusão da especificidade profissional e discorrer apenas sobre a análise e processamento de áudios que podem ser aplicados a qualquer pessoa que tenha interesse no campo de pesquisa.

O objetivo geral deste trabalho de graduação é utilizar conceitos de Aprendizado de Máquina (ML) e Processamento da Linguagem Natural (PLN) para desenvolver um software com uso das linguagens de programação Python e Javascript, no qual deverá ser capaz de extrair métricas relevantes sobre o processamento de dados de voz, como fonte das métricas a serem obtidas para as análises de voz e fluência.

A aplicação Speech Snake permitirá a interação do usuário para carregar o arquivo nos formatos MP3 ou WAV e os resultados provenientes da análise, que será feita por um algoritmo, serão apresentados por gráficos e métricas descritivas baseados em conhecimentos multidisciplinares adquiridos durante o período de curso em conjunto com pesquisas relacionadas à análise de voz e fluência.

O tratamento envolve duas categorias:

- Analisar voz: Efetua análises direcionadas a respiração, pronúncia e entonação de frases.
- Analisar fluência: Efetua análises relacionadas a consistência e fluidez da fala, com atenção para possíveis gaguejos e estabilidade em uma conversa.

O projeto foi elaborado a partir da pesquisa bibliográfica e da metodologia de natureza aplicada e exploratória sobre a relação entre voz e fluência e a utilização de ferramentas e técnicas, assim como, linguagens de programação e suas bibliotecas, direcionadas ao tratamento sonoro. Para uma compreensão mais adequada nesta área, consideramos o estudo de alguns tópicos:

- Processamento de áudios;
- Conversão de fala para texto;
- Extração de dados de áudio;
- Reconhecimento da fala;
- Características fonéticas;
- Conversão de áudios;
- Intensidade e análise de áudios;
- Sintetização de voz;
- Eliminação de ruídos;
- entre outros.

O arquivo de áudio deverá conter uma gravação pré-definida (trava-línguas) que deverá ser carregada na aplicação para que a análise dos dados seja realizada pelo algoritmo, com intuito de processar e apresentar os resultados em espectrogramas que demonstram a visualização gráfica do tempo e frequência de voz do áudio analisado.

2 REFERENCIAL TEÓRICO

O principal objetivo do ensino da língua é desenvolver a competência comunicativa para aplicar adequadamente nas diversas situações que envolvem sequências linguísticas gramaticais na construção de orações, frases, também habilidade de produção e compreensão textual. A importância do conhecimento de como a língua está constituída e como funciona para o uso adequado junto com habilidades de raciocínio e pensamento científico são importantes para vários campos do conhecimento humano (TRAVAGLIA, 2009).

A partir deste contexto apresentamos o que é voz, fluência, a relação com a análise linguística, o reconhecimento da fala e as técnicas aplicadas para o desenvolvimento do projeto.

2.1 Voz

O distúrbio que limita a comunicação de forma negativa é caracterizado por alterações na qualidade, frequência, intensidade ou esforço vocal que causa impacto na qualidade de vida emocional e social do indivíduo. Este distúrbio é denominado disfonia e pode ser dividido em diferentes etiologias, como por exemplo, orgânica, organofuncionais ou funcionais (MORETI; ZAMBON; BEHLAU, 2014).

- Orgânica - afeta a estabilidade na produção vocal a partir do comprometimento da fonte glótica ou trato vocal que pode acarretar maior esforço de fonação.
- Organofuncionais ou funcionais - geralmente possuem alteração da voz pela presença de sintomas frequentes, mas pouco relatados, em que o indivíduo se adaptou.

A presença de sintomas físicos é normalmente mencionada por indivíduos com qualquer distúrbio vocal, como dor ou desconforto na garganta, e são comuns para todas as tipologias de problemas de voz (MORETI; ZAMBON; BEHLAU, 2014).

2.2 Fluência

A fluência é caracterizada pela capacidade de manter uma fala fluída associada ao ritmo, entonação e velocidade, em busca do significado das palavras em relação às estruturas fonológicas, lexicais, morfológicas, sintáticas e semânticas da linguagem, que estão associadas a respiração, produção de voz, articulação, pensamento e transformação em ato audível para reconhecimento da linguagem, em que as alterações são denominadas disfluências e podem ser classificadas como normais (não gagas), anormais (gagas) ou ambíguas (algumas vezes normais e outras anormais) (SOUZA; CARDOSO, 2013).

- Disfluências normais (não gagas) - esporadicamente possuem prolongamento de sons, repetição de palavras ou sintagmas acompanhados de leve esforço motor em que a sequência, durabilidade e ritmo não são afetados e não há medo ao falar.
- Disfluências anormais (gagas) - identificadas pelo maior prolongamento de sons, repetições e esforço motor durante a fonação em que o indivíduo apresenta tensão muscular articulatória e medo de falar.

A fluência é a área em que é possível identificar as tipologias de disfluências típicas e atípicas para o tratamento precoce de transtornos relacionados que envolvem atividades de controle da fala e cognitivo-linguísticas de auto-percepção (SOUZA; CARDOSO, 2013).

2.3 Análise Linguística

Segundo Volker Noll (2008), as principais características fonéticas do português brasileiro em relação ao português europeu incluem a pronúncia das vogais, consoantes e entonação. Esses traços são influenciados por outras línguas, como as línguas indígenas e africanas, as diferentes regiões do Brasil, e ainda, fatores sociais e culturais que são importantes para a compreensão da língua falada e para o ensino de língua portuguesa.

O autor descreve essas características fonéticas da seguinte forma:

- Vogais: O português brasileiro possui cinco vogais que são pronunciadas de forma mais aberta do que no português europeu. Além disso, apresenta uma pronúncia mais nasalizada nas vogais, especialmente em regiões como o Nordeste e o Norte do país.
- Consoantes: O português brasileiro apresenta algumas diferenças na pronúncia das consoantes em relação ao português europeu. Como exemplo, podemos citar a letra "s", que é pronunciada como "sh" em algumas regiões do Brasil, enquanto em outras é pronunciada como um som mais parecido com "s". Já a letra "r" também é pronunciada de forma diferente em algumas regiões do país, com uma vibração mais fraca ou forte conforme a região.
- Entonação: O português brasileiro apresenta uma entonação mais melódica do que o português europeu, com um maior uso de inflexões vocais ascendentes e descendentes. Além disso, apresenta um maior uso de pausas e prolongamentos de vogais em algumas regiões.

Com a intenção de reduzir o impacto das variações linguísticas na análise de voz e fluência realizada pela aplicação, estipulamos o seguinte trava-línguas popular como padrão para a gravação do áudio que será analisado: *“Três tigres tristes para três pratos de trigo. Três pratos de trigo para três tigres tristes.”*

Dessa maneira, o usuário deve gravar o áudio com a pronúncia do trava-línguas, com limite máximo de 30 segundos (3MB/3.539KB de tamanho), nos formatos MP3 ou WAV, em ambiente silencioso para evitar falhas na análise quanto a possíveis interferências de ruídos externos.

2.4 Processamento de Áudios

Para acompanhar a rápida evolução tecnológica sucedida nos sistemas no decorrer dos anos, foi necessário o aperfeiçoamento de muitos sistemas, e entre eles, um dos estudos emergentes é o processamento de áudios. O estudo aprofundado neste campo viabiliza poderosas ferramentas, softwares e análises deste tipo de dado. (TAFNER; MALCON, 1996.)

Existem muitos aplicativos e métodos de captura, assim como, algoritmos inteligentes que são capazes de extrair informações de sons.

Máquinas não conseguem processar ou interpretar um áudio, elas apenas entendem o sistema booleano de Código Binário (0 ou 1), o que acontece é que existem diversos conversores de dados para binário que utilizam boa parte do processamento nesta ação (SENAI-SP EDITORA, 2018).

Os computadores e equipamentos eletrônicos em geral costumam armazenar o código binário em unidades de bits, no qual um bit equivale a 1 ou 0, e isso é gravado na memória para realizar tarefas como ligar e desligar.

A capacidade de armazenamento varia em relação ao dispositivo, mas as escalas utilizadas para quantificar os bits são:

- 1 bit equivale a uma unidade de binário;
- 1 byte equivale a 8 bits;
- 1 kilobyte equivale a 1024 bytes;
- 1 megabyte equivale a 1.000.000 bytes;
- 1 gigabyte equivale a 1.000.000.000 bytes.

E assim sucessivamente, até escalas de Terabytes e Petabytes, nas quais possuem uma quantidade enorme de bytes com diversas possibilidades de combinação de bits. Atualmente poucas empresas no mundo trabalham na escala de Petabytes, como por exemplo, Google, Amazon, IBM, Fórmula 1, Petrobras, entre outras.

Existem dois tipos de sinais em que os dispositivos eletrônicos conseguem trabalhar, são eles: o sinal digital e o sinal analógico.

Um sinal analógico possui variações senoidais em sua tensão, ou seja, ele aumenta e diminui gradativamente até chegar em seus valores máximos e mínimos, e isto segue de forma contínua e ondular (SENAI-SP EDITORA, 2018).

