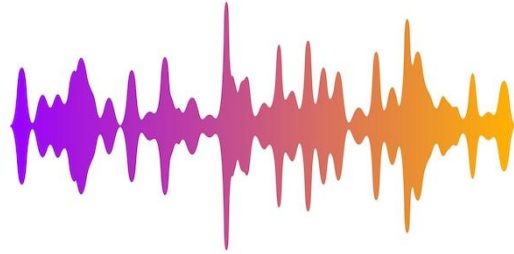


# Análise e Processamento de Sinais de Áudio



Disciplina: Processamento de Sinais  
Professora Regente: Carla Diniz Lopes Becker  
Mestranda: Dieine Bernieri Schiavon

# Currículo

---



- Bacharel em Informática Biomédica (UFCSPA)
- Mestranda em TI e Gestão em Saúde (UFCSPA)
- Bolsista Capes
- Pós-Graduanda em Eng. de Software Aplicada a Ciência de Dados (Instituto de Informática - UFRGS)
- Fonoaudióloga, especialista em Linguagem (UFRGS)



(11) Dieine Bernieri Schiavon | LinkedIn

dieineb@ufcspa.edu.br  
dieine.schiavon@ufrgs.br



<http://lattes.cnpq.br/2974158130212395>

# Resumo

---



## Parte 1: Transformada de Fourier de Tempo Curto (STFT)

- **Introdução à STFT:** análise de sinais no domínio do tempo e frequência
- **Aplicações:** como a STFT é utilizada na análise de áudio

# Resumo



---

## Parte 2: Geração de Espectrogramas com a Biblioteca Python Librosa

- **Introdução ao Librosa:** Apresentação da biblioteca, suas funcionalidades e vantagens para a análise de áudio
- **Passo a Passo:** Demonstração prática no Notebook Colab sobre como gerar espectrogramas a partir de sinais de áudio

# Resumo

---



## Parte 3 - Casos de Uso

- **Aplicações em Deep Learning:** uso de espectrogramas como entrada para modelos na classificação de sinais biológicos

