



# Modelo De Formantes Para Síntese de Vogais

**Bernardo Monteiro - a46149**

**José Cavaleiro - a43936**

Relatório da Unidade Curricular de Aplicações de Processamento de Sinal  
Mestrado em Engenharia Eletrotécnica e de Computadores  
Escola Superior de Tecnologia e Gestão

Supervisão Científica:  
Prof. Dr. João Paulo Teixeira

Bragança  
2025





# Modelo De Formantes Para Síntese de Vogais

**Bernardo Monteiro - a46149**

**José Cavaleiro - a43936**

Relatório da Unidade Curricular de Aplicações de Processamento de Sinal  
Mestrado em Engenharia Eletrotécnica e de Computadores  
Escola Superior de Tecnologia e Gestão

Supervisão Científica:  
Prof. Dr. João Paulo Teixeira

Bragança  
2025



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Aparelho Fonador Humano</b>	<b>3</b>
2.1	Principais Estruturas do Aparelho Fonador . . . . .	3
<b>3</b>	<b>Reprodução e Análise do Som</b>	<b>5</b>
3.1	Análise Acústica com Praat . . . . .	5
<b>4</b>	<b>Desenvolvimento da Produção de Fala</b>	<b>6</b>
4.1	Modelo de Produção de Fala . . . . .	6
4.1.1	Frequência Fundamental ( $F_0$ ) . . . . .	7
4.1.2	Gerador de Impulsos, retirado do software Matlab . . . . .	8
4.1.3	Modelo Matemático do Impulso Glotal ( $G_{(z)}$ ) . . . . .	8
4.1.4	Modelo Matemático do Trato Vocal ( $V_{(z)}$ ) . . . . .	9
4.1.5	Modelo de Radiação Labial ( $R_{(z)}$ ) . . . . .	11
<b>5</b>	<b>Resultados</b>	<b>12</b>
5.1	Sinal final . . . . .	12
5.2	Espectograma . . . . .	13
<b>6</b>	<b>Conclusão</b>	<b>14</b>
	<b>Bibliografia</b>	<b>15</b>
<b>A</b>	<b>Código</b>	<b>A1</b>
<b>B</b>	<b>Reprodução do Som</b>	<b>B1</b>

# Capítulo 1

## Introdução

Quando interagimos, a comunicação verbal destaca-se, sendo a fala uma das formas mais importantes de expressão. Para articular sons de forma clara, é necessário que o nosso sistema vocal funcione corretamente, o que exige a colaboração de diversos órgãos e processos do corpo. Dentro deste sistema, os formantes são fundamentais: eles representam os pontos altos de ressonância no trato vocal, ajudando-nos a distinguir os diferentes sons da fala.

Neste estudo, vamos explicar e investigar o que é o modelo de formantes na produção da fala, analisando as suas bases teóricas e como podemos representá-lo matematicamente. Para isso, detalharemos as principais partes do aparelho fonador e os modelos matemáticos utilizados para simular o seu funcionamento, focando na frequência fundamental, nos geradores de impulsos e nos modelos do trato vocal e da radiação labial.

Compreender como criar e interpretar a fala humana é útil para o desenvolvimento de tecnologias que reconhecem e sintetizam a voz. Assim, este trabalho tem como objetivo ajudar a perceber como as vogais são produzidas e a sua importância para o estudo da fala e da comunicação.

# Capítulo 2

## Aparelho Fonador Humano

Neste capítulo, exploramos a anatomia e o funcionamento do aparelho fonador, destacando os principais órgãos envolvidos e o seu papel na produção do som.

### 2.1 Principais Estruturas do Aparelho Fonador

Quando falamos da produção de som, é essencial compreendermos como se processa a produção da fala, pois, no nosso corpo, existem alguns órgãos ou partes deles responsáveis por funções fundamentais para a emissão do som. Na Figura 2.1, podemos observar alguns dos mais importantes, tais como:

- **Pulmões** – fornecem o fluxo de ar necessário para a produção do som;
- **Pregas vocais (cordas vocais, situadas na laringe)** – vibram para gerar sons sonoros ou abrem-se completamente para produzir sons como o “s”;
- **Faringe** – atua como ressonador, ajudando a moldar o som proveniente das pregas vocais;
- **Cavidade nasal** – intervém na produção de sons nasais, como o “m”.