O sinal digital por sua vez, só pode assumir valores booleanos que podem ser chamados de ligado ou desligado, a transição da tensão do valor máximo ao mínimo ocorre de forma tão rápida que pode ser considerado algo quase instantâneo. Por questão de conveniência, o uso de aparelhos eletrônicos é bem mais vantajoso do que o uso do sinal analógico, visto que é muito mais simples trabalhar apenas com dois níveis de tensão do que com vários. Há certas vantagens que definiram o sinal digital como predominante na maioria dos sistemas eletrônicos da atualidade, são

elas: o Tratamento, Armazenamento e Transmissão de informações (SENAI-SP EDITORA, 2018).

O ruído é o principal motivo da utilização de sinais digitais na conversão do som. Em uma música com sinal analógico, além das ondulações lineares, por ser algo volátil, a ocorrência de oscilações indesejadas pode prejudicar a melodia com interferências ou chiados. No caso do sistema digital, os ruídos também se tornam irrelevantes, um exemplo comparativo deste com o analógico são os CDs e os discos de vinil.

Os sistemas eletrônicos são orientados a níveis lógicos através de portas de entrada e saída, controlados por portas lógicas baseadas na tabela-verdade: OU, E, INVERSORA, NOU, NE, OU EXCLUSIVO, NOU EXCLUSIVO ou COINCIDÊNCIA (SENAI-SP EDITORA, 2018).

O som é armazenado e processado de forma digital nos eletrônicos e convertido para sinal analógico no processo de emissão de músicas em caixas de som ou fones de ouvido. Esta alteração é feita por conversores D/A (digital para analógico). Logo, para o efeito contrário, são utilizados conversores A/D para transformar o sinal analógico em digital e gravar os sons (SENAI-SP EDITORA, 2018).

Figura 1. Conversão de sinal analógico para digital

Fonte: PIERRI, R. (2022)

Um conversor A/D ao converter sons, utiliza um método de partição na frequência (Hz) da voz, no qual captura pontos específicos das ondas analógicas

geradas pelo áudio. Na gravação de todos os pulsos de tensão, serão compactados e armazenados, mas vale ressalvar que ao realizar a conversão novamente para sons analógicos, as ondas retornarão retificadas por causa da digitalização feita anteriormente (SENAI-SP EDITORA, 2018).

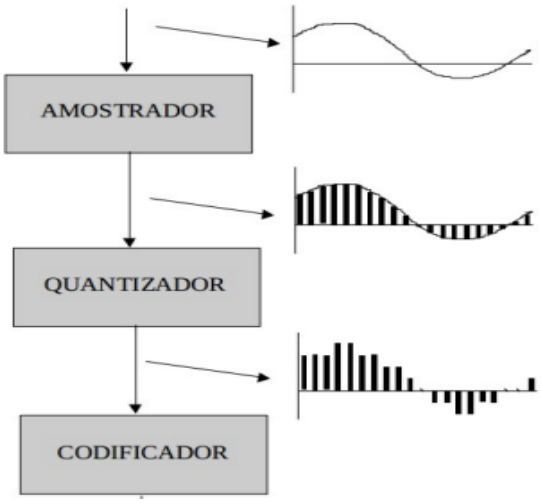
Uma onda sonora convertida por A/D corta todas as ondulações existentes e consequentemente descartadas, o que auxilia no tratamento de ruídos, mas existe a chance de cortar pontos importantes do sinal, então deve-se colocar o máximo de pontos possíveis nas ondulações (SENAI-SP EDITORA, 2018).

Segundo Deise (2016), cada ponto gravado no conversor é armazenado em uma variável no processador e estes tornam-se dados chamados de *Pulse Code Modulation* (PCM), que agregam os conceitos mencionados anteriormente em três etapas:

1. Amostragem;
2. Quantização;
3. Codificação.

Na amostragem, o sinal analógico é dividido em pulsos lineares com um tempo fixo. Ao chegar na quantização, os pulsos são divididos e arredondados para sinais digitais e enfim, na codificação os sinais são transformados em bits (DEISE, 2016.

Figura 2. Etapas do PCM



Fonte: DEISE, M. (2016, p.3)

O desenvolvimento da biometria foi um dos principais precursores para a evolução nos métodos digitais de captação de amostras humanas pelo método

comportamental, no qual está o reconhecimento da voz. A captação da fala é feita por um microfone que realiza a transformação em sinal digital e compara com outro que foi gerado e armazenado em uma base de dados (FERREIRA, 2003).

Uma das formas de transformar o som de fala humana em sinais digitais envolve o uso do *Linear Predictive Coding* (LPC), considerado uma das maneiras mais eficazes e práticas de extração de parâmetros de sinal de voz (FERREIRA, 2003).

2.5 Métricas Importantes no Processamento de Áudios

Com o surgimento de novas tecnologias emergentes focadas no reconhecimento de voz, execução de tarefas pela fala e outras aplicações nesta área que envolve Inteligência artificial, buscam formas cada vez mais eficientes do processamento de áudio que se tornou de suma importância para possibilitar o uso dos dados extraídos de forma mais eficiente (BARBACENA, 2010).

Existem diversas métricas e dados importantes que podem ser extraídos de um áudio. Consideramos as seguintes:

- Tempo (s): Duração que o áudio possui, seus principais valores podem ser atribuídos nas métricas de milissegundos (ms), segundos (s), minutos (m) e horas (h) (SENAI-SP EDITORA, 2018).
- Frequência (Hz): É uma métrica comum para a engenharia, na qual se baseia na repetição de ciclos de ondas geradas por algum sinal emissor em decorrer do tempo. Possui variações que podem ser aplicadas de acordo com o uso, mas geralmente é dada em Hertz (Hz) nas literaturas (SENAI-SP EDITORA, 2018).
- Amplitude (Pa): Mensuração da pressão sonora/magnitude de um áudio, é calculado na unidade de Pascal. São os picos de onda sonora a partir de seus valores médios (BARBACENA, 2010).
- Frequência Harmônica: São múltiplos inteiros de F0 que se juntam para formar vibrações das moléculas de ar (BARBACENA, 2010).
- Frequência Formantes: Usada para identificar a posição articulatória da vogal falada. A ressonância é utilizada para definir a qualidade vocal através do timbre. Muito importante para cantores e sons melódicos (BARBACENA, 2010).

- Decibel (dB): Métrica que mensura a intensidade extraída de um dado de áudio, de forma geral é a altura do som. Também possui valores genéricos para idade e gênero (BARBACENA, 2010).
- Timbre: Não é exatamente uma métrica aferida e sim uma análise harmônica geral, que utiliza as métricas anteriores para visualizar o áudio de forma espectral com a intensidade no formato de cores (BARBACENA, 2010).
- Taxa de Amostragem (Hz): É uma das características mais importantes do áudio. Efetua a divisão do áudio em pontos de dados, também é mensurado em Hz, mas a frequência deste é voltada para o armazenamento do áudio, que influencia diretamente na qualidade sonora (BARBACENA, 2010).
- Razão Harmônico-Ruído (HNR): Medida da proporção entre a energia das componentes harmônicas e a energia das componentes não-harmônicas (ruído) de um sinal de voz. Muito usada para análise qualitativa da clareza de voz (BARBACENA, 2010).
- Coeficientes do Cepstrum: São representações compactas e informativas das características espectrais de um áudio (BARBACENA, 2010).

2.6 Outros Métodos de Reconhecimento da Fala

Uma das áreas de pesquisa do Processamento da Linguagem Natural está focada no estudo de métodos para a conversão da fala para o texto na linguística computacional. Isso é possível graças ao desenvolvimento de algoritmos especializados que conseguem realizar essa transformação.

Como descrito nos serviços AWS (2022), existem dois tipos de transcrição da fala-texto: a literal, em que a replicação do que foi falado é reproduzido textualmente com precisão e fidelidade ao som original, e a adaptada, que busca uma maior compreensão do que foi dito para retornar uma frase corrigida e sem problemas, como vícios de linguagem.

Os sistemas *Speech-to-text* cresceram muito nos últimos anos, com grandes empresas que entraram a fundo no tema, como Google, Amazon, Microsoft, entre outras. Eles também utilizam sinais digitais de áudio e voz para entender um áudio em conjunto com a linguagem natural para processá-los (IBM, 2022).

Algoritmos que envolvem o reconhecimento da fala precisam obter a maior acurácia possível na tradução de um áudio para um texto, visto que o sentido e coesão das frases depende disso. Para tal, são utilizadas muitas técnicas e ferramentas de processamento da linguagem natural, modelos de Marcov Ocultos, N-gramas, redes neurais entre outros (IBM, 2022).

Ao captar as vibrações do emissor das palavras, as ondas sonoras são filtradas e divididas por fonemas da língua tratada em questão, em seguida ocorre uma série de validações a partir de comparações com palavras já conhecidas para retornar a melhor acurácia de igualdade e, só então, replicar como texto (AWS, 2022).

