



UNIVERSIDADE FEDERAL DE OURO PRETO
ESCOLA DE MINAS
COLEGIADO DO CURSO DE ENGENHARIA DE CONTROLE E
AUTOMAÇÃO - CEAU



JOSÉ LUCAS MENDES PEREIRA JÚNIOR

**DESENVOLVIMENTO DE UM *SOFTWARE WEB* PARA
PROCESSAMENTO E ANÁLISE ACÚSTICA DA VOZ**

**MONOGRAFIA DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E
AUTOMAÇÃO**

Ouro Preto, 2022

JOSÉ LUCAS MENDES PEREIRA JÚNIOR

**DESENVOLVIMENTO DE UM *SOFTWARE WEB* PARA
PROCESSAMENTO E ANÁLISE ACÚSTICA DA VOZ**

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheiro de Controle e Automação.

Orientador: Adrielle C. Santana, Dr.Sc.

Coorientador: Carla A. Vasconcelos, Dr.Sc.

**Ouro Preto
Escola de Minas – UFOP
2022**

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

P436d Pereira Junior, Jose Lucas Mendes.

Desenvolvimento de um software web para processamento e análise acústica da voz. [manuscrito] / Jose Lucas Mendes Pereira Junior. - 2022. 38 f.: il.: color..

Orientadora: Profa. Dra. Adrielle de Carvalho Santana.

Coorientadora: Dra. Carla Aparecida de Vasconcelos.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Escola de Minas. Graduação em Engenharia de Controle e Automação .

1. Processamento de sinais biomédicos. 2. Fourier, Transformadas de. 3. Sistemas de processamento da fala. 4. Acústica - Análise. I. Santana, Adrielle de Carvalho. II. Vasconcelos, Carla Aparecida de. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 681.5

Bibliotecário(a) Responsável: Maristela Sanches Lima Mesquita - CRB-1716



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA CONTROLE E
AUTOMACAO



FOLHA DE APROVAÇÃO

José Lucas Mendes Pereira Júnior

Desenvolvimento de um Software Web para Processamento e Análise Acústica da Voz

Monografia apresentada ao Curso de Engenharia de Controle e Automação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de bacharel em Engenharia de Controle e Automação

Aprovada em 17 de Junho de 2022

Membros da banca

- Profa. Dra. Adrielle de Carvalho Santana – Orientadora (Departamento de Engenharia de Controle e Automação, Universidade Federal de Ouro Preto)
Dra. Carla Aparecida de Vasconcelos - Coorientadora (Superintendência Central de Perícia Médica e Saúde Ocupacional, SEPLAG - MG)
Prof. Dr. Danny Augusto Vieira Tonidandel - Convidado (Departamento de Engenharia de Controle e Automação, Universidade Federal de Ouro Preto)
Prof. MSc. Karen Maria de Paula - Convidada (Centro Universitário Planalto - UNIPLAN, Distrito Federal)

Adrielle de Carvalho Santana, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 20/06/2022



Documento assinado eletronicamente por **Adrielle de Carvalho Santana, PROFESSOR DE MAGISTERIO SUPERIOR**, em 20/06/2022, às 11:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0346378** e o código CRC **2017C7D4**.

Dedico este trabalho à minha filha Maria Catarina, que embora ainda tão pequena, já é a maior razão de todo o meu esforço e dedicação.

AGRADECIMENTOS

Agradeço a Deus, que em sua infinita bondade me concedeu a graça deste momento tão importante em minha vida e a Virgem Santíssima, Mãe de Deus pela sua poderosíssima intercessão. Agradeço aos meus familiares, meus pais, José Lucas e Nice, a minhas irmãs Eduarda e Cecília, que desde o início da minha graduação me apoiaram e me deram total suporte. Aos amigos de UFOP, aos professores e em especial a turma 16.2, por me acolherem tão bem em Ouro Preto, pelo apoio constante e por fazerem desta experiência tão agradável. A orientadora, Profa. Adrielle e a coorientadora Carla, pela oportunidade, apoio, paciência e por contribuir tão efetivamente para que fosse possível a conclusão desta monografia. Finalmente, agradeço a minha esposa Adria, que sempre esteve do meu lado, me dando todo o suporte necessário, me motivando, encorajando e me ajudando a cultivar o melhor de mim.

“Quem acende uma luz é o primeiro a se beneficiar da claridade.” (G. K. Chesterton)

RESUMO

O processamento e análise acústica da voz humana é uma atividade importante tanto para a identificação de disfunções vocais quanto para realização de perícia trabalhista e criminal. Nesse sentido, falta hoje para os profissionais um software intuitivo, com interface amigável e com a apresentação de um espectrograma de boa qualidade para a análise visual. Assim, este projeto propõe o desenvolvimento de um *software* acessível, de baixo custo, para processamento e análise acústica de voz que utilize técnicas de processamento de sinais comuns na engenharia biomédica tais como a transformada de Fourier. O desenvolvimento deste projeto deu início com a identificação dos requisitos da aplicação com o auxílio de um profissional da área de análise acústica da voz. Em seguida foram levantadas as técnicas de processamento de sinais biomédicos necessárias para implementar as funcionalidades desejadas. Com isso, foram avaliadas algumas opções de linguagens e bibliotecas para desenvolvimento de *software* tanto *back-end* e *front-end*, onde foi escolhido o Streamlit, *framework* de código aberto que utiliza linguagem Python. O resultado prático foi a construção de uma aplicação capaz de receber e gravar áudios para análise e apresentar como resultado principal um espectrograma para auxiliar na análise acústica da voz, além da possibilidade de visualização do detector de pitch. Foram realizados alguns testes das funcionalidades da aplicação além da resolução e manipulação das visualizações, utilizando vozes femininas e masculinas. Os gráficos gerados possuem uma resolução satisfatória, podendo apresentar uma performance ainda maior para áudios gravados em ambientes com poucos ou nenhum ruído.

Palavras-chaves: Processamento de sinais biomédicos. Transformada de Fourier. Processamento de fala. Análise acústica.

ABSTRACT

The processing and acoustic analysis of the human voice is an important activity both for the identification of vocal dysfunctions and for carrying out labor and criminal expertise. In this sense, professionals today lack an intuitive software, with a friendly interface and with the presentation of a good quality spectrogram for visual analysis. Thus, this project proposes the development of an accessible, low-cost *software* for voice acoustic processing and analysis that uses signal processing techniques common in biomedical engineering such as the Fourier transform. The development of this project began with the identification of application requirements with the help of a professional in the field of voice acoustic analysis. Then, the biomedical signal processing techniques necessary to implement the desired functionalities were raised. With that, some options of languages and libraries were evaluated for the development of *software* both *back-end* and *front-end*, where Streamlit was chosen, *framework* of open source which uses Python language. The practical result was the construction of an application capable of receiving and recording audios for analysis and presenting as main result a spectrogram to assist in the acoustic analysis of the voice, in addition to the possibility of viewing the pitch detector. Some tests of the application's functionalities were carried out, in addition to the resolution and manipulation of the visualizations, using female and male voices. The generated graphics have a satisfactory resolution, being able to present an even greater performance for audios recorded in environments with little or no noise.

Key-words: Biomedical signal processing. Fourier transform. speech processing and acoustic analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Espectrograma da sílaba /da/	14
Figura 2 – Representação de um sinal no domínio no tempo e sua respectiva transformada de fourier	15
Figura 3 – STFT com diferentes resoluções de tempo-frequência	17
Figura 4 – Importando bibliotecas	19
Figura 5 – Esboço da aplicação	20
Figura 6 – Trecho Exemplo de implementação - Sidebar Menu	20
Figura 7 – Funções para geração do espectrograma e pitch	21
Figura 8 – Gravação do arquivo de áudio com duração pré-definida	22
Figura 9 – Utilizando Parselmouth	22
Figura 10 – Tela inicial Black mode	24
Figura 11 – Tela inicial Ligth mode	25
Figura 12 – Tela inicial - Configurações de estilo	25
Figura 13 – Espectrograma vogal sustentada A Voz masculina - comp. Janela 0.025 s	26
Figura 14 – Espectrograma vogal sustentada A Voz masculina - ampliado	26
Figura 15 – Espectrograma vogal sustentada A Voz feminina	27
Figura 16 – Espectrograma vogal sustentada A Voz feminina - ampliado	27
Figura 17 – Espectrograma vogal sustentada E Voz masculina - comp. Janela 0.03s	28
Figura 18 – Espectrograma vogal sustentada E Voz masculina - comp. Janela 0.03s ampliado	28
Figura 19 – Espectrograma vogal sustentada E Voz feminina	29
Figura 20 – Espectrograma vogal sustentada O Voz masculina comp. Janela 0.012	29
Figura 21 – Espectrograma vogal sustentada O Feminino	30
Figura 22 – Áudio sendo gravado	30
Figura 23 – Áudio gravado - Vogal I Sustentada	31
Figura 24 – Áudio gravado - Vogal A sustentada	31
Figura 25 – Áudio gravado - Vogal I sustentada - Cor do espectrograma: afmhot	32

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Contextualização	9
1.2	Objetivos gerais e específicos	10
1.2.1	<i>Geral</i>	10
1.2.2	<i>Específicos</i>	10
1.3	Justificativa do trabalho	10
1.4	Estrutura do texto	10
2	REVISÃO DA LITERATURA	12
2.1	Produção da fala humana	12
2.2	Análise acústica da voz	12
2.2.1	<i>Análise espectrográfica da voz</i>	13
2.3	Análise de Fourier	14
3	DESENVOLVIMENTO	18
3.1	Caracterização da área de estudo	18
3.2	Classificação do trabalho	18
3.3	Desenvolvimento da aplicação <i>web</i>	18
3.4	Acesso ao código fonte da aplicação	23
4	RESULTADOS	24
4.1	Apresentação da aplicação	24
4.2	Testes	25
5	CONCLUSÃO	33
6	REFERÊNCIAS	35

1 INTRODUÇÃO

1.1 Contextualização

A voz é, dentre as ferramentas primárias a mais imediata que o ser humano possui para interagir com o seu meio (BEHLAU; AZEVEDO; PONTES, 2010) . Analisando a partir do ponto de vista fisiológico, a voz é produzida por um conjunto de órgãos e músculos. As pregas vocais são duas faixas de tecido elástico localizadas lado a lado na laringe, um pouco acima da traqueia. Quando se está em silêncio, as pregas permanecem abertas, o que cria uma via, e é por meio dessa via que podemos respirar. Quando se emite som, o ar que sai dos pulmões é forçado através das pregas vocais, fazendo com que elas vibrem.