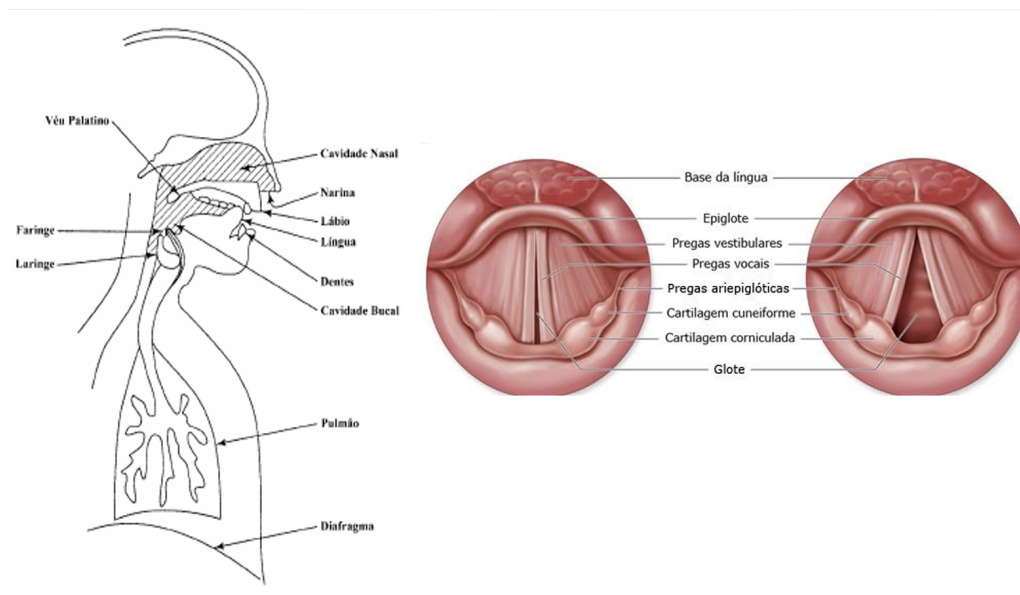


Figura 2.1: Esquema do aparelho fonador humano.



# Capítulo 3

## Reprodução e Análise do Som

Este capítulo aborda a reprodução e análise do som, com foco na utilização do software Praat para examinar características acústicas das vogais.

### 3.1 Análise Acústica com Praat

Para o desenvolvimento deste projeto, é essencial considerar particularidades que não são perceptíveis apenas pela audição. Características acústicas únicas, como os formantes e suas respectivas larguras de banda, não podem ser identificadas sem o auxílio de ferramentas específicas.

Neste caso foi utilizado o software Praat, onde foram gravadas as cinco vogais, permitindo a análise desses parâmetros. A Figura 3.1 apresenta o espectrograma da vogal “a”, no qual é possível observar suas características acústicas.

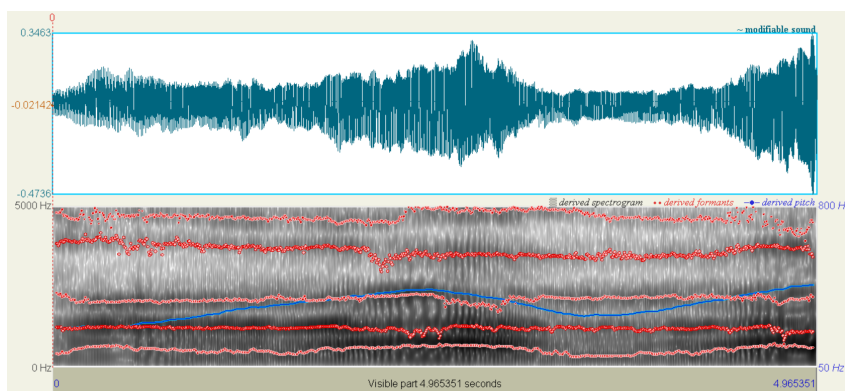


Figura 3.1: Características acústicas vogal "a", retirado do software Praat

# Capítulo 4

## Desenvolvimento da Produção de Fala

Este capítulo aborda o desenvolvimento do modelo de produção de fala, detalhando os componentes essenciais, como o impulso glotal, o trato vocal e a radiação labial.