O uso preciso da captação de fala para texto possibilita o tratamento dos áudios com maior eficácia, uma vez que torna viável a comparação com outros modelos e aponta melhorias em relação a vícios de linguagem, gagueiras e outros pontos relacionados à fonética (IBM, 2022).

Assim como existem métodos para escutar a voz e armazenar com o apoio do reconhecimento da fala, também há a sintetização da voz. Esta realiza o efeito de emissão de texto para fala. Isto é possível graças a processos em que o texto é segmentado e tratado em sua estrutura com ênfase gramatical, para em seguida, ser analisado de acordo com o idioma e fonemas pertencentes. Com isso, ocorre a etapa na qual será realizado o processamento necessário para sincronizar e editar a fala de forma melódica, a fim de deixá-la fluida e, posteriormente, transformar em um sinal sonoro (IBM, 2022).

2.7 Impactos da Inteligência Artificial no Processamento de Áudio

A abordagem sobre a Inteligência Artificial (IA) é muito comum ao falarmos de tecnologia, com razão, pois estes algoritmos inteligentes têm revolucionado rapidamente muitos conceitos e metodologias no decorrer dos anos. A IA está presente em ramos de estudo diversificados e o cenário é semelhante nos casos que contêm áudios, pois é responsável por funções muito úteis, como o aumento da acurácia na reprodução, armazenamento ou comparação de sons (KERSCHBAUMER, 2018).

Diferentemente de um *chatbot*, que possui um script com um banco de dados com uma gama de perguntas e respostas pré-definidas em um sistema de progressão

arbóreo, a Inteligência Artificial (IA) é mais sofisticada ao ponto de analisar o que foi dito e interpretar com base em conhecimentos provenientes de aprendizado, ou seja, à medida que recebe mais informações, ela se adapta e aumenta a efetividade nas respostas de acordo com o seu grau de desenvolvimento (LIPPMANN, 1997).

O desenvolvimento de algoritmos inteligentes abrange muitas áreas da neurociência, uma vez que o objetivo é recriar as funções de um cérebro humano de forma digital. Há uma grande semelhança entre ações de uma máquina e uma pessoa visto que ambos seguem uma sequência lógica e ordenada para realizar qualquer tipo de ação ou raciocínio.

Ainda existem lacunas no âmbito que uma máquina não consegue imitar com perfeição um cérebro humano, então como forma de suprir essa questão, é necessário a criação de algoritmos voltados ao aprendizado de máquina, que como o nome diz, realizam uma série de treinamentos para que a Inteligência Artificial (IA) consiga aprender de modo autônomo. Geralmente o foco é específico para alguma função e a forma com que são realizados os ensinamentos são baseados em testes de bases de dados pré-estabelecidas como treinamento, até alcançar a precisão aproximada de 100% nos resultados. Após isto, é realizado um teste com bases não treinadas para verificar a acurácia em um teste real (KERSCHBAUMER, 2018).

O cérebro possui muitos nódulos e cada um tem sua função estimulante no corpo humano através de memórias que são responsáveis por realizar as ações, como por exemplo, formar os fonemas que farão uma sinapse por função que deve ser ativada sob demanda dos sensores receptores humanos (visão, audição e tato). Quando há interferências que impedem esta comunicação dá-se o nome de patologias. Entre elas estão aquelas que restringem de muitas maneiras as formas com que as pessoas se comunicam com a boca, como a gagueira, a inconsistência na frequência, a volatilidade e a durabilidade da voz. Neste caso é importante o tratamento adequado realizado por um profissional habilitado (RIBEIRO, 2020).

Sistemas especialistas são fundamentais para o desenvolvimento de inteligências artificiais com foco específico, pois conduzem o algoritmo no processamento de informações. São baseados em conhecimento capaz de simular os procedimentos que uma pessoa especializada seguiria para chegar a determinada conclusão, isto é construído de forma regrada e acompanhada (KERSCHBAUMER, 2018).

O desenvolvimento de um sistema especialista deve ser feito com acompanhamento de profissionais da área específica junto aos programadores, para que possibilite a criação de uma aplicação realmente útil e consistente. Em uso, deve auxiliar e facilitar a tomada de decisões sem substituir a mão de obra humana qualificada para aquela função (KERSCHBAUMER, 2018).

É preciso tratar o desenvolvimento de sistemas especializados de forma delicada, conforme mencionado anteriormente, na voz existem muitos tipos de dados que podem ser coletados e tratados para obter informações relevantes aos profissionais, no entanto, tais sistemas não substituem a competência profissional.

2.8 Utilização da Análise de Dados na Aplicação Speech Snake

A partir das métricas obtidas e em conjunto com as bibliotecas dedicadas ao processamento de dados de voz, foram escolhidas algumas análises relevantes em relação ao tratamento do áudio.

Podemos destacar a análise das Frequências Formantes através do mapeamento realizado no Primeiro Formante (F1) e Segundo Formante (F2), em seguida, para as representações gráficas selecionadas: WaveForm, Frequência Fundamental (F0) e Espectrograma.

Os formantes representam as frequências naturais de ressonância do trato vocal, na posição articulatória específica da vogal falada. Os formantes são geralmente expressos através do seu valor médio e designados de forma progressiva F1, F2, F3, em que a descrição das vogais, quase nunca ultrapassa a identificação dos três primeiros formantes que determinam a qualidade vocal, em termos acústicos, e sua identidade em termos auditivos (BEHLAU, 2004).

De acordo com a Fonoaudiologia e Medicina (USP, 2023), o comprimento do trato vocal de um adulto médio mede em torno de 17cm. Esses tubos têm a propriedade de reforçar, por efeito de ressonâncias, os sons com 500Hz, 1500Hz, 2500Hz e 3500Hz, e assim indefinidamente. Esta característica é transferida ao som laríngeo, que reforça os harmônicos ao redor destas frequências que formam faixas de concentração de energia, conhecidas como Formantes (F1, F2, F3, F4…).

A partir de testes identificamos a importância de inserir parâmetros que pudessem ser usados de forma geral, com a importância de destacar que essa

alteração não se trata de regra absoluta, mas é adequada pela diferença de sexo e o ambiente que o áudio foi gravado. Desse modo, consideramos os valores de Tafner (1996) e parametrizamos F1 na faixa de 300Hz a 900Hz, enquanto os valores de F2 na faixa de 800Hz a 2800Hz, para a voz ser considerada clara. Caso contrário, a voz é considerada confusa.

Em estudo posterior, consideramos a necessidade de aplicar 25% de range no Primeiro Formante (F1). Essa alteração é relacionada a possíveis interferências no fundo do áudio e velocidades, todavia essa mudança não ofereceu impacto significativo no resultado da análise.

Waveforms ou Gráfico de Ondas, é uma análise gráfica simples, porém muito útil para identificação rápida de alguns padrões e ruídos. É composto por métricas de amplitude em decorrência do tempo, que permite a compreensão de características do sinal de áudio, como picos de volume, transições, duração, padrões e outros aspectos relevantes (BARBACENA, 2010).

Espectrograma é um gráfico capaz de mostrar dinamicamente a forma espectral de energia por meio da aplicação da transformada de Fourier de curta duração do sinal (STFT), que realiza a divisão em tempo e analisa a distribuição de energias em frequências variadas. Dessa forma, é um elemento visual importante para área de processamento do áudio, engenharia de áudio e análises de áudio em geral, permite identificar características do sinal, frequências variadas, padrões temporais e sobreposições de sons. Os eixos são divididos em tempo (horizontal), frequência (vertical) e amplitude representada por cores que variam de acordo com cada literatura, mas sempre como uma representação visual de intensidade que é importante para estudar e compreender diferentes tipos de sinais e linearidade expressiva de um áudio (BARBACENA, 2010).

Na aplicação Speech Snake, a STFT é realizada de forma automatizada através de funções da biblioteca Librosa.

Frequência Fundamental (F0) é um gráfico que demonstra a variação da frequência de sinal mais alto ou baixo ao longo de um áudio. A criação do gráfico é feita com uso do método YIN baseado em autocorrelação, presente na documentação Librosa, apoiada no artigo “YIN, um estimador de frequência fundamental para fala e música" por Alain de Cheveigné e Hideki Kawahara, publicado no *Journal of the Acoustical Society of America* em 2002.

2.9 Bibliotecas Utilizadas

Para o desenvolvimento do *back-end* da aplicação foi utilizada a linguagem de programação Python em conjunto com Javascript para o *front-end*.

As bibliotecas utilizadas para a análise de dados em Python podem ser divididas em duas categorias, áudio e matemática/estatística.