A voz é uma característica própria do indivíduo. A produção da voz está relacionada a fatores culturais e psicossociais, além de biológicos e genéticos. A personalidade, o estado emocional e a forma de expressar as emoções também diferenciam a voz (GOULART; COOPER, 2003).

Avanços nos estudos da voz humana nas últimas décadas foram possíveis com o surgimento de técnicas confiáveis e objetivas que contribuíram para a avaliação dos distúrbios vocais (ZITTA, 2005). Com o advento da tecnologia, a análise vocal acústica utiliza-se de *softwares* que possibilitam mensurar vários aspectos do sinal de voz analisado (VIEIRA; BIASE; PONTES, 2006).

Recentemente esta avaliação tornou-se mais comum no Brasil com estudos abrangentes. Os dados da análise acústica fornecem ao(à) fonoaudiólogo(a) informações que possibilitam o controle da evolução do(a) paciente em fonoterapias, o registro das condições vocais no pré e pós-operatório em cirurgias laríngeas (CÔRTEZ; GAMA, 2010; KELCHNER et al., 2010; HOSOKAWA *et al.* , 2014) e também em processos judiciais, auxiliando na identificação dos Distúrbios Vocais Relacionados ao Trabalho (DVRTs) e na identificação e comparação de falantes na área criminal.

Tem-se no mercado alguns *softwares* que buscam auxiliar os(as) profissionais dessa área a realizar tais análises. No *software* GRAM, por exemplo, é possível visualizar claramente características de uma voz normal, disfonia de grau 2 do tipo áspero-soprosa e disfonia de grau 2 para 3 do tipo rouco-soprosa. Todos estes produtos atendem até certo ponto, no entanto, muitos destes *softwares*, como o GRAM, já são obsoletos, não mais atualizados, sem suporte para novos sistemas operacionais e com poucas funcionalidades embora possuam uma qualidade suficiente para avaliações qualitativas da voz. Outros *softwares* disponíveis hoje para auxílio à análise acústica ou são muito caros e não tão confiáveis ou gratuitos com uma qualidade espectrográfica inadequada para a análise de voz.

Logo, é possível verificar a necessidade do desenvolvimento de um *software* para o processamento e análise acústica da voz, que seja de baixo custo, com interface amigável, compatível com os sistemas operacionais mais recentes e que permita uma excelente qualidade das imagens espectrográficas.

1.2 Objetivos gerais e específicos

1.2.1 Geral

Desenvolvimento de um *software web* para processamento e análise acústica da voz.

1.2.2 Específicos

- Desenvolver uma aplicação *web* de baixo custo e amigável com boa qualidade das imagens espectrográficas para realizar processamento e análise da voz.
- Fornecer na aplicação opção de visualizar a forma de onda da voz e o detector de pitch.
- Realizar teste das funcionalidades da aplicação.

1.3 Justificativa do trabalho

Os *softwares* disponíveis hoje no mercado para realização da análise acústica da voz atendem até certo ponto, pois muitos destes já se tornaram obsoletos, não mais atualizados, sem suporte para novos sistemas operacionais e com poucas funcionalidades embora possuam uma qualidade suficiente para avaliações qualitativas da voz. Além disso, algumas opções de *softwares* disponíveis hoje para auxílio à análise acústica ou são muito caros e não tão confiáveis ou gratuitos com uma qualidade espectrográfica inadequada para a análise de voz.

Este projeto visa então o desenvolvimento de uma aplicação *web* de baixo custo para processamento e análise da voz. Além disso, com a pesquisa contribui-se no processo de divulgação das técnicas de análise acústica da voz e desenvolvimento de aplicações *web*. Espera-se atrair a atenção para o tema e contribuir para novas ideias e métodos que atendam as necessidades desse mercado. Além disso, as conclusões deste trabalho podem servir de base para outros estudos na área.

1.4 Estrutura do texto

No capítulo 1 foi apresentada a contextualização e justificativa do trabalho bem como seus objetivos. No capítulo dois foi realizado uma revisão bibliográfica para conhecimento do funcionamento das técnicas de processamento de sinais necessárias para implementar as funcionalidades desejadas. No capítulo 3 tem-se a descrição do processo de desenvolvimento do software. No capítulo 4 foram apresentados os principais resultados, bem como os testes

realizados. E por fim, no capítulo 5 tem-se as conclusões deste trabalho e as recomendações para projetos futuros.

2 REVISÃO DA LITERATURA

Neste capítulo são abordados os conceitos básicos de produção da voz humana e também são apresentadas técnicas gerais para análise acústica da voz, com ênfase no espectrograma, técnica utilizada no desenvolvimento do software para processamento e análise acústica de voz apresentado nesta monografia.

2.1 Produção da fala humana

Segundo Aggarwal, Gochhayat e Singh (2021), a produção da fala humana ocorre por meio da "vibração das cordas vocais". Os principais componentes envolvidos nesse processo incluem pulmões, traqueia, laringe, cavidade faríngea, cavidade oral e cavidade nasal. O sistema funciona da seguinte forma: o ar expelido dos pulmões passa pela traqueia e entra em contato com as cordas vocais dentro da laringe (este fenômeno também é chamado de excitação de entrada), as cordas vocais vibram.

Para produzir a fala, o corpo humano não dispõe de nenhum órgão específico, mas utiliza um conjunto de diversas estruturas orgânicas em funcionamento coordenado, chamado aparelho fonador (JÚNIOR, 2007). Segundo o mesmo autor, o aparelho fonador funciona da seguinte forma: Os pulmões, brônquios e traqueia são os órgãos respiratórios que produzem a corrente de ar necessária à fonação; a laringe e cordas vocais são responsáveis por produzir a vibração utilizada na fala; e a faringe, boca e fossas nasais funcionam como caixas de ressonância. A cavidade bucal possui vários obstáculos à passagem do ar, que são responsáveis pelos diversos sons da linguagem.

Segundo Goulart e Cooper (2002) a voz é uma característica própria de cada pessoa, uma vez que sua produção está relacionada a fatores biológicos e genéticos, mas também culturais e psicossociais. O autor ainda complementa que até mesmo o estado emocional e a forma de se expressar também influenciam na voz.

2.2 Análise acústica da voz

Um sinal sonoro $s(n)$ é dado pela convolução temporal de um sinal de excitação $e(n)$, que representaria fisicamente as vibrações das cordas vocais; com a resposta ao impulso de um filtro $h(n)$ que representa o aparelho vocal, modulando o sinal excitatório (SANTANA, 2021).

Considerando que $S(\Omega)$, $E(\Omega)$ e $H(\Omega)$ são a resposta em frequência de $s(n)$, $e(n)$ e $h(n)$, respectivamente, o espectro de magnitude deste modelo pode ser dado por (ALCAIM; OLIVEIRA, 2011):

$$|S(\Omega)|_{dB} = 10\log_{10}|S(\Omega)| = 20\log_{10}(|E(\Omega)| \cdot |H(\Omega)|) = |E(\Omega)|_{dB} + |H(\Omega)|_{dB}$$

(2.1)

Segundo Boualem e Boashash (2016), existem duas características acústicas úteis na análise de um sinal de voz sonora: frequência fundamental e formante. A frequência fundamental é geralmente a componente de frequência mais baixa do sinal; representa a frequência de vibração das cordas vocais durante a produção do som. O formante é uma concentração de energia acústica em torno de uma determinada frequência na onda da fala; cada formante corresponde a uma ressonância no trato vocal.

Na produção da fala, são enfatizados os harmônicos do sinal de excitação (frequência fundamental da fala – F0) próximos às frequências ressonantes do trato vocal. Essas ressonâncias são típicas de cada pessoa dando suas características particulares de fala. Os formantes são os harmônicos de frequência fundamental nas ressonâncias naturais do aparelho vocal (HUANG et al., 2001).

Para Bailey e Snowling (2002) a percepção da fala requer a capacidade de determinar a forma espectral, detectar e discriminar modulação de amplitude e modulação de frequência fundamental e espectral, e fazê-lo com resolução temporal que engloba tanto as mudanças relativamente lentas que se estendem por todo um enunciado quanto as relativamente rápidas. Além disso, como a fala raramente é ouvida isolada de outros sons, o ouvinte deve separar o sinal de interesse dos ruídos.

Atualmente, a avaliação fonoaudiológica da voz pode ser feita por meio da avaliação perceptivo-auditiva, ou da análise acústica, que são medidas ou gráficos gerados por programas de computador (VALETIM, 2009). A análise acústica realiza medidas do sinal sonoro vocal, oferecendo informações do ponto de vista objetivo e quantitativo (GAMA, 1997).

Segundo Júnior (1998), as opções mais frequentes de análise acústica do sinal de uma voz disfônica são: medidas de perturbação da frequência fundamental e amplitude da voz, análise espectrográfica da voz; diversas maneiras propostas para determinar a razão sinal-ruído e processamento do sinal com fins de extrair informações específicas.

Nesta monografia é utilizado o método da análise espectrográfica da voz por meio do espectrograma que será discutido com mais detalhes na seção 2.2.1.

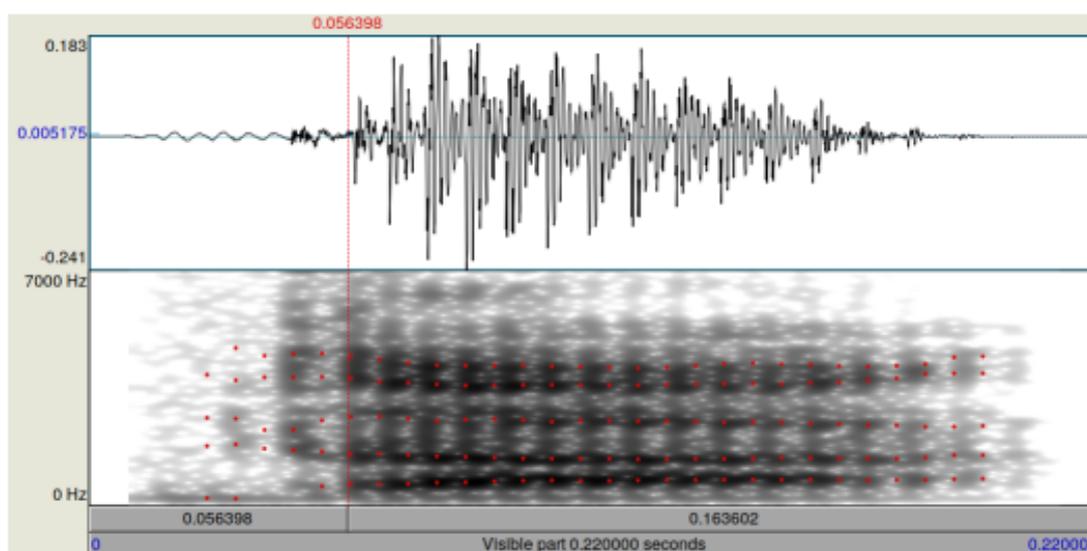
2.2.1 Análise espectrográfica da voz

Um espectrograma é uma representação visual da distribuição de energia acústica por meio de frequências ao longo do tempo. O eixo horizontal de um espectrograma normalmente representa o tempo, o eixo vertical representa as etapas de frequência discretas e a quantidade de

potência detectada é representada como a intensidade em cada ponto de frequência de tempo (LAMPERT; O'KEEFE, 2010).