### 4.1 Modelo de Produção de Fala

A produção da fala pode ser modelada matematicamente através de diferentes abordagens, cada uma procurando representar com maior precisão o funcionamento do aparelho fonador humano. O modelo de produção da fala baseia-se na interação entre três componentes principais: **modelo do impulso glotal**  $G(z)$ , o **trato vocal**  $V(z)$  e a **radiação labial**  $R(z)$ . A Figura 4.1 demonstra o modelo completo da produção da fala onde estão identificadas as 3 componentes mais importantes, para o nosso caso estamos a excluir a parte do gerador do ruído.

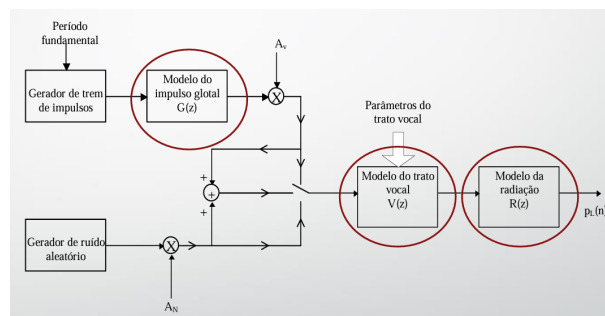


Figura 4.1: Modelo da produção da fala completo

Em termos matemáticos, o modelo da produção da fala pode ser descrito pela equação:

$$H(z) = G(z)V(z)R(z) \quad (4.1)$$

onde:

- $G(z)$  representa o modelo do impulso glotal;
- $V(z)$  é o modelo do trato vocal;
- $R(z)$  corresponde ao modelo de radiação nos lábios.

#### 4.1.1 Frequência Fundamental ( $F_0$ )

A frequência fundamental vocal ( $F_0$ ) é a medida da vibração das cordas vocais em Hertz e reflete a eficiência do sistema fonatório. Varia com fatores como idade, sexo, estado emocional e hábitos de vida. Geralmente, a  $F_0$  das mulheres é mais alta do que a dos homens. Estudos mostram que a média da  $F_0$  nem sempre distingue vozes patológicas das normais, mas a sua variabilidade pode ser útil na avaliação de problemas vocais [1].

O nosso sinal  $F_0$  foi representada com o valor médio de 175 e a sua amplitude é de 75, tal como podemos ver na equação 4.2 e na Figura 4.2 onde em baixo representados.

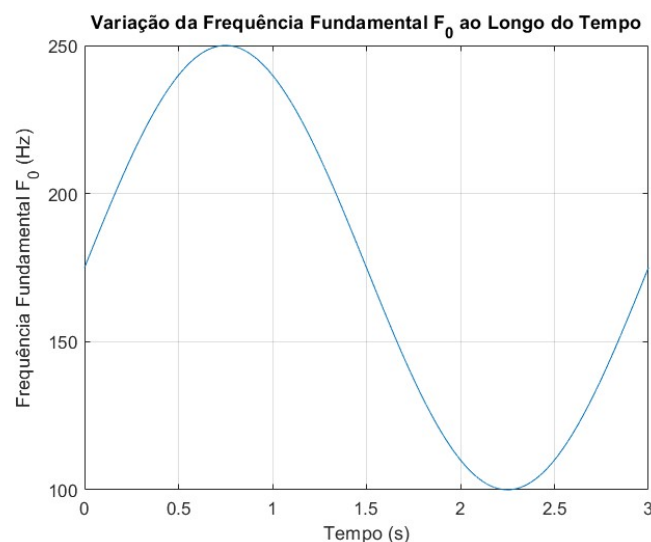


Figura 4.2: Frequência Fundamental, retirado do software Matlab

### Equação da frequência fundamental:

$$F_0 = 175 + 75 \times \sin(2\pi \times n \times T \times 1/3) \quad (4.2)$$

#### 4.1.2 Gerador de Impulsos, retirado do software Matlab

O gerador de impulsos glotais gera uma sequência de impulsos espaçados por um intervalo de tempo definido por  $T_0 = \frac{1}{F_0}$ . Este gerador é um modelo que simula a voz humana, pois este representa as cordas vocais, a abrir e fechar em pulsos unitários.