Bibliotecas são pacotes de recursos utilizados para determinada função, conforme definições a seguir:

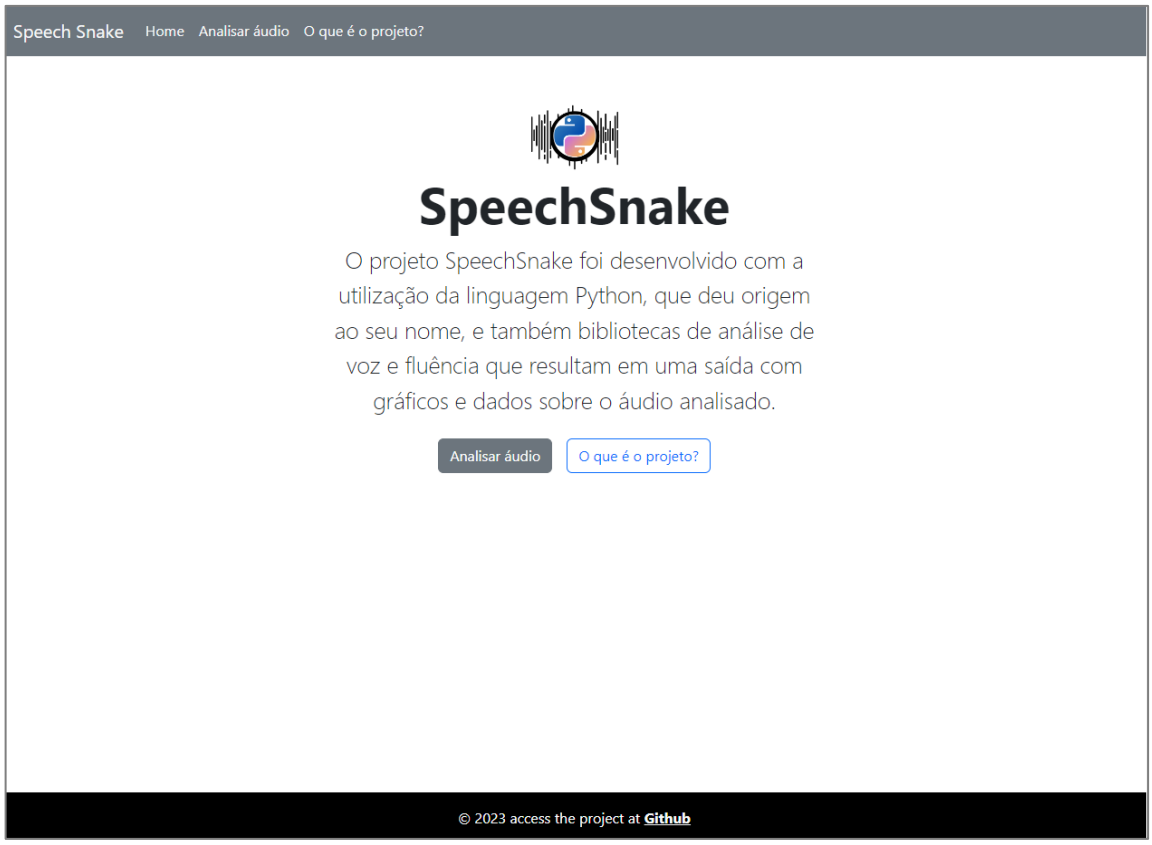
- Librosa: Biblioteca em Python usada para análise de sinal de música e áudio que fornece as estruturas necessárias para a transformação e manipulação de dados de áudio (LIBROSA, 2015).
- Parselmouth: Biblioteca em Python para trabalhar com o *software* Praat (utilizado para análise acústica) através de uma API para estudos da fala e Processamento de Linguagem Natural (PLN) de maneira eficiente (PARSELMOUTH, 2022).
- Pydub: É uma biblioteca focada na edição de áudios criada para simplificar o processamento desde a importação até a manipulação e edição. Seu foco está na facilidade de utilização em comparação às demais. (PYDUB, 2021).
- Soundfile: Também é uma biblioteca para o processamento e análise de dados, mas em um nível mais baixo. Seu principal diferencial está na facilidade de acesso e controle dos metadados gerados (SOUNDFILE, 2023).
- Matplotlib: Biblioteca de visualização de dados em Python que possui ampla variedade de funções e ferramentas para criação de diversos gráficos de modo estático, animado e interativo, com a possibilidade de incorporar interfaces a partir de outras bibliotecas (MATPLOTLIB, 2023).
- Numpy: Biblioteca matemática e estatística com alto poder de processamento para realizar operações com matrizes, *arrays* e cálculos complexos de forma automatizada e otimizada (NUMPY, 2023).
- BackBlaze: Oferece soluções de backup e armazenamento de dados em nuvem de forma confiável, possui compatibilidade com diversas integrações e ferramentas de desenvolvimento (BACKBLAZE, 2023).

- b2sdk: SDK (*Software Development Kit*, em português: kit de desenvolvimento de *software*) em Python utilizado para acessar os recursos do Backblaze B2 *Cloud Storage* por meio de interface simples e intuitiva com garantia de proteção e integridade dos dados armazenados em nuvem (B2-SDK-PYTHON, 2020).
- Flask: Microframework em Python para o desenvolvimento de aplicativos web de forma simples. No microframework Flask o “micro” é devido ao núcleo simples, porém extensível, que permite a adição de funcionalidades em uma aplicação para integração de banco de dados, validação de formulário, manipulação de upload, tecnologias de autenticação, etc (FLASK, 2013).
- Bootstrap: Framework de código aberto para desenvolvimento web amplamente utilizado devido a documentação abrangente e suporte multiplataforma que facilitam a criação de sites. Fornece componentes, estilos e temas personalizáveis para criação de interfaces web responsivas (*front-end*) (BOOTSTRAP, 2023).

3 APLICAÇÃO SPEECH SNAKE

Com o uso das técnicas, ferramentas tecnológicas e pesquisas realizadas neste trabalho de graduação, foi desenvolvida a aplicação Speech Snake. O fluxo do programa consiste nas seguintes telas com suas respectivas funcionalidades:

Figura 3. Tela inicial da aplicação - Home

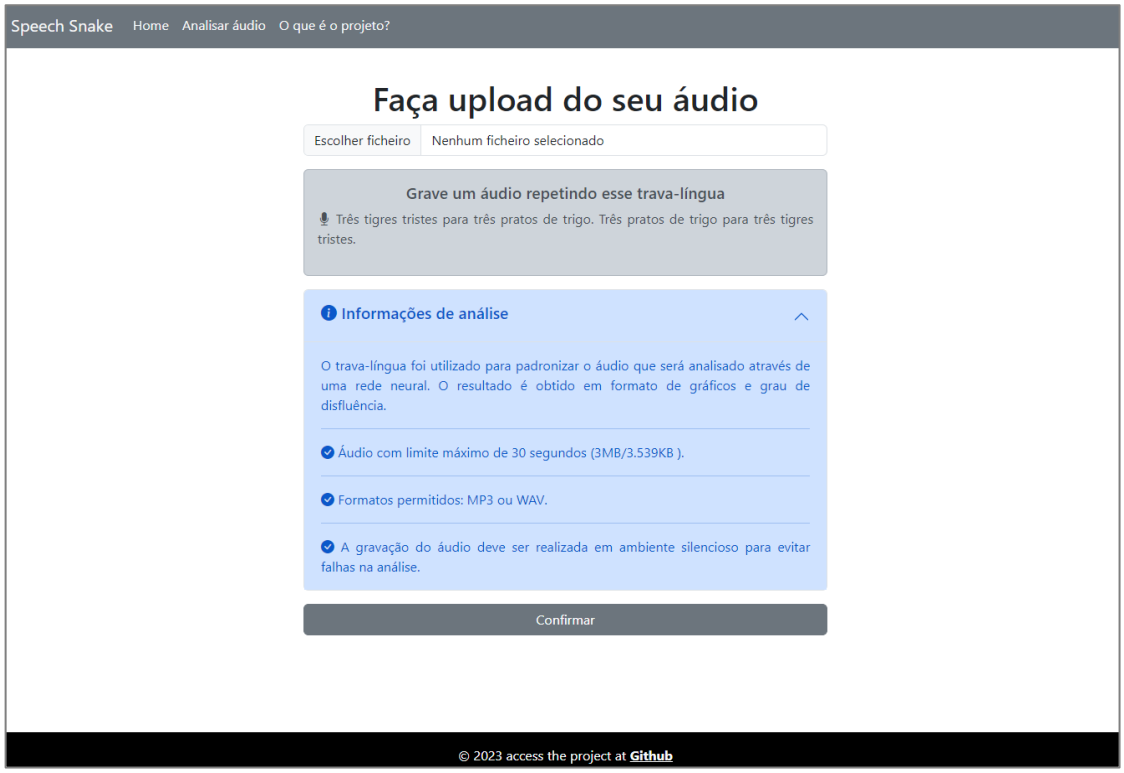


Fonte: SPEECH SNAKE (2023)

Upload de Áudio (Analisar áudio): Nessa tela, o usuário pode carregar um áudio com duração máxima de 30 segundos (3MB/3.539KB de tamanho) após a gravação do trava-línguas estipulado. O objetivo é permitir que o usuário envie o arquivo de áudio gravado em formato MP3 ou WAV.