A Figura 1 ilustra um espectrograma obtido para a sílaba /da/ no *software* Praat. Os formantes da vogal “a” desta sílaba podem ser visualizados em linhas horizontais pretas, marcadas pelo *software* com pontos vermelhos, mostrando um relativo acúmulo de energia nestas frequências (BOERSMA; WEENINK, 2018).

Figura 1 – Espectrograma da sílaba /da/



Fonte: Santana (2020)

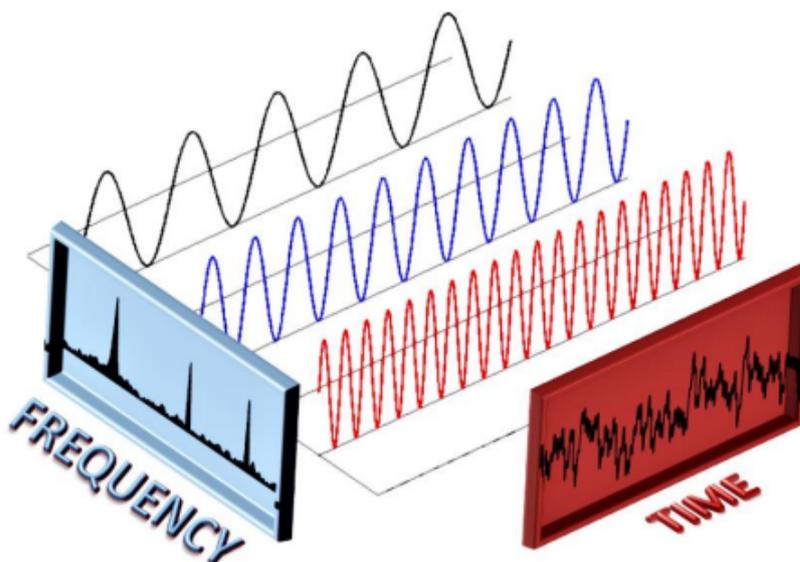
Segundo Pontes et al. (2002) a análise espectrográfica é uma das principais ferramentas análise acústica da voz, uma vez que apresenta visualmente as características acústicas da emissão e complementa a avaliação perceptivo-auditiva. O mesmo autor ainda explica que a espectrografia traduz padrões sonoros em padrões visuais gráficos que apresentam a frequência fundamental e os harmônicos de uma voz, assim como as zonas de forte intensidade dos sons da fala. Para realizar tal processo, utiliza-se uma análise matemática denominada “análise de Fourier”, que determina os componentes da onda sonora.

2.3 Análise de Fourier

Segundo Briggs e Henson (1995), no século XIX um matemático e físico francês chamado Jean Baptiste Joseph Fourier (1768-1830) demonstrou que qualquer forma de onda pode ser representada por uma somatória de senóides e cossenóides de diferentes frequências, amplitudes e fases. Isso pode ser compreendido por meio da Figura X, que ilustra a transformada de Fourier de um sinal, onde é possível visualizá-lo no domínio do tempo e sua respectiva transformada no domínio da frequência (PINTO, 2018). Esse processo ficou conhecido como a análise de

Fourier, uma família de técnicas matemáticas, todas elas baseadas na decomposição de sinais em senóides.

Figura 2 – Representação de um sinal no domínio no tempo e sua respectiva transformada de fourier



Fonte: Pinto (2018)

Os sinais podem ser classificados segundo vários critérios, por exemplo: sinais contínuos ou discretos e sinais periódicos ou aperiódicos. Estes dois critérios levam aos quatro elementos da família de transformadas de Fourier, que são utilizados para tratar cada caso. Para sinais contínuos e aperiódicos se aplica a Transformada de Fourier; para sinais contínuos e periódicos a Série de Fourier; para sinais discretos e aperiódicos a Transformada de Fourier de Tempo Discreto (DTFT); por fim, para os sinais discretos e periódicos utiliza-se a Transformada Discreta de Fourier (Discrete Fourier Transform - DFT) (JENKINS, 1999).

A transformada de Fourier funciona calculando o produto escalar entre o sinal e ondas senoidais de diferentes frequências. As ondas senoidais têm três características: frequência (o número de oscilações de onda por um certo período de tempo, medido em ciclos por segundo ou hertz), amplitude (corresponde à altura da onda, marcada pela distância entre o ponto de equilíbrio da onda até a crista) e fase (o tempo de a onda senoidal, medida em radianos ou graus). Dessa forma, o resultado da transformada de Fourier é uma representação tridimensional (3-D) dos dados da série temporal em que as três dimensões são frequência, amplitude e fase (COHEN, 2014).

A maioria dos sinais de fala são processos não estacionários com múltiplos componentes que podem variar em tempo e frequência (BOUALEM BOASHASH, 2016). Entretanto, segundo Hlawatsch e Boudreaux-Bartels (1992), a transformada de Fourier não mostra explicitamente a localização no tempo dos componentes de frequência, mas alguma forma de localização no tempo pode ser obtida usando uma pré-janela adequada. Para isso utiliza-se a transformada de

Fourier de curta duração (Short-Time Fourier Transform - STFT) que é uma representação linear de tempo-frequência (TFR) usada para apresentar mudanças no sinal que variam com o tempo.

A STFT é um método alternativo para extrair informações de potência e fase em tempo-frequência. A STFT é uma extensão simples da transformada de Fourier que aborda as suas duas principais limitações para análises tempo-frequência: a transformada de Fourier obscurece mudanças variantes no tempo na estrutura de frequência do dados, e assume que os dados são estacionários durante a série temporal (COHEN, 2014).

A STFT é um método bem conhecido para a análise tempo-frequência de sinais não estacionários (NURUZZAMAN; BOYRAZ; JALALI, 2006). A abordagem STFT é realizar uma Transformada de Fourier em apenas uma pequena seção (janela) de dados por vez, mapeando assim o sinal em uma função bidimensional (2-D) de tempo e frequência (BUSTAMI et al., 2007).

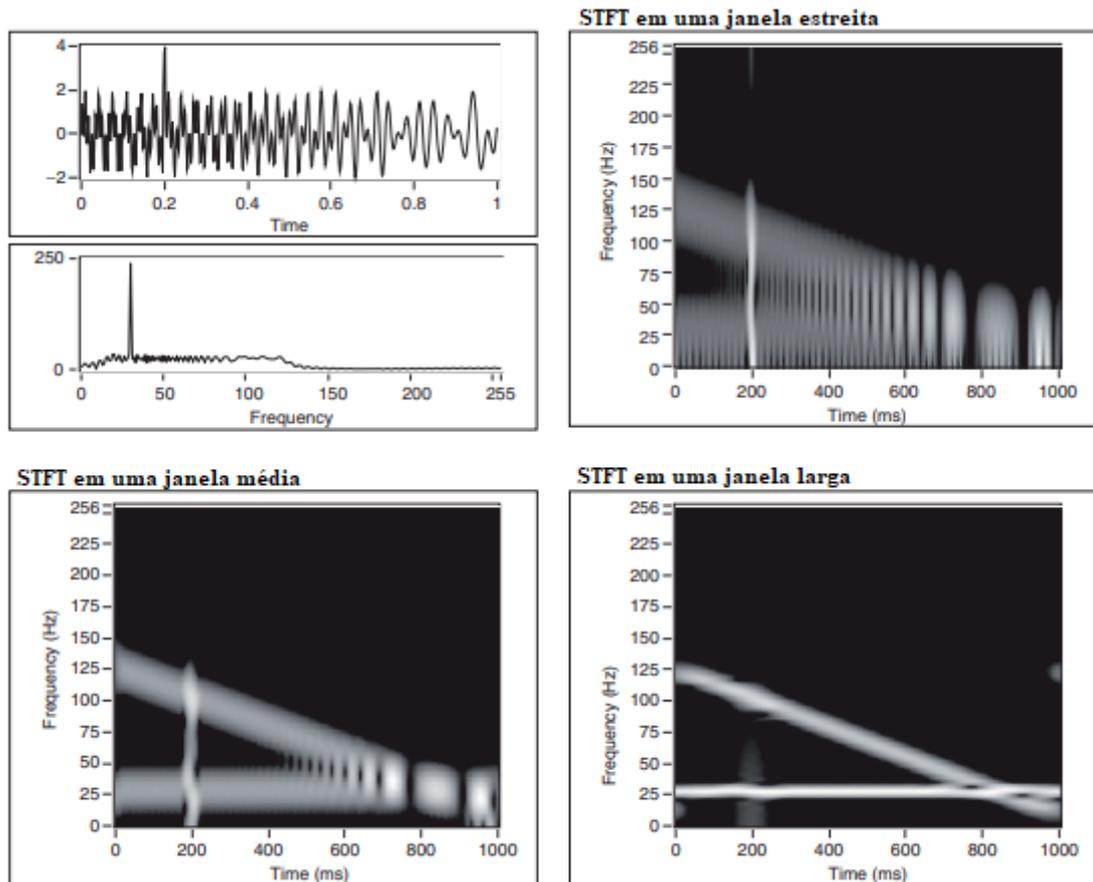
Nuruzzaman, Boyraz e Jalali (2006) explicam que em uma STFT, o sinal é segmentado aplicando-se uma janela a ele. A largura desta janela deve ser curta o suficiente para que o sinal dentro da janela seja considerado estacionário. O tamanho da janela determina a resolução temporal da análise tempo-frequência e estabelece quão bem um evento não estacionário pode ser localizado no tempo. Segundo Krishnan (2021), a versão discreta da STFT de um sinal $x(n)$ pode ser expressa como:

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m} \quad (2.2)$$

onde $w(n)$ é a janela de análise, que é considerada diferente de zero apenas no intervalo $[0, N-1]$.