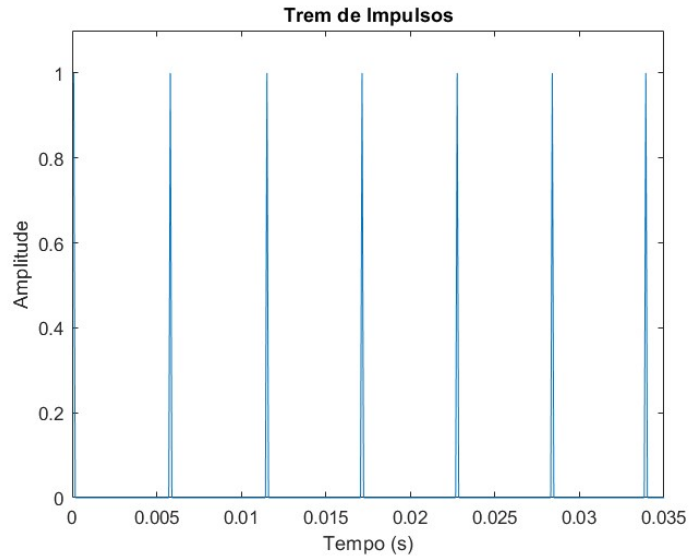


Figura 4.3: Trem de Impulsos, retirado do software Matlab

#### 4.1.3 Modelo Matemático do Impulso Glotal ( $G_{(z)}$ )

O modelo matemático do impulso glotal, é uma parte fundamental na síntese da fala, pois este é responsável por simular a vibração gerada pelas cordas vocais durante a produção de som. Para realizar este modelo é necessário utilizar os impulsos unitários do gerador (4.1.2) e aplicar a equação 4.3. Dependendo do valor de “a”, deve gerar uma sequência de impulsos glotais com as formas da Figura 4.4 Cada forma reproduzirá um sinal com diferente timbre de voz [2].

$$G_{(z)} = \frac{-aeln(a)z^{-1}}{(1 - az^{-1})^2} \quad (4.3)$$

O valor de “ $a$ ” escolhido para produzir os nossos impulsos glotais foi o  $a = 0.85$ .

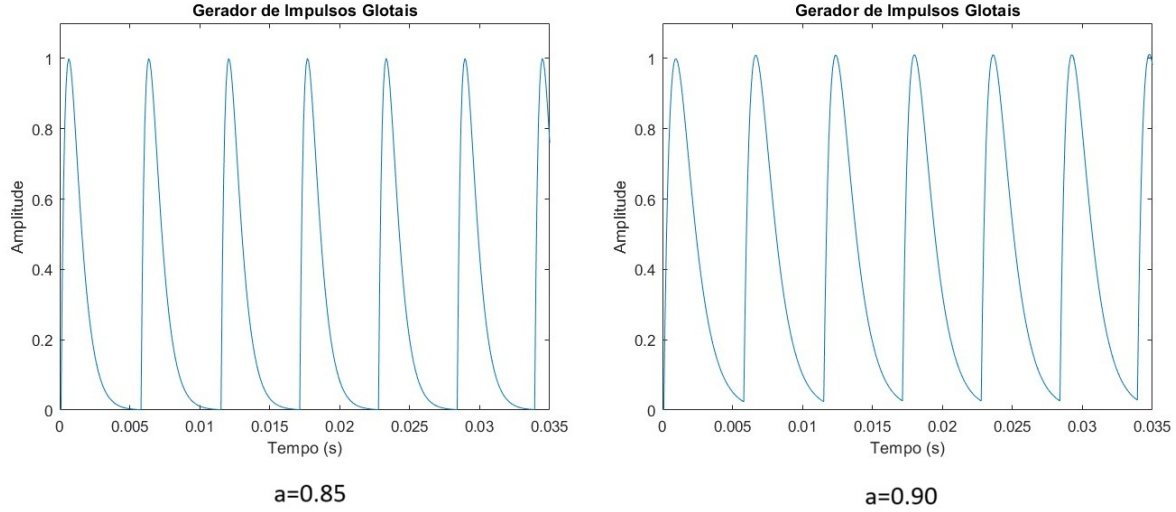


Figura 4.4: Impulsos glotais com diferentes valores de “ $a$ ”, retirado do software Matlab

#### 4.1.4 Modelo Matemático do Trato Vocal ( $V_{(Z)}$ )

Para a realização deste modelo, é necessário conhecer os formantes e as respetivas larguras de banda de cada vogal, que se encontram apresentados na Tabela 4.1. Estes valores foram obtidos através do software Praat.