Figura 4

Figura 4. Tela de upload do áudio - Analisar áudio



Fonte: SPEECH SNAKE (2023)

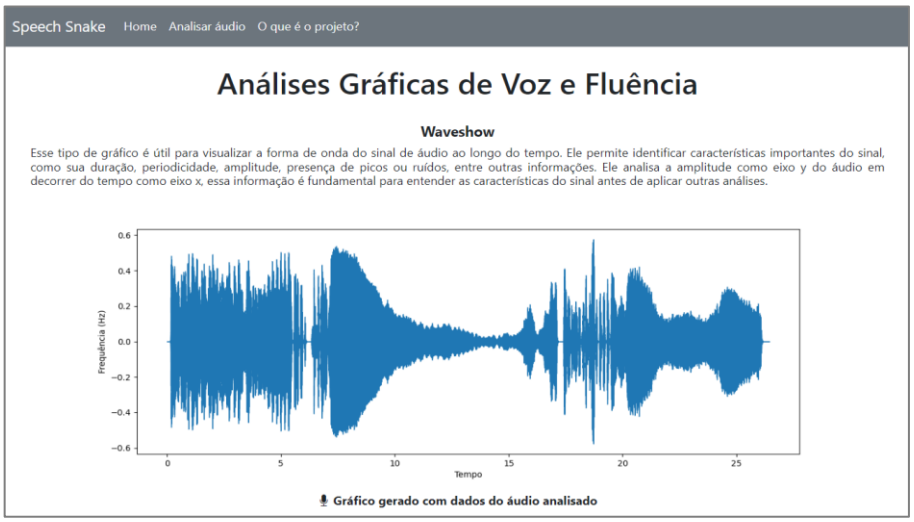
Análise do Áudio: O áudio enviado é coletado dentro do container de armazenamento de arquivo com uso do identificador disponível na *BackBlaze* (armazenamento em nuvem). Em seguida, é realizada a análise de Frequência Formante que utiliza o algoritmo disponível na biblioteca Parselmouth. Com apoio das parametrizações de Tafner (1996), que consideram o valor de F1 na faixa de 300Hz a 900Hz e os valores de F2 na faixa de 800Hz a 2800Hz, a classificação do áudio é definida como "Voz Clara" ou "Voz Confusa".

Figura 5. Tela de análise do áudio

Fonte: SPEECH SNAKE (2023)

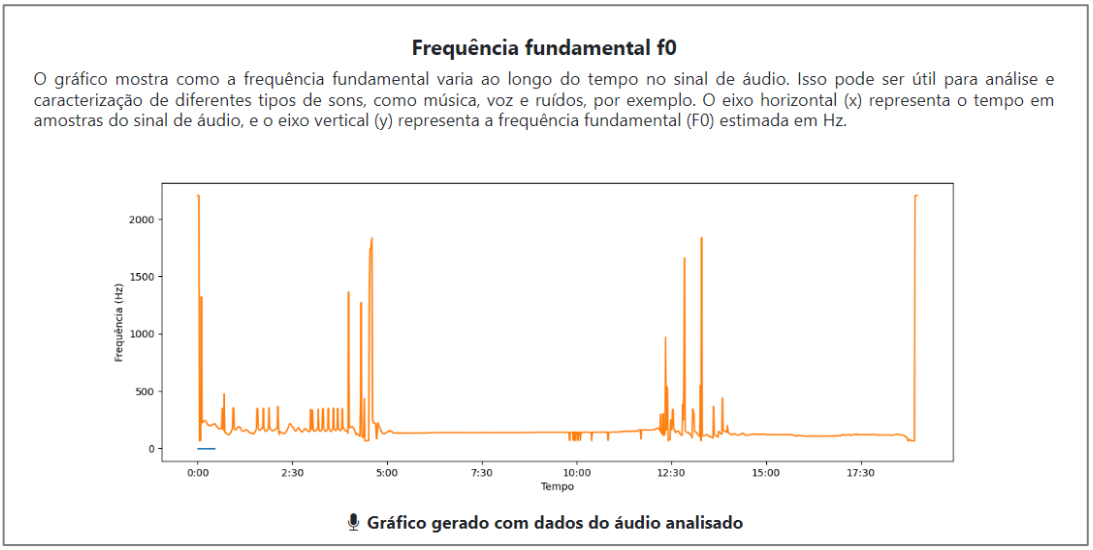
Visualização dos Gráficos: Nessa etapa são gerados os gráficos de acordo com a rotina de visualização, a partir do uso da biblioteca Librosa. A exibição é realizada por meio da Matplotlib. Os gráficos resultantes são: Waveshow, Espectrograma e Frequência Fundamental (F0).

Figura 6. Tela de análise do áudio - Gráfico Waveshow



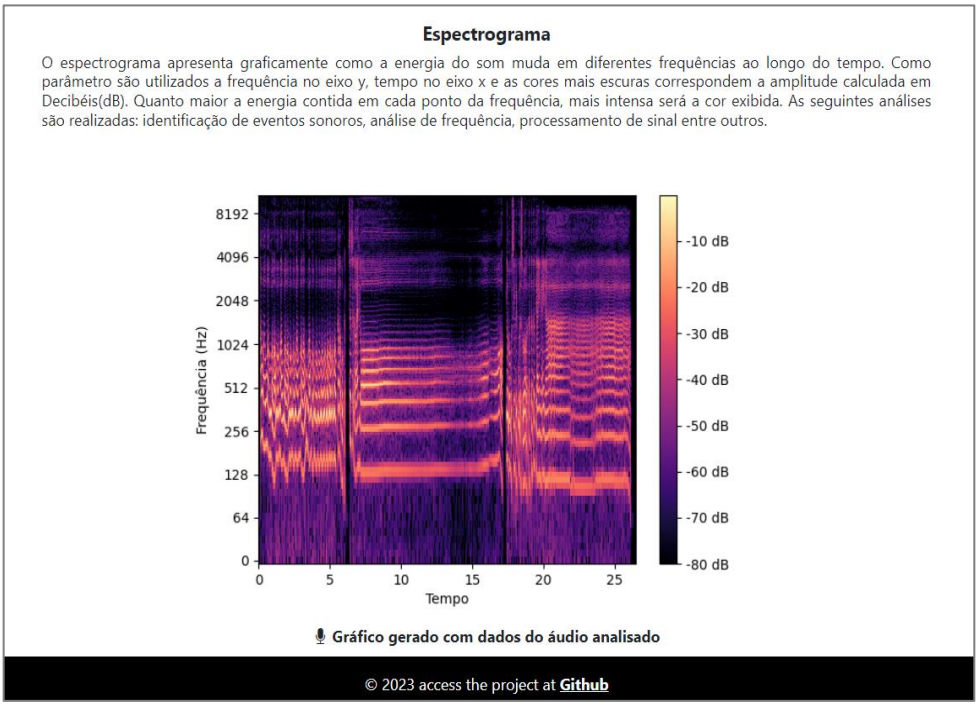
Fonte: SPEECH SNAKE (2023)

Figura 7. Tela de análise do áudio - Gráfico Frequência Fundamental (F0)



Fonte: SPEECH SNAKE (2023)

Figura 8. Tela de análise do áudio - Gráfico Espectrograma



Fonte: SPEECH SNAKE (2023)

A aplicação web também possui uma tela com explicações sobre o projeto e métricas extraídas de áudio que devem ser consideradas para compreensão das análises.

Figura 9. Tela sobre o projeto - O que é o projeto?

Fonte: SPEECH SNAKE (2023)

O fluxo possibilita que o usuário carregue o arquivo de áudio, obtenha a análise de Frequência Formante e visualize os gráficos gerados segundo as métricas extraídas no processamento do áudio enviado, assim como, a definição de "Voz Clara" ou "Voz Confusa".

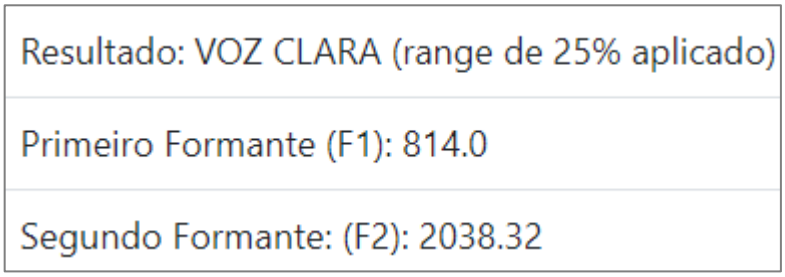
4 ANÁLISES E RESULTADOS

A aplicação web Speech Snake possibilita a realização de análises de formantes e exibição gráfica dos detalhes de um áudio processado a partir de um trava-línguas parametrizado pela equipe de pesquisa.

Os testes realizados alcançaram os resultados seguintes.