Segundo Kehtarnavaz (2008), existe um *trade-off* entre resolução de tempo e frequência na STFT. Em outras palavras, embora uma janela de tempo de largura estreita resulte em uma melhor resolução no domínio do tempo, ela gera uma resolução ruim no domínio da frequência e uma janela de tempo mais longa fornece melhor precisão e resolução de frequência à custa da diminuição da precisão temporal. Conforme apresentado por Tonidandel e Araujo (2017), temos o chamado efeito de espalhamento recíproco (ou escalamento) da transformada de Fourier. Esta propriedade aponta o fato de que se um sinal sofre uma compressão no domínio do tempo, ocorrerá uma expansão (espalhamento) do espectro no domínio da frequência e vice-versa. Essa propriedade é também conhecida como “Princípio da Incerteza” na teoria da transformada de Fourier. Três espectrogramas ilustrando diferentes resoluções de tempo-frequência são mostrados na Figura 2, referentes ao sinal original no domínio do tempo e sua Transformada de Fourier ilustrados no canto superior esquerdo.

Figura 3 – STFT com diferentes resoluções de tempo-frequência



Fonte: Kehtarnavaz (2008)

A resolução de frequência de uma transformada de Fourier é definida pelo número de pontos na série temporal. Portanto, quanto maior a janela de tempo, mais frequências podem ser extraídas e, portanto, maior a resolução de frequência. Já a resolução temporal é sempre definida pela taxa de amostragem de dados e, portanto, não depende apenas da duração do segmento de tempo (COHEN, 2014). Segundo o mesmo autor, a taxa de amostragem refere-se ao número de vezes por segundo que os dados são adquiridos.

3 DESENVOLVIMENTO

3.1 Caracterização da área de estudo

Esta monografia tem como objetivo o desenvolvimento de uma aplicação *web* acessível para processamento e análise da voz. Para tal buscou-se desenvolver uma aplicação *web* de baixo custo e amigável com boa qualidade das imagens espectrográficas para realizar processamento e análise da voz.

3.2 Classificação do trabalho

O presente trabalho é classificado como pesquisa exploratória aplicada, pois tem o intuito de auxiliar fonoaudiólogos(as) e profissionais da área na análise do estado da voz de seus pacientes.

Além disso, o projeto é de cunho tecnológico, onde a pesquisa é qualitativa, pois a aplicação busca coletar dados, realizar análise e apresentar o resultado para o usuário.

A pesquisa realizada nesta monografia utilizou-se de livros, artigos científicos, teses, documentação em sites de banco e do Governo, além de documentação específica de linguagens de programação de desenvolvimento. Por isso pode ser caracterizada como pesquisa documental, uma vez que além de utilizar de contribuições de diversos autores sobre determinado assunto, ainda utilizada de materiais que não receberam ainda um tratamento analítico ou que ainda podem ser reelaborados de acordo com os objetos da pesquisa.

3.3 Desenvolvimento da aplicação *web*

O desenvolvimento deste projeto inicia-se com a identificação dos requisitos da aplicação com o auxílio de uma profissional da área de análise acústica da voz. O segundo passo foi escolher fazer uma aplicação *web* em vez de software ou aplicativo para que fosse possível rodar em qualquer sistema operacional.

Em seguida, com base nos requisitos, foram levantadas as técnicas de processamento de sinais biomédicos necessárias para implementar as funcionalidades desejadas. A partir deste levantamento foi realizado uma revisão bibliográfica para conhecimento do funcionamento destas técnicas, em especial a STFT.

Para que o desenvolvimento da aplicação, além de procurar uma ferramenta que fosse de código aberto, um outro ponto muito importante considerado foi a facilidade de se trabalhar com as técnicas de processamento dos sinais de voz, como, por exemplo, a transformada de Fourier.

Avaliando algumas opções de linguagens e bibliotecas para desenvolvimento de *software*

tanto *back-end* e *front-end*, identificou-se no *Python* uma linguagem que oferece uma gama enorme de bibliotecas e funcionalidades voltadas para a área de tratamento de sinais. Além disso optou-se por utilizar o Streamlit, que é um *framework* de código aberto que utiliza linguagem *Python*. Este, além de ser gratuito e acessível é de fácil desenvolvimento e manutenção, uma vez que foi criado justamente para ajudar cientistas de dados a colocarem em produção seus projetos sem a necessidade de conhecer ferramentas de *front-end* ou de *deploy* de aplicações.

O ambiente de desenvolvimento integrado (IDE) utilizado foi o VisualStudio Code, que é um editor de código-fonte desenvolvido pela Microsoft para Windows, Linux e macOS.

Dando início à construção do código, foi feita a importação das bibliotecas a serem utilizadas conforme figura 3:

Figura 4 – Importando bibliotecas

```
import streamlit as st
import librosa
import matplotlib.pyplot as plt
import numpy as np
import librosa.display
from PIL import Image
from scipy.io.wavfile import write
import amfm_decompy.pYAAPT as pYAAPT
import amfm_decompy.basic_tools as basic
import sounddevice
```

Fonte: Arquivo pessoal (2022)

Para geração das visualizações, foi utilizado a biblioteca *Parselmouth* (JADOUL; THOMPSON; BOER, 2018). Essa biblioteca baseia-se na base de código do famoso *software* Praat, que implementa uma variedade enorme de algoritmos de processamento de fala e fonética. Muitas dessas funções podem ser acessadas por meio do *Parselmouth*. Para isso, *Parselmouth* utiliza a biblioteca *pybind11*, o que permite expor a funcionalidade C/C++ do Praat como uma interface *Python*.

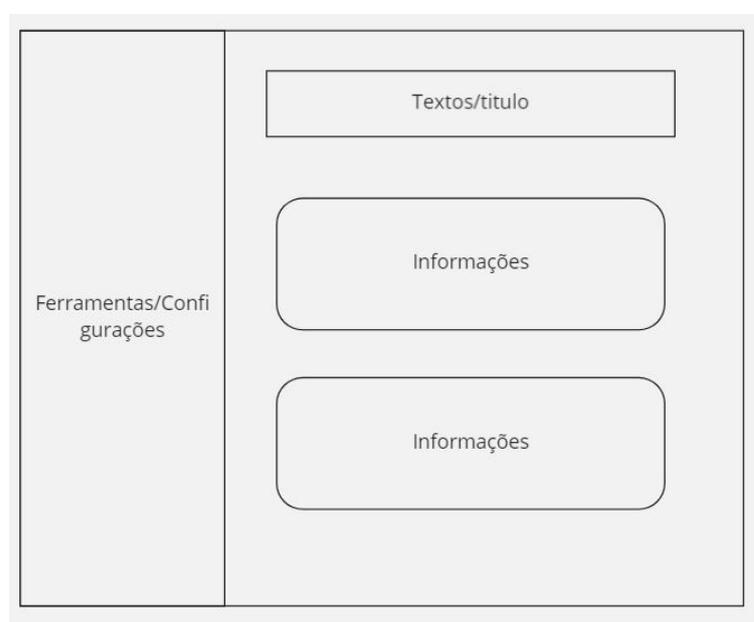
Alguns profissionais já tentaram portar a funcionalidade do Praat para o *Python*, porém o *Parselmouth* se mostrou superior. Segundo a comunidade desenvolvedora, no objetivo de fornecer uma interface completa e simples para o código interno do Praat. Enquanto outros projetos envolvem a linguagem de script do Praat ou reimplementam partes da funcionalidade do Praat em *Python*, o *Parselmouth* acessa diretamente o código C/C++ do Praat (o que significa que os algoritmos e sua saída são exatamente os mesmos do Praat) e fornece acesso eficiente aos dados do programa, mas também fornece uma interface que não é diferente de qualquer outra biblioteca *Python*.

Infelizmente, a biblioteca *Parselmouth* ainda possui uma documentação pouco organizada, ainda em processo de criação e melhoria. Trata-se de uma biblioteca relativamente nova,

com uma comunidade crescente, e, por isso, não se encontra muito material disponível para consulta e estudo. Logo, muitas das funcionalidades ainda não são devidamente descritas e apresentadas, o que dificulta a sua utilização. Para este projeto, foram utilizados parâmetros básicos para criação das visualizações. Pretende-se com o melhoramento da biblioteca, avançar no aperfeiçoamento da sua utilização e melhoria da aplicação proposta neste trabalho.

Dando continuidade, iniciou-se então o desenvolvimento do *front-end* da aplicação. A nível de *layout* e desenvolvimento, inicialmente foi pensada uma interface simples e intuitiva, baseada em outras inúmeras aplicações, o que trará ao(a) usuário(a) uma grande familiaridade com o sistema. A figura 4 apresenta o esboço inicial da aplicação:

Figura 5 – Esboço da aplicação



Fonte: Arquivo pessoal (2022)

O Streamlit é uma ótima opção na construção *do front-end*, pois possibilita a criação rápida de componentes. Para criação da barra lateral (*Sidebar*), basta utilizar o atributo *sidebar*, juntamente com o componente que deseja adicionar, seja ele um título, botão, checkbox, dentre outros. Na figura 5 tem-se um trecho do código que exemplifica este caso:

Figura 6 – Trecho | Exemplo de implementação - Sidebar Menu

```
st.sidebar.title("Upload de arquivo")
uploaded_file = st.sidebar.file_uploader("Upload your input Audio file", type=["wav","ogg"])
st.sidebar.title("Gravar arquivo de áudio")
time_duration = st.sidebar.number_input('Informe o tempo de gravação (sec) que deseja realizar:', step=1, min_value=0, value=5)
gravar = st.sidebar.button("Gravar áudio")
```

Fonte: Arquivo pessoal (2022)

Foram dispostos em tela algumas funcionalidades e configurações passíveis de alteração por parte do(a) usuário(a), como campo para *setpoint* da frequência de amostragem, duração do áudio a ser gravado, definição do mapa de cores a ser utilizado, além de configurações avançadas como o comprimento da janela e definição da faixa dinâmica.