	<b>a</b>	<b>e</b>	<b>i</b>	<b>o</b>	<b>u</b>
<b>F1 (Hz)</b>	617.19	496.08	347.05	567.16	386.63
<b>B1 (Hz)</b>	169.37	70.05	31.62	97.16	16.02
<b>F2 (Hz)</b>	1158.21	1896.12	1953.07	889.32	801.66
<b>B2 (Hz)</b>	474.45	322.23	1137.30	93.68	174.54
<b>F3 (Hz)</b>	2206.21	2161.53	2320.75	2269.36	2458.37
<b>B3 (Hz)</b>	177.70	791.66	135.09	96.33	120.61
<b>F4 (Hz)</b>	3531.82	3734.36	3842.94	3370.87	3745.85
<b>B4 (Hz)</b>	669.63	230.06	163.02	153.93	204.35

Tabela 4.1: Tabela de formantes e larguras de banda das vogais analisadas.

Ao obter os formantes, aplicou-se a equação 4.4 para cada valor da Tabela 4.1. Ou seja, para a vogal "a", realizou-se a filtragem sucessiva do sinal de entrada utilizando um

filtro digital correspondente a cada formante e respetiva largura de banda. O sinal inicial, um impulso unitário, foi processado através de um sistema de quatro filtros ressonantes em série, cada um modelando um formante da vogal [2].

Em cada iteração do modelo, o numerador e o denominador do filtro foram calculados com base na frequência do formante e na sua largura de banda, seguindo a equação 4.4. Esse filtro destaca as frequências das ressonâncias naturais do trato vocal, simulando a estrutura espectral da vogal desejada. Por fim, o sinal foi normalizado para manter a amplitude dentro dos limites adequados, evitando distorções e garantindo uma representação correta do som sintetizado.

$$V_k(z) = \frac{1 - 2 \cdot \left| e^{-\frac{B_k \cdot T}{2}} \right| \cdot \cos(2\pi \cdot F_k \cdot T) + |e^{-B_k \cdot T}|}{1 - 2 \cdot \left| e^{-\frac{B_k \cdot T}{2}} \right| \cdot \cos(2\pi \cdot F_k \cdot T) \cdot z^{-1} + |e^{-B_k \cdot T}| \cdot z^{-2}} \quad (4.4)$$

O sinal resultante do modelo matemático do trato vocal está representado na Figura 4.5. A aplicação deste modelo permitiu a conversão do trem de impulsos, ilustrado na Figura 4.4, em um sinal com características acústicas mais próximas da fala humana.