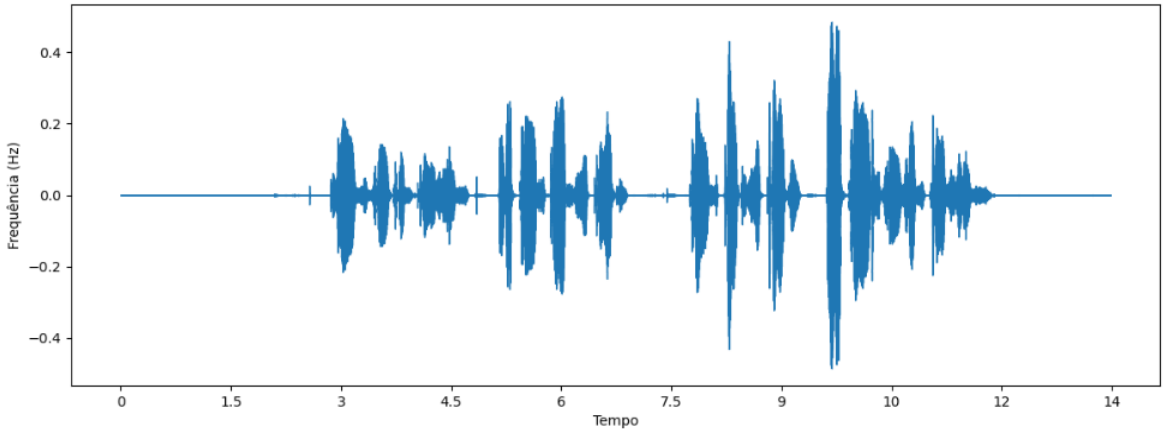
4.1 Teste 1: Áudio Com Velocidade Normal

Figura 10. Resultado das Formantes



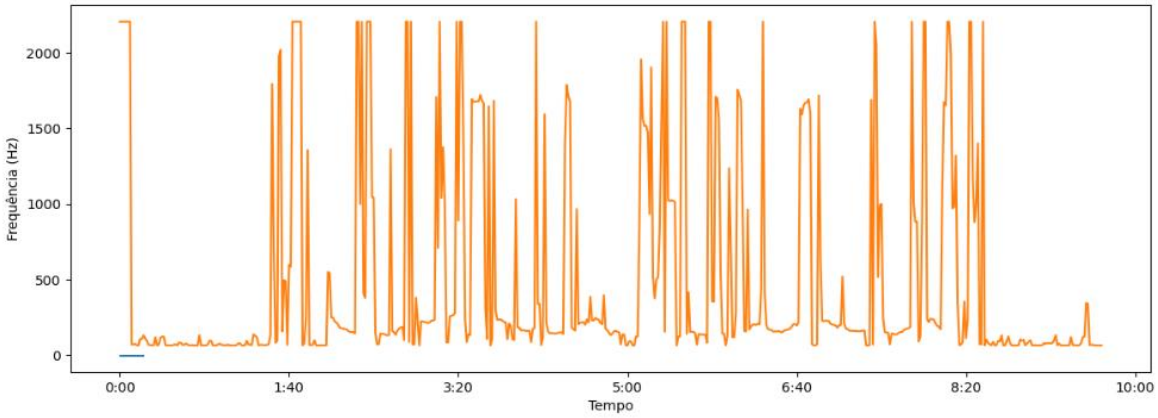
Fonte: SPEECH SNAKE (2023)

Figura 11. Resultado do gráfico Waveshow



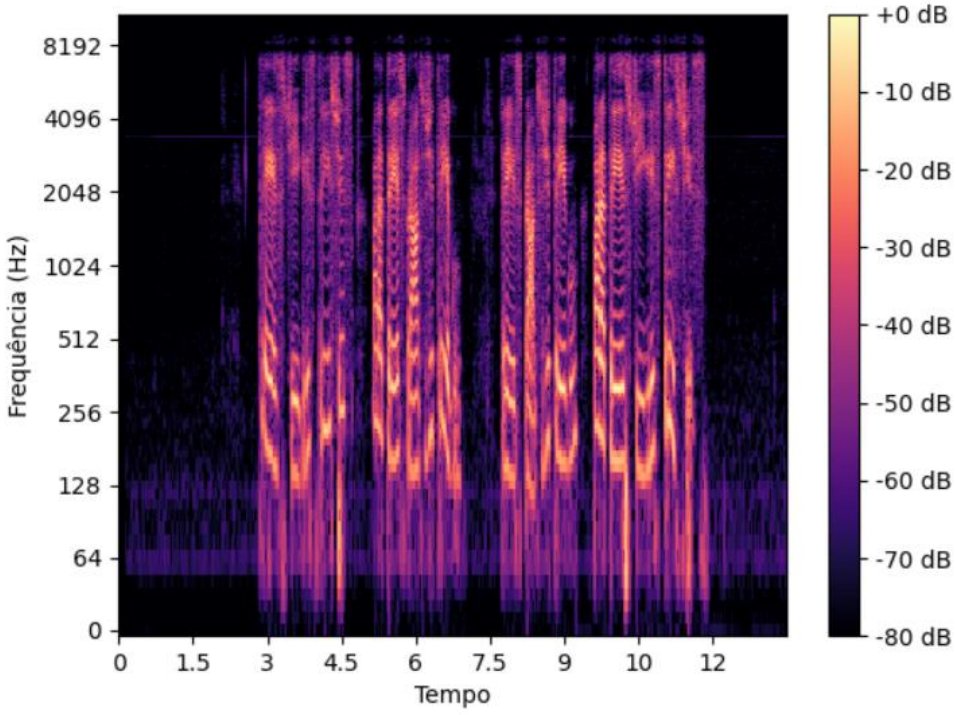
Fonte: SPEECH SNAKE (2023)

Figura 12. Resultado do gráfico de Frequência Fundamental (F0)



Fonte: SPEECH SNAKE (2023)

Figura 13. Resultado do gráfico Espectrograma



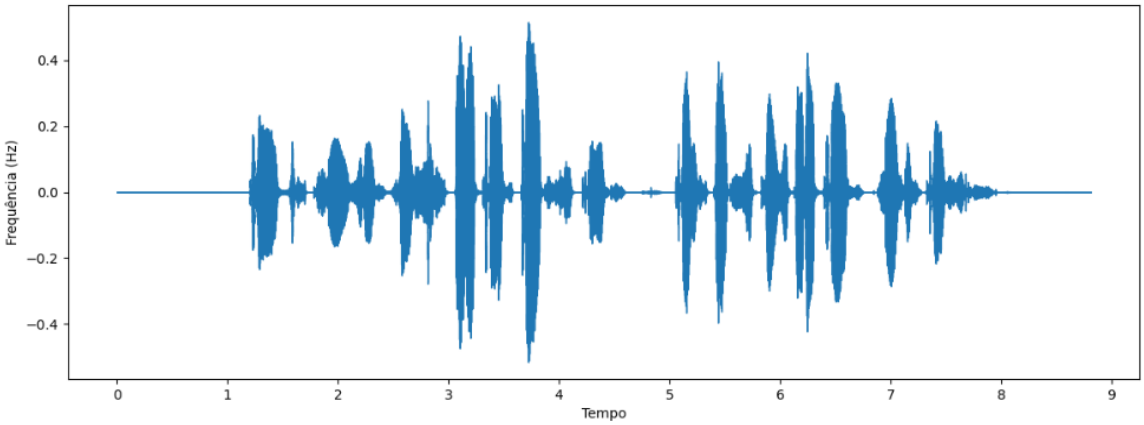
Fonte: SPEECH SNAKE (2023)

4.2 Teste 2: Áudio Com Velocidade Rápida

Figura 14. Resultado das Formantes

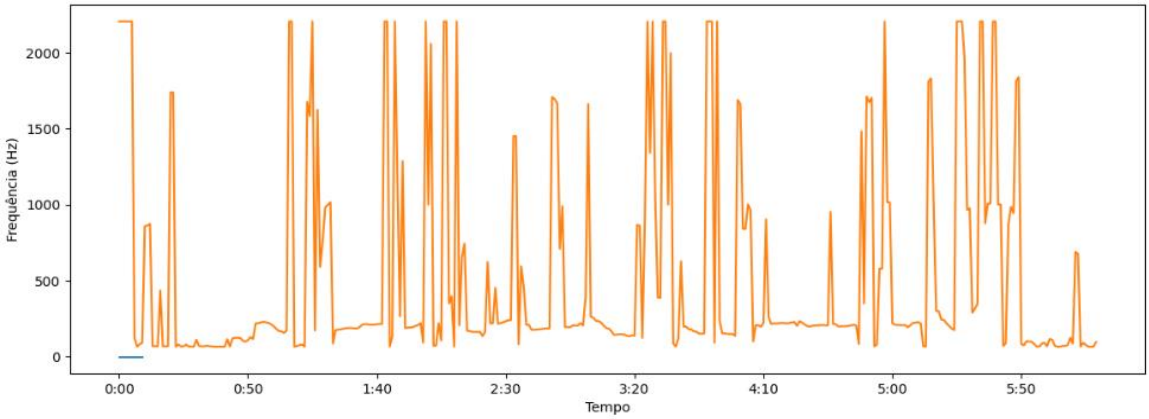
Fonte: SPEECH SNAKE (2023)