Seguindo a proposta do esboço inicial, na parte principal da aplicação será disponibilizado o resultado, ou seja, o *plot* do espectrograma para análise. Conforme dito anteriormente, a biblioteca Parselmouth trabalha utilizando o Praat em segundo plano e entrega a visualização da aplicação. Para integrar esta biblioteca, foram criadas duas funções fundamentais, responsáveis pelo "desenho" tanto do espectrograma, quanto da visualização do pitch em tela. As funções foram chamadas de *draw_spectrogram* e *draw_pitch*, ilustradas nas figuras figuras 6 e 7:

Figura 7 – Funções para geração do espectrograma e pitch

```

26 def draw_spectrogram(spectrogram, dynamic_range=st.session_state['dynamic_range']):
27     fig = plt.figure(figsize=(16,7))
28     X, Y = spectrogram.x_grid(), spectrogram.y_grid()
29     sg_db = 10 * np.log10(spectrogram.values)
30     plt.pcolormesh(X, Y, sg_db, vmin=sg_db.max() - dynamic_range, cmap=cmap_option)
31     plt.ylim([spectrogram.ymin, spectrogram.ymax])
32     plt.xlabel("time [s]")
33     plt.ylabel("frequency [Hz]")
34     plt.colorbar(format='%2.0f db')
35
36 def draw_pitch(pitch):
37     pitch_values = pitch.selected_array['frequency']
38     pitch_values[pitch_values==0] = np.nan
39     plt.plot(pitch.xs(), pitch_values, 'o', markersize=5, color='w')
40     plt.plot(pitch.xs(), pitch_values, 'o', markersize=2)
41     plt.grid(False)
42     plt.ylim(0, pitch.ceiling + st.session_state['pitch.ceiling'])
43     plt.ylabel("fundamental frequency [Hz]")

```

Fonte: Arquivo pessoal (2022)

Os elementos *st.session_state['dynamic_range']* e *st.session_state['pitch.ceiling']* são variáveis cujo estado está guardado para a próxima interação. O streamlit roda todo o código do início ao fim em todas as interações realizadas na tela, logo, é possível que alguma variável volte ao seu estado inicial se ela não for previamente salva. A função *session_state* garante que aquela variável permaneça com aquele último valor que ela recebeu na interação anterior.

Utilizando a biblioteca *sounddevice*, foi implementado também a possibilidade de gravação de um arquivo de áudio com duração previamente definida pelo usuário. Para isso, utilizou-se a função *rec*, que tem como atributos necessários a duração desejada multiplicada pela taxa de amostragem desejada, o número de canais, além da taxa de amostragem em si. O código vai aguardar o tempo necessário para gravação e em seguida irá gerar o arquivo de áudio *.wav* com um nome pré-definido pelo desenvolvedor. Esse nome será utilizado posteriormente para carregamento do arquivo de áudio recém gravado. Na figura 7 podemos visualizar o que foi descrito:

Figura 8 – Gravação do arquivo de áudio com duração pré-definida

```

print("Recording... ")
recording = sounddevice.rec(int(time_duration*st.session_state['freq_amostragem']), samplerate=st.session_state
['freq_amostragem'], channels=2)
with st.spinner('Gravando ... '):
    sounddevice.wait()

print("Done!")
write("output.wav", st.session_state['freq_amostragem'], recording)

```

Fonte: Arquivo pessoal (2022)

Como dito anteriormente, a biblioteca *Parselmouth* nos entrega o espectrograma finalizado após a passagem de alguns parâmetros. Não há a necessidade de realização do cálculo da transformada de Fourier a nível de desenvolvimento, tudo isso é feito usando o Praat na camada inferior. É necessário, porém, que se entenda o seu funcionamento e efeito da alteração dos parâmetros de entrada a função.

A seguir, é feita uma explicação mais detalhada do código básico para geração de um espectrograma utilizando Praat por meio da biblioteca Parselmouth:

Figura 9 – Utilizando Parselmouth

```

409     snd = parselmouth.Sound(uploaded_file.name)
410     # _____
411
412
413
414     pitch = snd.to_pitch()
415     pre_emphasized_snd = snd.copy()
416     pre_emphasized_snd.pre_emphasize()
417     spectrogram = pre_emphasized_snd.to_spectrogram(window_length=st.session_state['comp_janela'],
418                                                     maximum_frequency=st.session_state['max_freq'])
419     plt.figure()
420     draw_spectrogram(spectrogram)
421     plt.twinx()
422     draw_pitch(pitch)
423     plt.xlim([snd.xmin, snd.xmax])

```

Fonte: Arquivo pessoal (2022)

Referenciando-se nas linhas de código apresentadas na figura 8, inicialmente, Na linha 409 o nosso arquivo de áudio foi carregado utilizando a função *Sound*. Para utiliza-la, basta informar o nome do arquivo e se certificar de que esse arquivo esteja presente na pasta raiz do projeto. A função *Sound* retorna um objeto (salvo na variável *snd*) com todas as funções e propriedades necessárias. Para a geração do objeto *Pitch* executa-se a função *snd.to_pitch*, conforme linha 414. Esse objeto é utilizado como parâmetro para a função *draw* (citada anteriormente) na linha 421. Para geração do espectrograma, é possível antecipadamente enfatizar as intensidades da gravação, usando a função *pre_emphasize()* da linha 416. Com o retorno dessa função, pode-se aplicar a função *to_spectrogram()*, onde é possível passar os parâmetros de comprimento da janela utilizada e a frequência máxima a ser visualizada no espectrograma. No exemplo da figura 9 utiliza-se variáveis com estado previamente salvo, conforme explicado anteriormente. Por fim, basta chamar a função *draw_spectrogram* passando a reposta da função *to_spectrogram()*, executada na linha 417.

Também foram implementadas algumas lógicas que permitem ao usuário utilizar a aplicação de forma contínua, excluindo ou adicionando áudios, alterando configurações do espectrograma, e visualizando ou não outros componentes, como a forma de onda. Todas essas configurações estão disponíveis no menu de ferramentas na lateral esquerda da aplicação.

3.4 Acesso ao código fonte da aplicação

Todo o código desenvolvido pode ser acessado no repositório [SpectroApp](#) do Github. Certifique-se de acessar a branch *ft-parselmouth*, onde foram desenvolvidas todas as *features* utilizando a biblioteca Parselmouth.

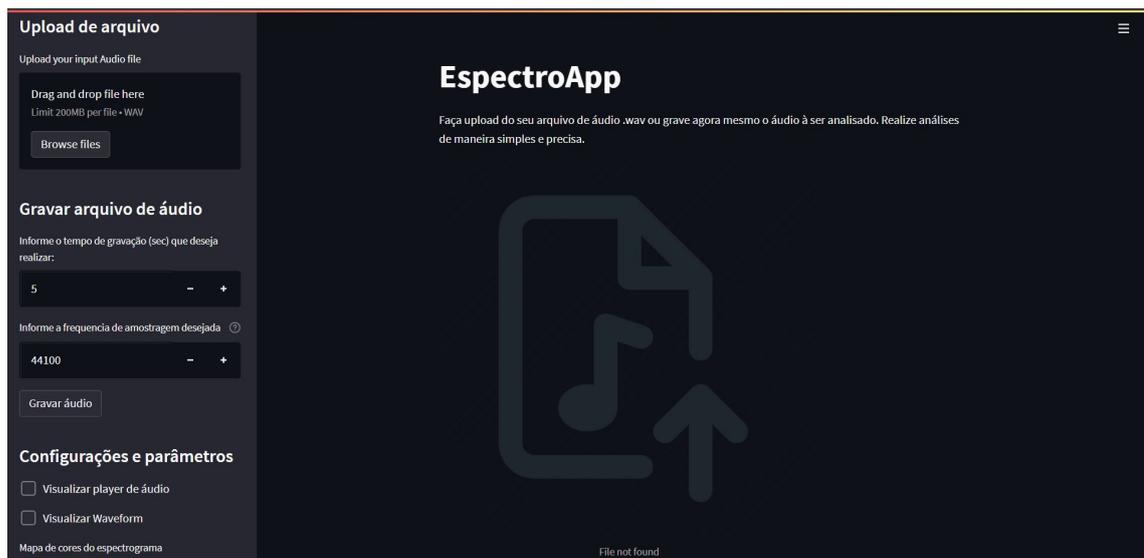
4 RESULTADOS

4.1 Apresentação da aplicação

Com a aplicação devidamente implementada, *bugs* e possíveis erros de fluxo de navegação corrigidos e previstos, foram realizados alguns testes para verificação das funcionalidades e possibilidades acerca das bibliotecas utilizadas.

A tela inicial da aplicação possui um *layout* simples e de fácil entendimento. Existe uma breve descrição e orientação acompanhada do indicativo de que nenhum áudio foi carregado até o momento. A barra lateral (*Sidebar*) aparece visível, sendo possível também ocultá-la. A aplicação possui a funcionalidade *light/dark mode*, onde o(a) usuário(a) pode deliberar acerca de qual aparência ele(a) quer que a aplicação tenha, de forma simples e rápida. A seguir, as figuras 9 e 10 apresentam o resultado final da aplicação em modo inicial:

Figura 10 – Tela inicial | Black mode



Fonte: Arquivo pessoal (2022)

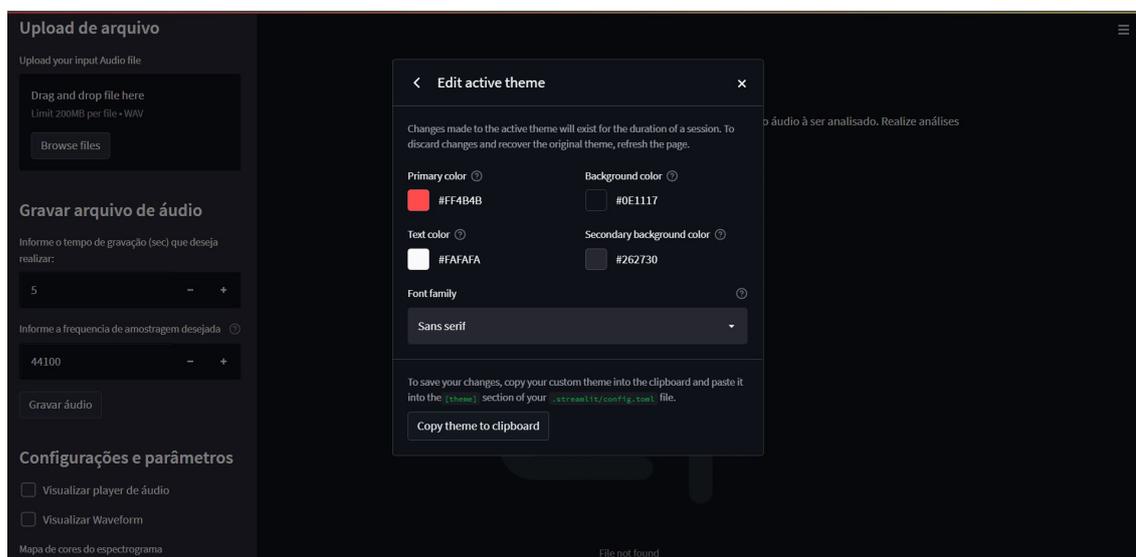
Figura 11 – Tela inicial | Ligth mode



Fonte: Arquivo pessoal (2022)