# Análise e Processamento de Sinais de Áudio

---



## Parte 1: Transformada de Fourier de Tempo Curto (STFT)

Parte 2: Geração de Espectrogramas com a Biblioteca Python Librosa

Parte 3 - Casos de Uso

# Transformada Discreta de Fourier - DFT

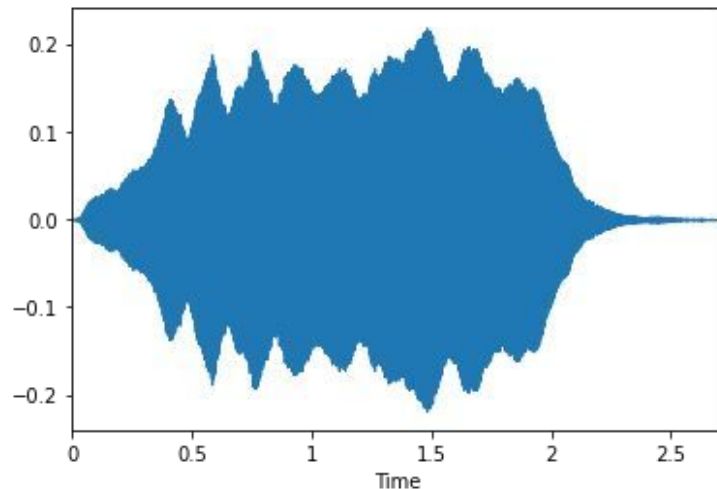
---



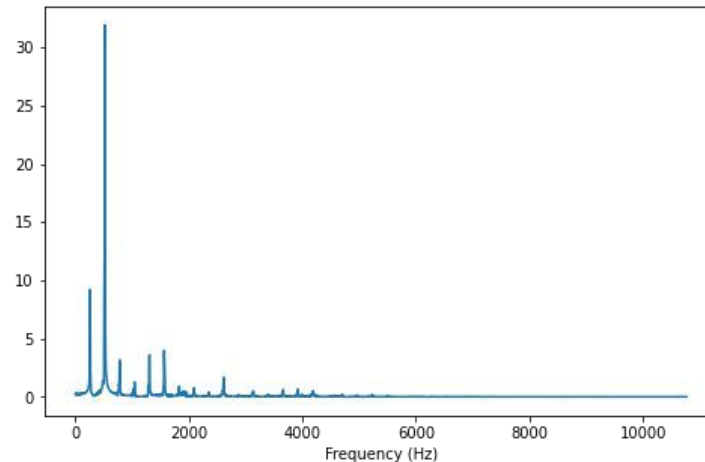
- Transforma os sinais no domínio do tempo em componentes do domínio da frequência
- Sequências finitas (ex: sinal de voz, fala, deglutição, respiração, batimentos cardíacos)

# Transformada Discreta de Fourier - DFT

---



DFT





# DFT X FFT

---



- DFT estabelece uma relação entre a representação no domínio do tempo X frequência
- *Fast Fourier Transform*, ou FFT, é um algoritmo computacional que reduz o tempo de computação e a complexidade de grandes transformações

# DFT X FFT

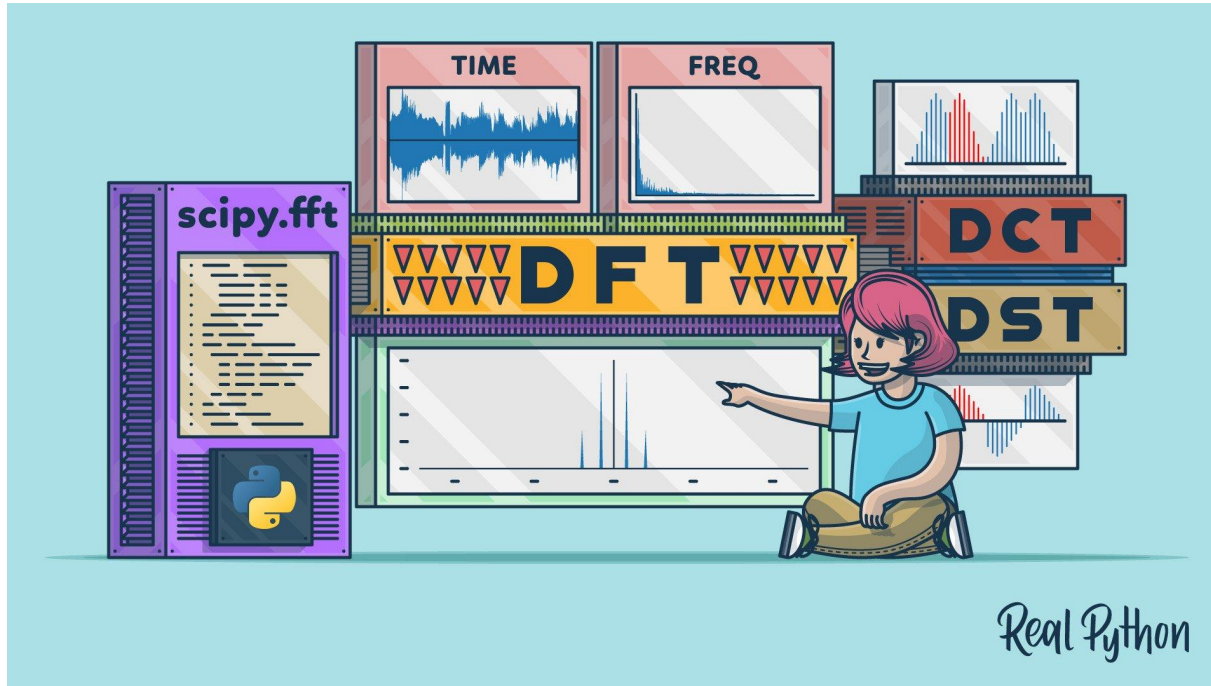


- A FFT reduz a complexidade computacional da DFT de:

Algoritmo	Complexidade
DFT	$O(N^2)$
FFT	$O(N \log_2 N)$

- A equação fundamental permanece a mesma, mas a FFT reestrutura o cálculo para ser mais eficiente computacionalmente

# DFT X FFT

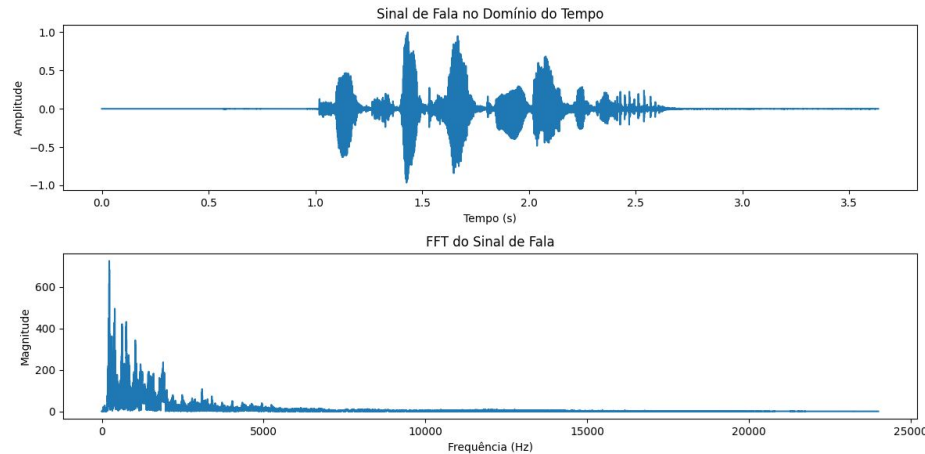


Ref: Fourier Transforms With scipy.fft: Python Signal Processing – Real Python



# Problema da Transformada Discreta de Fourier

**SABEMOS O QUÊ (Frequências)**

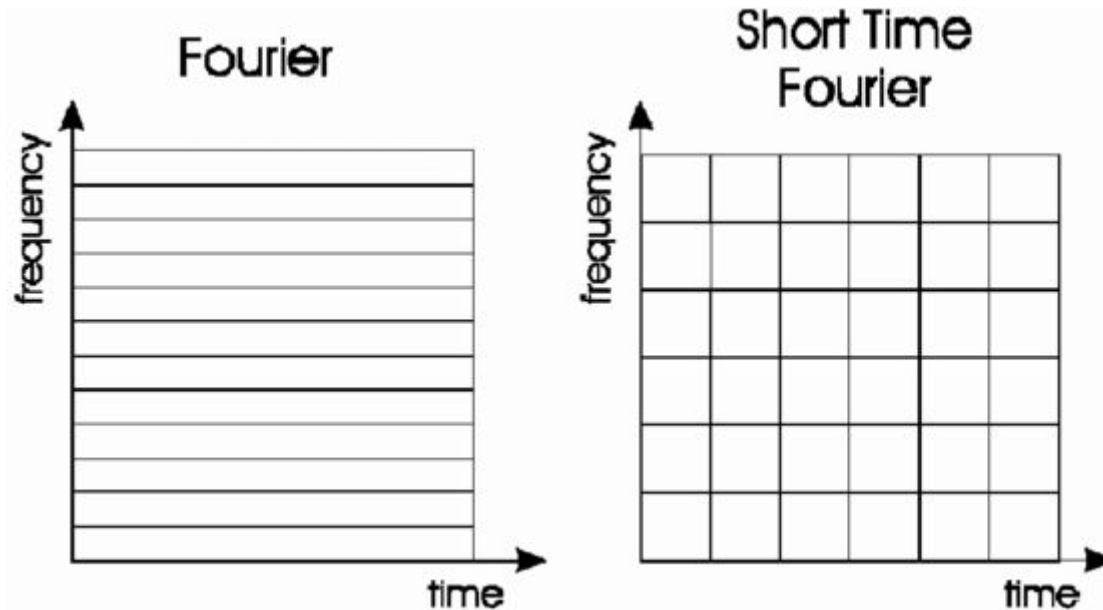


**MAS NÃO SABEMOS QUANDO  
(Tempo)**

# DFT – FFT - STFT



- A principal desvantagem da DFT é que não distingue efeitos variantes no tempo. Tornando a DFT inadequada para sinais cujas propriedades espectrais mudam com o tempo