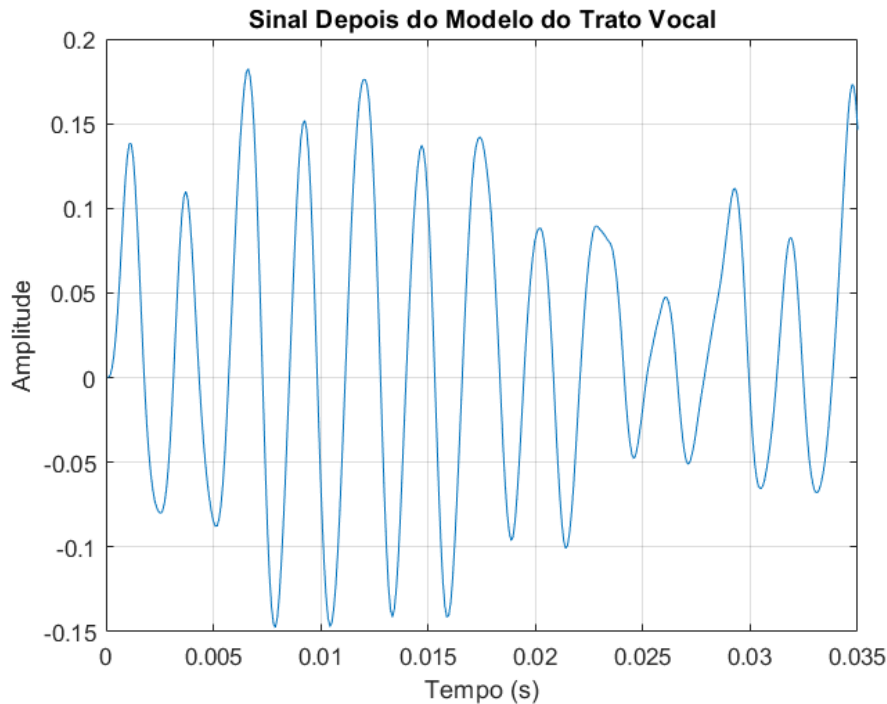


Figura 4.5: Sinal Obtido depois do trato Vocal, retirado do software Matlab

#### 4.1.5 Modelo de Radiação Labial ( $R_{(z)}$ )

O modelo de radiação nos lábios é representado pela fórmula matemática da 4.5, onde a mesma descreve como o som é reproduzido a partir dos lábios de uma pessoa.

O objetivo ao filtrarmos o sinal do modelo de radiação labial, é de ajustar ou igualar o sinal capturado de forma a melhorar a sua qualidade e torná-lo mais realista, isso é, torná-lo o mais parecido possível com a fala humana.

$$R_{(z)} = 1 - z^{-1} \tag{4.5}$$

# Capítulo 5

## Resultados

Este capítulo apresenta os resultados obtidos da simulação da produção de fala, incluindo o sinal final gerado e o seu espectrograma.

### 5.1 Sinal final

Através do código elaborado no software Matlab, Anexo [A](#), obtivemos o seguinte sinal final que está representado na Figura [5.1](#). Neste é possível observar, num intervalo de tempo reduzido, uma forma de onda que se assemelha ao espectro de uma gravação de voz. O áudio pode ser reproduzido no Anexo [B](#).

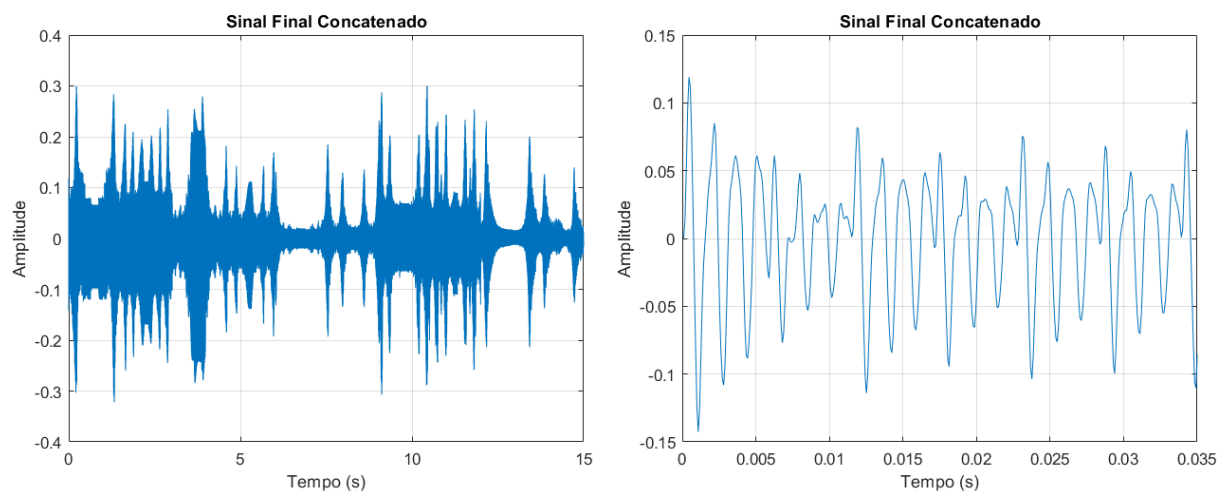


Figura 5.1: Sinal Final Obtido, retirado do software Matlab



## 5.2 Espectrograma

O espectrograma apresentado na Figura 5.2 representa a distribuição espectral do sinal de voz concatenado, contendo as vogais *a*, *e*, *i*, *o*, *u*. No eixo vertical, observa-se a frequência em  $kHz$ , enquanto o eixo horizontal indica a evolução temporal das vogais. A intensidade das componentes espectrais é representada por uma escala de cores, onde regiões em amarelo indicam maior potência do sinal, e azul, menor potência.