Ainda sobre a aparência da aplicação, o(a) usuário(a) também pode configurar outras características como cores de fonte primárias e secundárias e *background* além da fonte utilizada na aplicação. Para isso basta acessar o "menu sandwich" à direita, ir até *settings*, e acessar a opção "Edit active theme" ou "Editar tema ativo", conforme mostrado na figura 11:

Figura 12 – Tela inicial - Configurações de estilo



Fonte: Arquivo pessoal (2022)

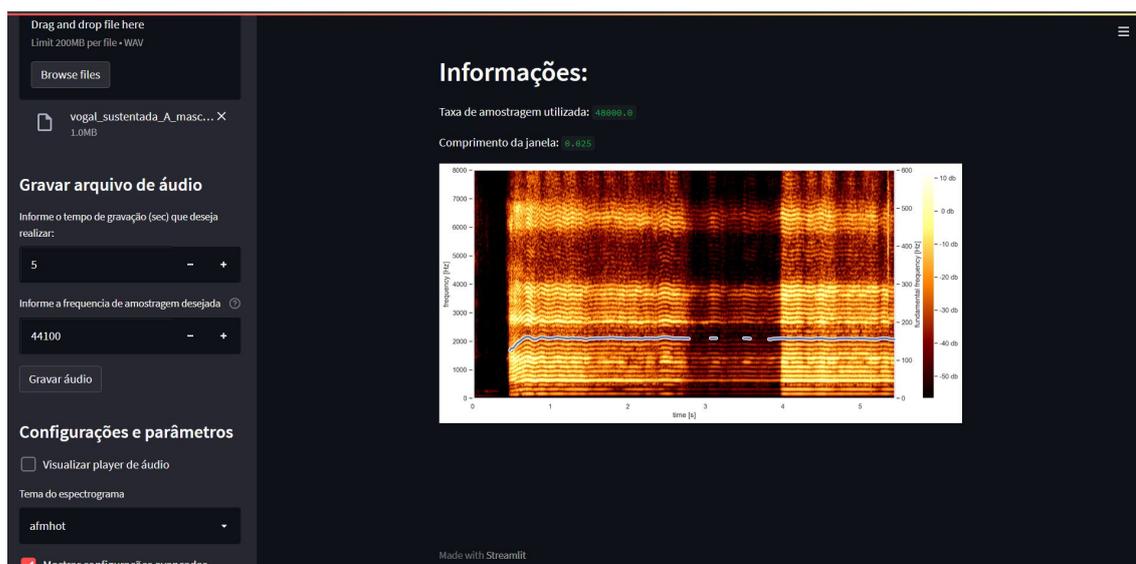
4.2 Testes

O primeiro teste foi realizado utilizando a vogal sustentada "A". Inicialmente, gravou-se externamente um arquivo de áudio emitido a vogal sustentada e então foi feito o *upload* do arquivo. Logo após a finalização do carregamento, o espectrograma é mostrado em tela de

maneira automática. Além disso, tem-se algumas informações pertinentes que permitem ao(a) usuário(a) entender quais configurações estão sendo utilizadas.

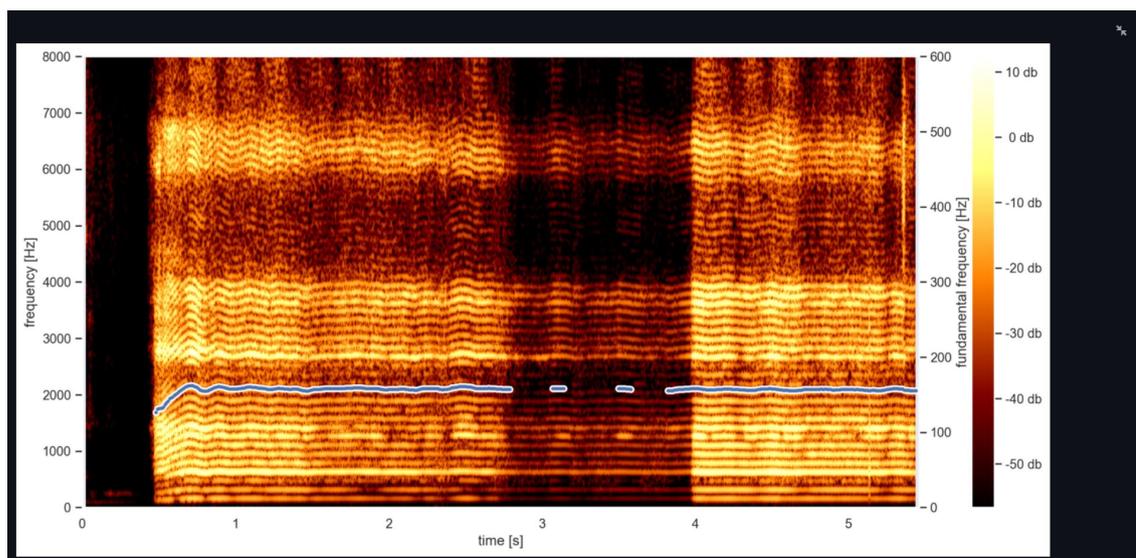
Foram realizados testes com duas pessoas, uma do sexo masculino e outra do sexo feminino. A seguir tem-se nas figuras 12, 13, 14 e 15 os espectrogramas resultantes. Algumas configurações foram alteradas apenas a nível de demonstração e teste das funcionalidades. Cada mudança é citada no título da respectiva imagem:

Figura 13 – Espectrograma vogal sustentada A | Voz masculina - comp. Janela 0.025 s



Fonte: Arquivo pessoal (2022)

Figura 14 – Espectrograma vogal sustentada A | Voz masculina - ampliado



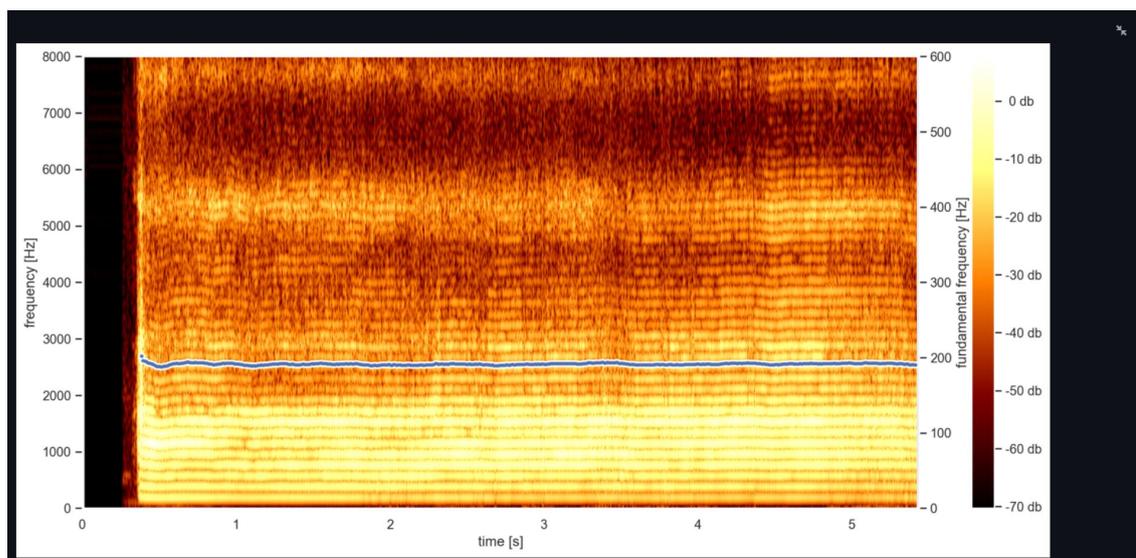
Fonte: Arquivo pessoal (2022)

Figura 15 – Espectrograma vogal sustentada A | Voz feminina



Fonte: Arquivo pessoal (2022)

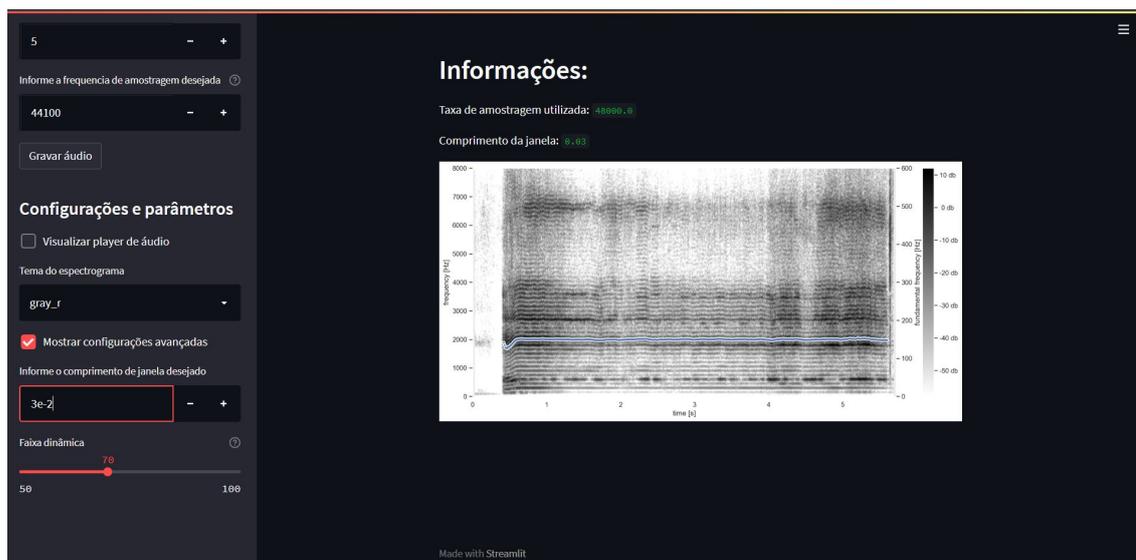
Figura 16 – Espectrograma vogal sustentada A | Voz feminina - ampliado