Ref: <https://github.com/musikalkemist>



# STFT

---

- A STFT segmenta sinais não estacionários em janelas de tempo curtas, onde o sinal é tratado como estacionário
- Permite analisar a evolução das frequências ao longo do tempo

# STFT – Transformada de Fourier de curta duração

A STFT representa um sinal do domínio do tempo-frequência por meio de cálculo de DFT em janelas curtas sobrepostas

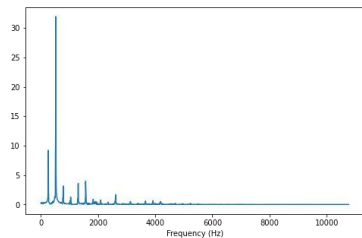
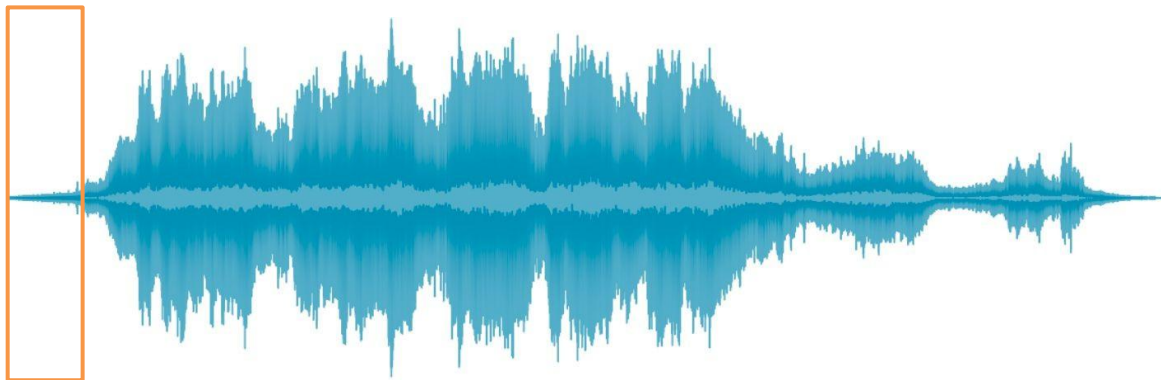


Ref:

[https://www.researchgate.net/figure/Figura-3-Transformacao-de-um-sinal-atraves-da-tecnica-STFT-Diferente-da\\_fig1\\_305983467](https://www.researchgate.net/figure/Figura-3-Transformacao-de-um-sinal-atraves-da-tecnica-STFT-Diferente-da_fig1_305983467)