Figura 15. Resultado do gráfico Waveshow



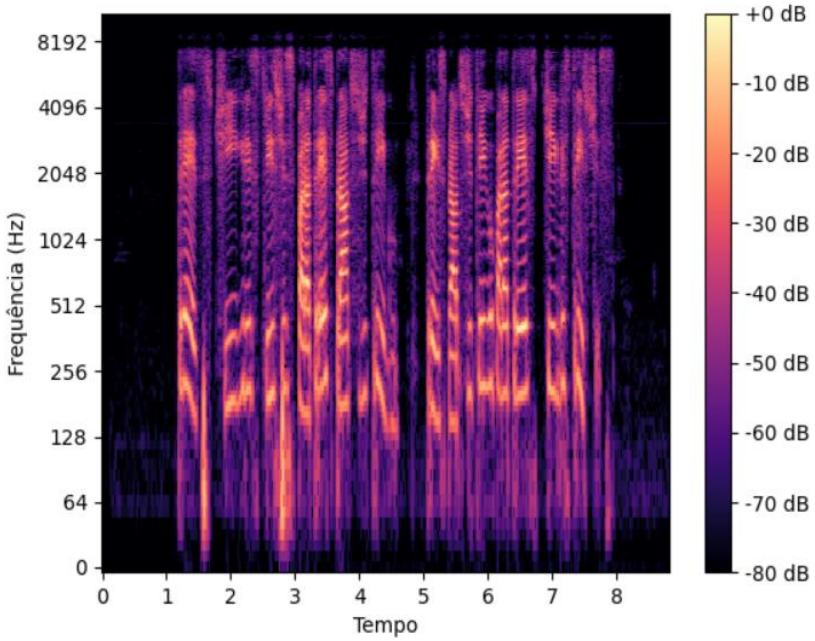
Fonte: SPEECH SNAKE (2023)

Figura 16. Resultado do gráfico de Frequência Fundamental (F0)



Fonte: SPEECH SNAKE (2023)

Figura 17. Resultado do gráfico Espectrograma



Fonte: SPEECH SNAKE (2023)

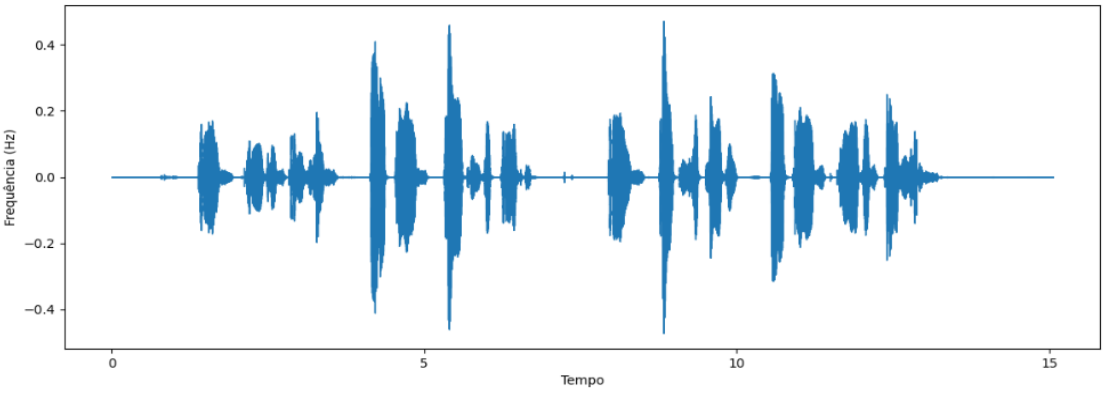
4.3 Teste 3: Áudio Com Velocidade Lenta

Figura 18. Resultado das Formantes

Resultado: VOZ CLARA (range de 25% aplicado)
Primeiro Formante (F1): 815.49
Segundo Formante: (F2): 2040.35

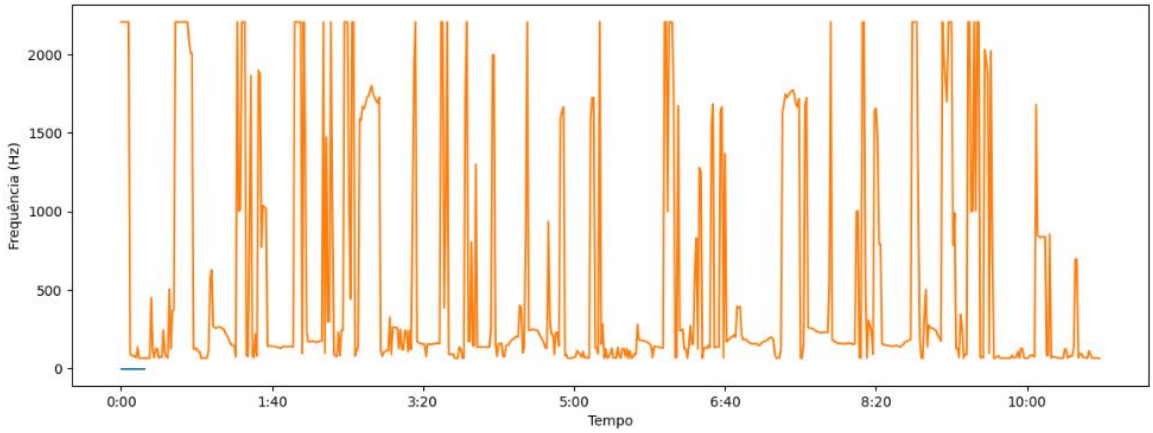
Fonte: SPEECH SNAKE (2023)

Figura 19. Resultado do gráfico Waveshow



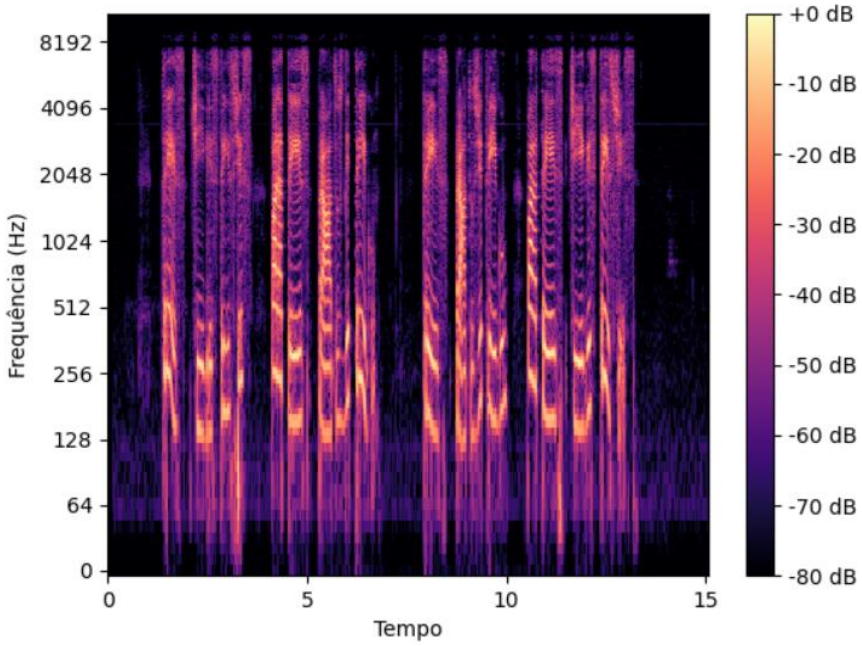
Fonte: SPEECH SNAKE (2023)

Figura 20. Resultado do gráfico de Frequência Fundamental (F0)



Fonte: SPEECH SNAKE (2023)

Figura 21. Resultado do gráfico Espectrograma



Fonte: SPEECH SNAKE (2023)

Com base nos testes realizados é possível observar uma clara alteração nos resultados dos formantes apresentados nos gráficos com relação ao processamento dos diferentes tipos de áudios, classificados como: lento, normal e rápido. Essas variações podem ser detalhadas através da análise desses elementos.

5 CONSIDERAÇÕES FINAIS

Durante o desenvolvimento do projeto, encontramos dificuldades para definir o público-alvo devido à complexidade da proposta. No entanto, ao realizarmos o levantamento da pesquisa, constatamos que existem poucas aplicações que realizam a análise dados de áudios com representações gráficas e textos explicativos para estudos. Diante disso, consideramos a ideia de criar um projeto que oferecesse uma interface amigável, dados relevantes, descrições detalhadas e que fosse *open-source*, a fim de contribuir com áreas que demonstrassem interesse no tema.

Com o desafio, notamos a necessidade de preencher uma lacuna no mercado para fornecer uma solução que atenda às demandas de pesquisadores que desejam explorar o uso de áudios em conjunto com elementos visuais e textuais. Acreditamos que essa abordagem pode ser inovadora e facilitará o aprendizado de conceitos complexos, além de ampliar as possibilidades de exploração sobre o assunto.