Fonte: Arquivo pessoal (2022)

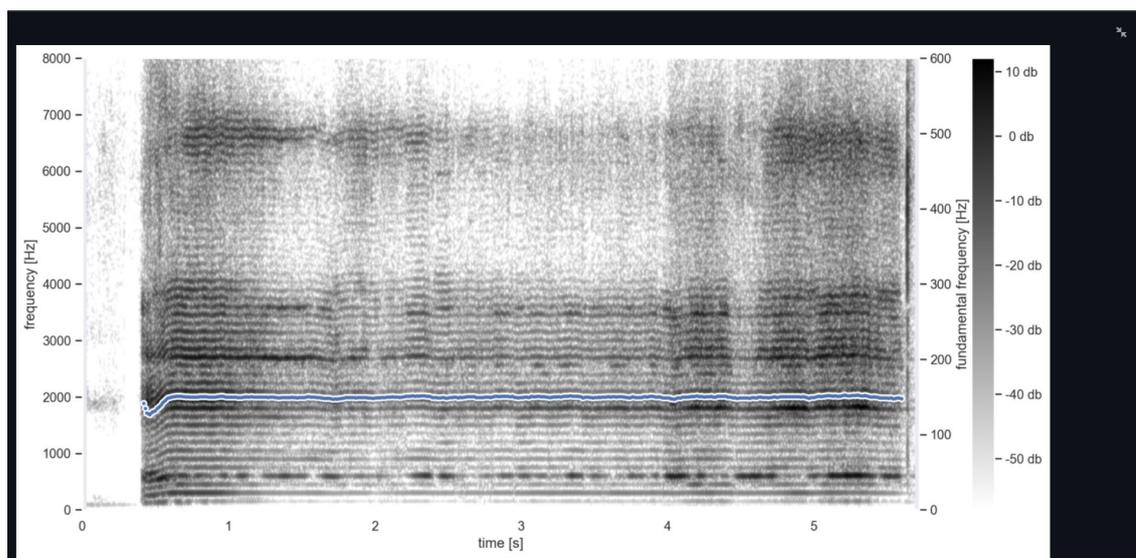
Como visto, é possível que o usuário amplie a visualização do espectrograma para auxiliar a análise clicando no ícone no canto superior direito do *plot*. A seguir são mostrados outros resultados, também envolvendo duas vozes (masculina e feminina) onde foram testados áudios emitindo as vogais O e E de forma sustentada. Novamente, foram aplicadas algumas alterações nos parâmetros disponíveis para edição apenas para exemplificação e teste de possibilidades:

Figura 17 – Espectrograma vogal sustentada E | Voz masculina - comp. Janela 0.03s



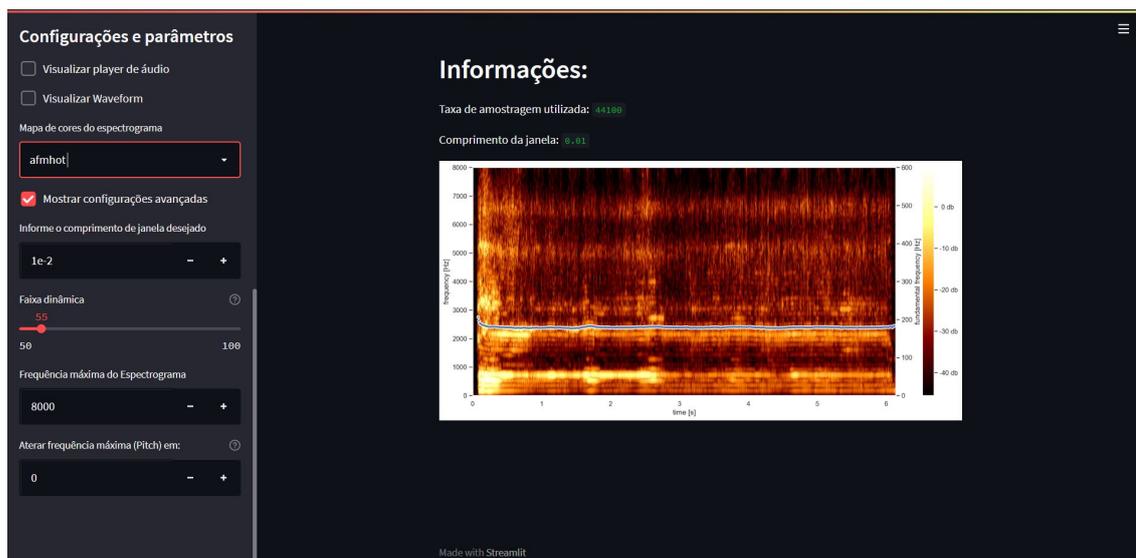
Fonte: Arquivo pessoal (2022)

Figura 18 – Espectrograma vogal sustentada E | Voz masculina - comp. Janela 0.03s ampliado



Fonte: Arquivo pessoal (2022)

Figura 19 – Espectrograma vogal sustentada E | Voz feminina



Fonte: Arquivo pessoal (2022)

Figura 20 – Espectrograma vogal sustentada O | Voz masculina comp. Janela 0.012



Fonte: Arquivo pessoal (2022)

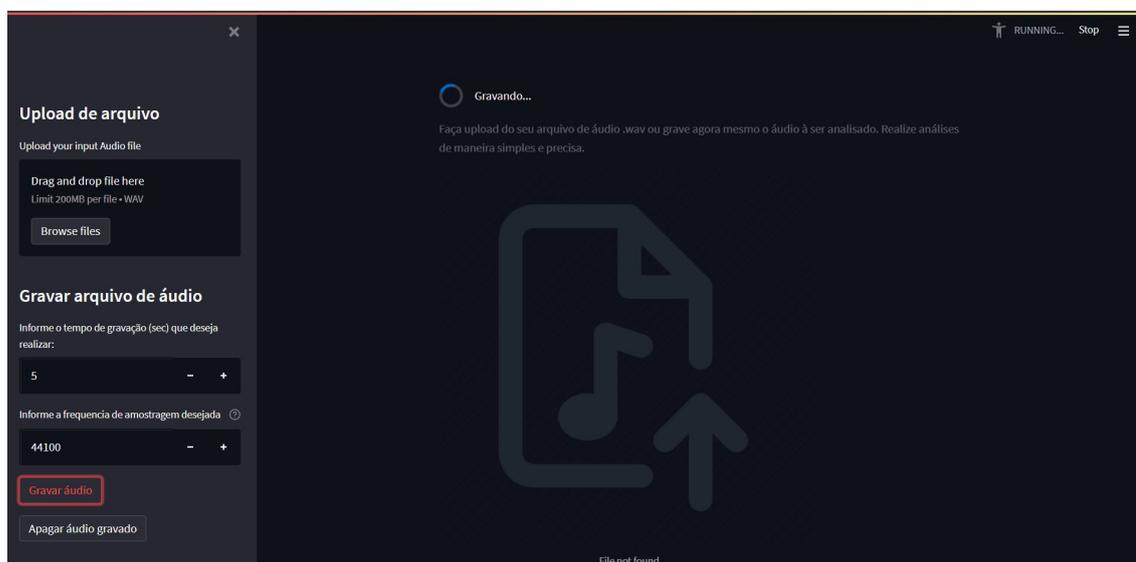
Figura 21 – Espectrograma vogal sustentada O | Feminino



Fonte: Arquivo pessoal (2022)

Para teste e validação da gravação do áudio dentro da aplicação, foi realizado o teste também emitindo a vogal sustentada A e I por um período de 5 segundos por uma voz masculina. Ao clicar no botão "Gravar áudio", temos um *spinner* indicando que o áudio está sendo gravado (figura 21). É possível verificar o resultado (tela padrão e ampliada) na figura 22, 23 e 24:

Figura 22 – Áudio sendo gravado



Fonte: Arquivo pessoal (2022)

Figura 23 – Áudio gravado - Vogal I Sustentada



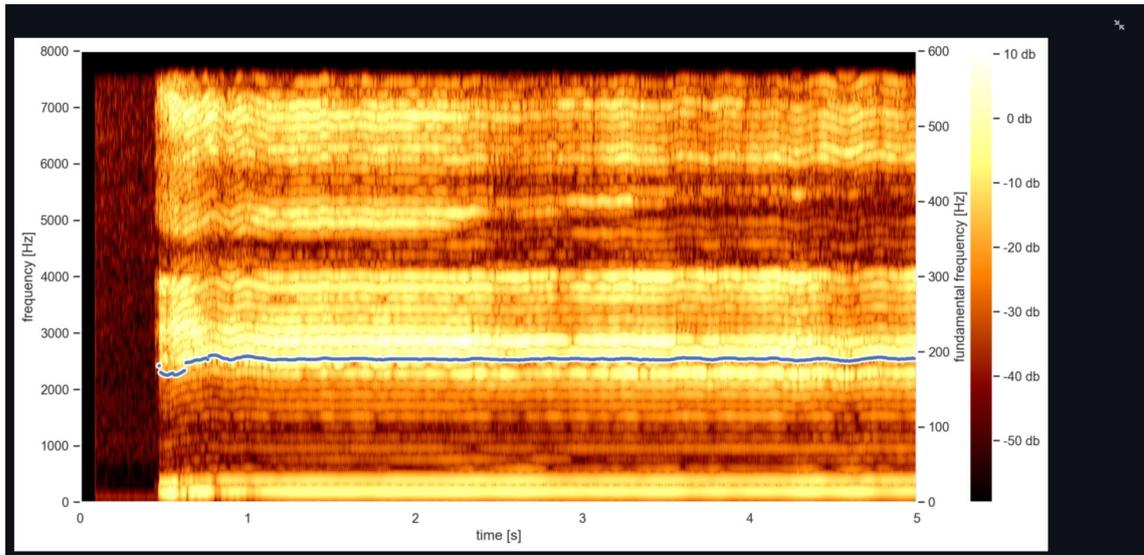
Fonte: Arquivo pessoal (2022)

Figura 24 – Áudio gravado - Vogal A sustentada



Fonte: Arquivo pessoal (2022)

Figura 25 – Áudio gravado - Vogal I sustentada - Cor do espectrograma: afmhot



Fonte: Arquivo pessoal (2022)

5 CONCLUSÃO

Este projeto teve como objetivo o desenvolvimento de uma aplicação *web* acessível para processamento e análise da voz.