As bandas mais destacadas correspondem aos formantes, que são picos de ressonância do trato vocal e caracterizam cada vogal. Nota-se que as vogais apresentam diferentes padrões espectrais, refletindo as diferenças nas configurações articulatórias durante a sua produção. As linhas verticais tracejadas marcam as transições entre as vogais.

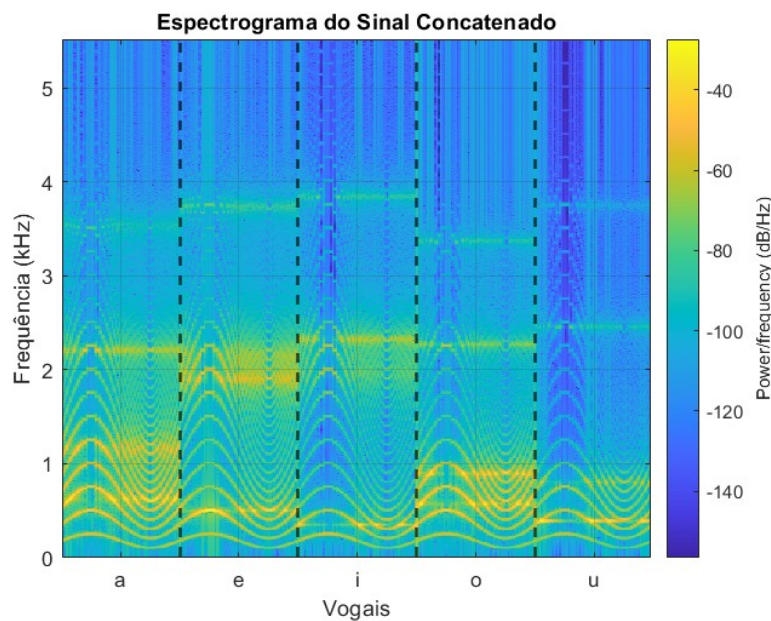


Figura 5.2: Espectrograma das 5 vogais, retirado do software Matlab

# Capítulo 6

## Conclusão

Ao examinar o padrão dos formantes, conseguimos desvendar a intrincada natureza da emissão da fala e o quão importante é o canal vocal na formatação do som. A investigação do sistema fonético humano, juntamente com métodos matemáticos e instrumentos como o programa Praat, permitiu a identificação das qualidades sonoras das vogais e a réplica computacional da sua criação.

Constatou-se que os formantes são cruciais na forma como ouvimos as vogais, sendo definidos pelas ressonâncias do canal vocal. Esquemas matemáticos, como os que especificam a frequência fundamental ( $F_0$ ), o gerador de impulsos, o canal vocal e a irradiação labial, possibilitam aproximar a emissão artificial da voz humana.

Esta análise enfatiza a relevância da compreensão da produção da fala tanto para a pesquisa linguística quanto para a criação de tecnologias de fala, como plataformas de síntese de voz e reconhecimento vocal.

# Bibliografia

- [1] João Paulo Teixeira, Débora Ferreira, and Susana Moreira Carneiro. Análise acústica vocal-determinação do jitter e shimmer para diagnóstico de patologias da fala. In *6<sup>o</sup> Congresso Luso-Moçambicano de Engenharia, 3<sup>o</sup> Congresso de Engenharia de Moçambique*, number 6<sup>o</sup>. INEGI, 2011.
- [2] João Paulo Teixeira. Modelo de formantes para síntese de vogais, 2025.

# Apêndice A

## Código

O Anexo A, denominado como “[Vogais](#)“, está disponível em ficheiro de Matlab e texto. Este ficheiro contém o código realizado para a obtenção do som das 5 Vogais

# Apêndice B

## Reprodução do Som

O Anexo B, denominado como “[Sinal Voz](#)“, está disponível em ficheiro de reprodução wav.