A aplicação web Speech Snake pode apresentar custos de grande escala devido ao armazenamento dos áudios em nuvem, então optamos por realizar somente testes em produção para que a disponibilidade da aplicação esteja adaptativa para possível hospedagem futura.

Portanto, acreditamos o desenvolvimento de uma solução com uma interface intuitiva, com análises de dados relevantes, descrições pertinentes e de natureza *open-source* irá atender a necessidade das comunidades interessadas nesse tipo de abordagem e contribuir com a propagação de conhecimento.

REFERÊNCIAS

AWS. **O que é conversão de fala em texto?** Guia de transcrição para iniciantes. [*s.l.*]: 2022. Disponível em: <https://aws.amazon.com/pt/what-is/speech-to-text/>. Acesso em: 9 nov. 2022.

BACKBLAZE. **B2 Cloud Storage: The Lowest Cost On Demand Storage As a Service.** [*s.l.*]: 2023. Disponível em: <https://www.backblaze.com/b2/cloud-storage.html>. Acesso em: 17 Mai. 2023.

BARBACENA, I. L. **MELHORIA DA QUALIDADE DA VOZ DE DEFICIENTES AUDITIVOS UTILIZANDO-SE CORREÇÃO DA FREQUÊNCIA FUNDAMENTAL**. Tese de Doutorado. Universidade Federal de Campina Grande. Campo Grande, 2010. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/xmlui/bitstream/handle/riufcg/3250/ILTON%20LUIZ%20BARBACENA%20-%20TESE%20PPGEE%202010.pdf?sequence=3&isAllowed=y>. Acesso em: 02 jun. 2023.

BEHLAU, M. et al. **Voz: o livro do especialista.** Avaliação de voz. Rio de Janeiro: Revinter, 2004. p. 56-60.

BOOTSTRAP. **Documentação Bootstrap.** São Francisco, 2023. Disponível em: <https://getbootstrap.com/>. Acesso em: 28 abr. 2023.

B2-SDK-PYTHON. **Quick Start Guide.** [*s.l.*]: 2020. Disponível em: <https://b2-sdk-python.readthedocs.io/en/master/quick_start.html#copy-file>. Acesso em: 17 Mai. 2023.

CHEVEIGNÉ, A.; KAWAHARA, H. **YIN, um estimador de frequência fundamental para fala e música.** Journal of the Acoustical Society of America. [*s.l.*]: 2002. Disponível em: <https://pubs.aip.org/asa/jasa/article-abstract/111/4/1917/547221/YIN-a-fundamental-frequency-estimator-for-speech?redirectedFrom=fulltext>. Acesso em: 01 Jun 2023.

DEISE, M. et al. **MODULAÇÃO POR CÓDIGO DE PULSO PCM.** Santa Catarina, 2016. Disponível em: <https://wiki.sj.ifsc.edu.br/wiki/images/3/34/PCM.pdf>. Acesso em: 9 nov. 2022.

FERREIRA, R. **Contents of Análise LPC - Linear Predictive Coding**. [*s.l.*]: 2003. Disponível em: <http://users.isr.ist.utl.pt/~ricardo/GSM/node6_ct.html#:~:text=Linear%20Predictive%20Coding%20%C3%A9%20uma%20maneira%20simples%2C%20r%C3%A1pida>. Acesso em: 9 nov. 2022.

FLASK. **Documentação Flask.** [*s.l.*]: 2013. Disponível em: <https://flask-ptbr.readthedocs.io/en/latest/index.html>. Acesso em: 28 abr. 2023.

IBM. **Cloud formation: O que é reconhecimento de fala?** [*s.l.*]: 2022. Disponível em: <https://www.ibm.com/br-pt/cloud/learn/speech-recognition#toc-o-que--rec-SMJbcsv>. Acesso em: 9 nov. 2022.

KERSCHBAUMER, R. **Sistemas Especialistas**. Santa Catarina, 2018. Disponível em: <https://professor.luzerna.ifc.edu.br/ricardo-kerschbaumer/wp-content/uploads/sites/43/2018/02/3-Sistemas-Especialistas.pdf>. Acesso em: 9 nov. 2022.

LIBROSA. **Documentação Librosa.** Austin Texas: SciPy, 2015. Disponível em: <https://librosa.org/doc/latest/index.html>. Acesso em: 28 abr. 2023.

LIPPMANN, R. P. **Speech recognition by machines and humans.** Speech Communication, v. 22, n. 1, p. 1-15, jul. 1997.

LÓSS, J. C. S.; SILVA, L. P.; CABRAL, H. L. T. B.; LIMA, W. L. F. **Distúrbios que afetam a linguagem.** Interfaces da Linguagem. p. 220. Campos dos Goytacazes, RJ: Brasil Multicultural, 2020. Disponível em: <http://brasilmulticultural.org/wp-content/uploads/2021/06/ebook-Interfaces-da-linguagem-1.pdf#page=220>. Acesso em: 01 Out.2022.

MATPLOTLIB. **Matplotlib: Visualization with Python.** [*s.l.*]: 2023. Disponível em: <https://matplotlib.org/>. Acesso em: 01 Mai. 2023.

MORETI, F.; ZAMBON, F; BEHLAU, M. **Sintomas vocais e autoavaliação do desvio vocal em diferentes tipos de disfonia.** Departamento de Fonoaudiologia, Universidade Federal de São Paulo – UNIFESP. São Paulo, 2014. Disponível em: < https://www.scielo.br/j/codas/a/SKpZhSB6bXf74LkCCLH4kdh/?format=pdf&lang=pt>. Acesso em: 05 Out.2022.

NOLL, V. **O português brasileiro: formação e contrastes.** São Paulo: Globo, 2008. 399 p.

PARSELMOUTH. **Parselmouth – Praat in Python, the Pythonic way.** [*s.l.*]: 2022. Disponível em: <https://parselmouth.readthedocs.io/en/stable/>. Acesso em: 17 Mai. 2023.

PIERRI, R. **Como é que o som de uma música é convertido em sinal digital e armazenado na memória do computador?** Nets&Nuts: Eletrônica, programação e variedades. [*s.l.*]: 2019. Disponível em: <https://nets-nuts.com.br/como-e-que-o-som-de-uma-musica-e-convertido-em-sinal-digital-e-armazenado-na-memoria-do-computador/#:~:text=Como%20%C3%A9%20que%20o%20som%20de%20uma%20m%C3%BAsica>. Acesso em: 9 nov. 2022.

RIBEIRO, R. **A importância da neurociência para a Fonoaudiologia.** Projetando Neurociência. [*s.l.*]: 2020. Disponível em: <https://projetandoneurociencia.org/a-importancia-da-neurociencia-para-a-fonoaudiologia/>. Acesso em: 9 nov. 2022.

SANTOS, SUELAINE; THOMÉ, ANTONIO (1998). **Uso de técnicas Neurais para o reconhecimento de comandos de voz.** [s.l: s.n.]. Disponível em: <https://rmct.ime.eb.br/arquivos/RMCT_1_tri_1998/uso_tec_neurais_recon_cmd_voz.pdf>. Acesso em: 2 jun. 2023.

SENAI-SP EDITORA. **Eletrônica digital.** [*s.l.*]: SESI SENAI Editora, 2018. p. 9-26.

SOUZA, R. L.; CARDOSO, M. C. A. F. **Fluência e Prosódia: Aspectos Diferenciais Frente aos Distúrbios.** Centro Universitário Metodista do IPA de Porto Alegre, RS, 2013. Disponível em: <https://periodicos.unifesp.br/index.php/neurociencias/article/view/8166/5698>. Acesso em: 15 Out.2022.

TAFNER. A. M. **Reconhecimento de palavras faladas isoladas usando redes neurais**. Dissertação. Universidade Federal de Santa Catarina. Santa Catarina, 1996. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/158028>. Acesso em: 01 Jun. 2023.

TRAVAGLIA, L. C. **Gramática e interação: uma proposta para o ensino de gramática**. 14 ed. São Paulo: Cortez, 2009.

USP. Fonoaudiologia e Medicina. **Formantes**. Universidade de São Paulo. São Paulo, 2023. Disponível em: <https://voz.fob.usp.br/pt/formantes/>. Acesso em: 01 Jun. 2023.

APÊNDICE A – Link Speech Snake

SPEECH SNAKE. São Paulo, 2023. Disponível em: < https://speech-snake.herokuapp.com/>.

APÊNDICE B – Áudios dos Testes

SPEECH SNAKE. **Áudio Normal.** São Paulo, 2023. Disponível em: <https://speechsnakeaudios.s3.us-east-005.backblazeb2.com/Normal.wav>.

SPEECH SNAKE. **Áudio Rápido.** São Paulo, 2023. Disponível em: <https://speechsnakeaudios.s3.us-east-005.backblazeb2.com/Rapido.wav>.

SPEECH SNAKE. **Áudio Lento.** São Paulo, 2023. Disponível em: <https://speechsnakeaudios.s3.us-east-005.backblazeb2.com/Lento.wav>.