O resultado prático foi a construção de uma aplicação capaz de receber e gravar áudios para análise e apresentar como resultado principal um espectrograma para auxiliar na análise acústica da voz, além da possibilidade de visualização da forma de onda do áudio carregado e o detector de pitch. Os gráficos gerados possuem uma resolução satisfatória, podendo apresentar um desempenho ainda maior para áudios gravados em ambientes com poucos ou nenhum ruído. Além disso, foram realizados testes para avaliar as funcionalidades implementadas na aplicação.

Durante o desenvolvimento, identificou-se na linguagem escolhida (*Python*) uma forte e ágil ferramenta para a análise de sinais, auxiliando na construção de visualizações, análises e cálculos matemáticos de forma simples e clara. O mesmo se conclui a respeito da ferramenta Streamlit. Por meio dela foi possível a criação de uma interface robusta, de fácil entendimento e fluidez, facilitando o processo de análise acústica da voz de forma geral.

Para geração das visualizações, foi utilizada a biblioteca *Parselmouth* que tem como base o código do famoso *software* Praat, que implementa uma variedade enorme de algoritmos de processamento de fala e fonética. Como dito, infelizmente, a biblioteca *Parselmouth* ainda possui uma documentação pouco organizada e que ainda está em processo de criação e melhoria. Porém, ela se mostrou muito eficiente na geração dos espectrogramas, sendo uma biblioteca de fácil interpretação, além da fácil utilização. Trata-se de uma biblioteca relativamente nova, com uma comunidade crescente, e por isso, não se encontra muito material disponível para consulta e estudo. Logo, muitas das funcionalidades ainda não são devidamente descritas e apresentadas, o que dificulta a sua utilização. Para este projeto, foram utilizados parâmetros básicos para criação das visualizações. Pretende-se com o melhoramento da biblioteca, avançar-se no aperfeiçoamento da sua utilização e melhoria da aplicação.

Recomenda-se para trabalhos futuros a adição de novas funcionalidades a aplicação. Um exemplo muito interessante seria a possibilidade de geração do espectrograma durante a captação da voz em tempo real. Uma outra opção, seria a possibilidade de gravação da voz por período indeterminado, possibilitando a pausa e retomada da gravação. Também pretende-se implementar a funcionalidade de recorde do áudio gravado, possibilitando que o(a) usuário(a) informe qual trecho de áudio ele(a) deseja analisar. Além disso, embutir no *software* sugestões de diagnósticos a partir da análise automática das imagens com base em um conjunto de dados de vozes normais e disfuncionais. Possibilitar obter leituras de tempo e frequência ao clicar sobre um ponto qualquer do traçado do pitch.

Além das recomendações apresentadas, sugere-se obter a opinião de profissionais da

área sobre o funcionamento da aplicação, a fim de realizar ajustes e melhorias que contribuam ainda mais para o processo de análise acústica da voz.

Por fim, espera-se com esse projeto, contribuir com o profissionais de fonoaudiologia, bem como outro profissionais da área da saúde em termos de acessibilidade a um *software* de baixo custo que auxilie no processo de análise acústica da voz.

6 REFERÊNCIAS

BEHLAU, M.; AZEVEDO, R.; PONTES, P. Conceito de voz normal e classificação das disfonias. In: BEHLAU, M. **Voz: o livro do especialista**, 1. Rio de Janeiro: Revinter, 2001a. cap. 2, p. 53-79.

BOUALEM BOASHASH (ed.). **Time-Frequency Signal Analysis and Processing**. 2. ed. [S. L.]: Boualem Boashash, 2016.

BUSTAMI, Fadzlul Rahimi Ahmad et al. The Application Of Short Time Fourier Transform And Image Processing Techniques To Detect Human Heart Abnormalities. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING AND INFORMATICS INSTITUT TEKNOLOGI BANDUNG, 6., 2007, Indonesia. **Anais [...]**. Indonesia: Proceedings Of The International Conference On Electrical Engineering And Informatics Institut Teknologi Bandung, 2007. p. 450-453.

CARMO JÚNIOR, José Roberto do. **Melodia e Prosódia: um modelo para a interface música-fala com base no estudo comparado do aparelho fonador e dos instrumentos musicais reais e virtuais**. 2007. 192 f. Tese (Doutorado) - Curso de Linguística, Universidade de São Paulo, São Paulo, 2007.

COHEN, Mike X. **Analyzing Neural Time Series Data: theory and practice**. Cambridge: The Mit Press, 2014.

CÔRTEZ, Marcela Guimarães; GAMA, Ana Cristina Côrtes. Análise visual de parâmetros espectrográficos pré e pós-fonoterapia para disfonias. **Revista da Sociedade Brasileira de Fonoaudiologia**, [S.L.], v. 15, n. 2, p. 243-249, 2010. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1516-80342010000200016>.

GAMA, Ana Cristina Côrtes. **Análise Acústica de Voz: Estudo comparativo entre mulheres sem queixa vocal e com disfonia**. 1997. 65 f. Dissertação (Mestrado) - Curso de Mestre em Distúrbios da Comunicação Humana: Campo Fonoaudiológico, Universidade Federal de São Paulo, São Paulo, 1997. Disponível em: <https://core.ac.uk/download/pdf/37718758.pdf>. Acesso em: 24 abr. 2022.

GAYOTTO LH da C. **Dinâmicas de Movimento da Voz**. Revista Distúrbios da Comunicação. 2006; v. 17, n. 3: 41-49.

GOULART, Diana; COOPER, Malu. **Por Todo Canto: método de técnica vocal**. [S.I.]: G4 Edições, 2003.

HARTMANN, William. Acoustic Signal Processing. **Springer Handbook Of Acoustics**, [S.L.], p. 503-530, 2007. Springer New York. http://dx.doi.org/10.1007/978-0-387-30425-0_14.

HLAWATSCH, F.; BOUDREAUX-BARTELS, G.F. Linear and quadratic time-frequency signal representations. **Ieee Signal Processing Magazine**, [S.L.], v. 9, n. 2, p. 21-67, abr. 1992. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/79.127284>.

HOSOKAWA, Kiyohito; OGAWA, Makoto; HASHIMOTO, Michiko; INOHARA, Hide-nori. Statistical Analysis of the Reliability of Acoustic and Electroglottographic Perturbation Parameters for the Detection of Vocal Roughness. **Journal Of Voice**, [S.L.], v. 28, n. 2, p. 9-16, mar. 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.jvoice.2013.07.005>.

JADOUL, Yannick; THOMPSON, Bill; BOER, Bart de. Introducing Parselmouth: a python interface to praat. **Journal Of Phonetics**, [S.L.], v. 71, p. 1-15, nov. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.wocn.2018.07.001>.

KEHTARNAVAZ, Nasser. Frequency Domain Processing. **Digital Signal Processing System Design**, [S.L.], p. 175-196, 2008. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-374490-6.00007-6>.

KELCHNER, Lisa N. et al. Perceptual Evaluation of Severe Pediatric Voice Disorders: rater reliability using the consensus auditory perceptual evaluation of voice. **Journal Of Voice**, [S.L.], v. 24, n. 4, p. 441-449, jul. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.jvoice.2008.09.004>.

KRISHNAN, Sri. Adaptive analysis of biomedical signals. **Biomedical Signal Analysis For Connected Healthcare**, [S.L.], p. 129-156, 2021. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-813086-5.00001-3>.

LAMPERT, Thomas A.; O'KEEFE, Simon E.M.. A survey of spectrogram track detection algorithms. **Applied Acoustics**, [S.L.], v. 71, n. 2, p. 87-100, fev. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.apacoust.2009.08.007>.

MATEO, Carlos; TALAVERA, Juan Antonio. Short-time Fourier transform with the window size fixed in the frequency domain. **Digital Signal Processing**, [S.L.], v. 77, p. 13-21, jun. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.dsp.2017.11.003>.

MONTEIRO, Luisa; FERREIRA, Nuno. **Pertubações nas cordas vocais**. 2011. Disponível em: <https://hmsportugal.wordpress.com/2011/09/09/perturbacoes-das-cordas-vocais-2/>. Acesso em: 02 jun. 2022.

MOURA, Jose. What is signal processing? [President's Message. **Ieee Signal Processing Magazine**, [S.L.], v. 26, n. 6, p. 6-6, nov. 2009. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/msp.2009.934636>.

NURUZZAMAN, A.; BOYRAZ, O.; JALALI, B.. Time-Stretched Short-Time Fourier Transform. **Ieee Transactions On Instrumentation And Measurement**, [S.L.], v. 55, n. 2, p. 598-602, abr. 2006. Institute of Electrical and Electronics Engineers (IEEE).

PINTO, Artur Alves. **PROCESSAMENTO DE SINAIS SONOROS PARA IDENTIFICAÇÃO DE NOTAS MUSICAIS UTILIZANDO O RASPBERRY PI**. 2018. 99 f.

Monografia (Especialização) - Curso de Engenharia Mecatrônica, Centro Federal de Educação Tecnológica de Minas Gerais, Divinópolis, 2012.

PONTES, Paulo A. L. et al. Características das vozes roucas, ásperas e normais: análise acústica espectrográfica comparativa. **Revista Brasileira de Otorrinolaringologia**, [S. L.], v. 68, n. 2, p. 182-191, abr. 2002.

SCHUCK JÚNIOR, Adalberto. **ANÁLISE E PROCESSAMENTO DE SINAIS DE VOZ DISFÔNICA ATRAVÉS DA TRANSFORMADA W A VELET DISCRETA**. 1998. 213 f. Tese (Doutorado) - Curso de Engenharia de Minas, Metalúrgica e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1998.

TONIDANDEL, Danny Augusto Vieira; ARAËJO, Antônio Emílio Angueth de. **Invertendo Domínios: o conceito de transformada**. Rio de Janeiro: Ciência Moderna, 2017.

VIEIRA, Vanessa Pedrosa; BIASE, Noemi de; PONTES, Paulo. Análise acústica e perceptiva au-ditiva versus coaptação glótica em alteração estrutural Mínima. **Acta Orl.** [S.I], p. 174-180. jan. 2006.

ZITTA, Simone Maganhotto. **Análise perceptivo-auditiva e acústica em mulheres com nódulos vocais**. 2005. 111 f. Dissertação (Mestrado) - Curso de Mestrado em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná, Curitiba, 2005.