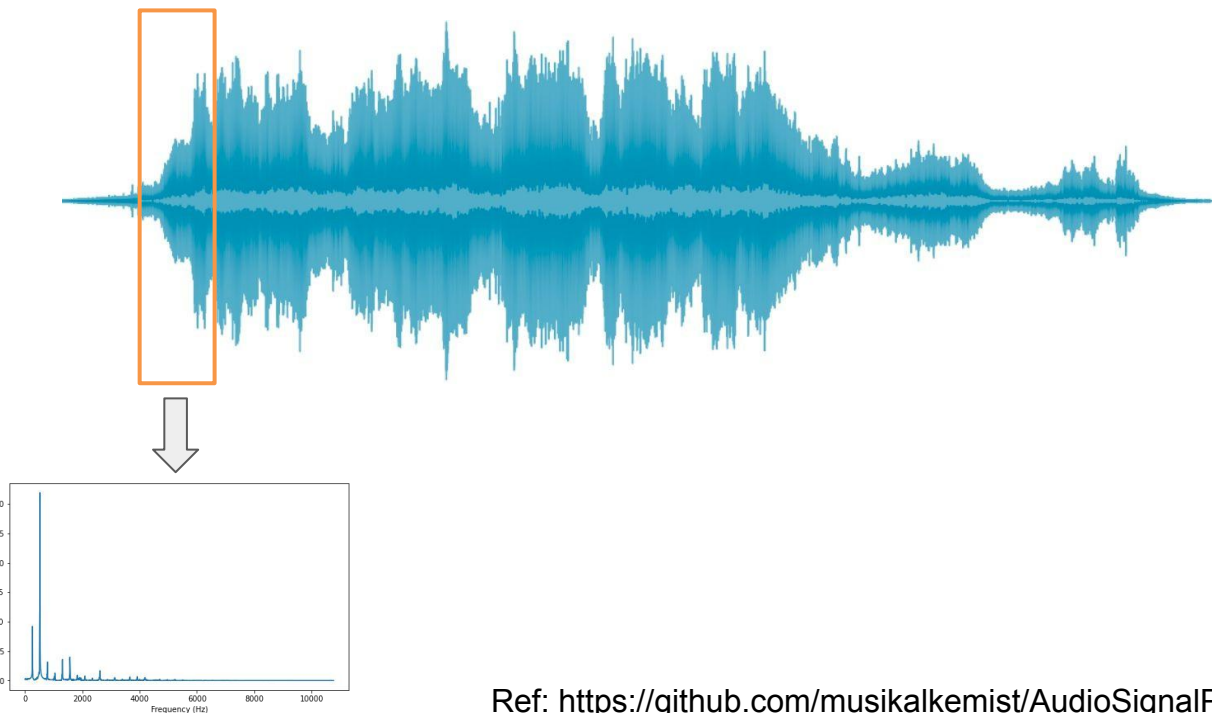


# STFT



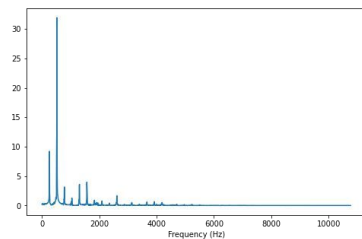
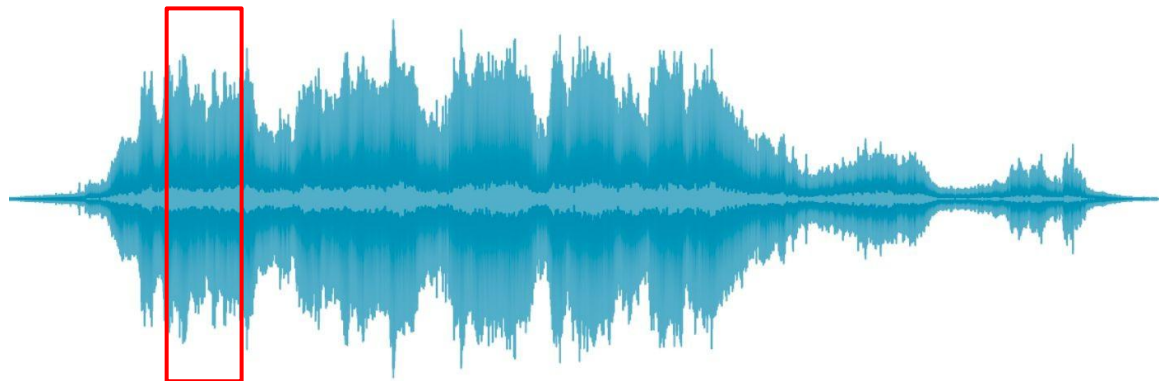
Ref: <https://github.com/musikalkemist/AudioSignalProcessingForML>

# STFT intuition



Ref: <https://github.com/musikalkemist/AudioSignalProcessingForML>

# STFT



Ref: <https://github.com/musikalkemist/AudioSignalProcessingForML>

# Windowing (Janelamento)

---



- Aplicar função de janelamento ao sinal

$$x_w(k) = x(k) \cdot w(k)$$

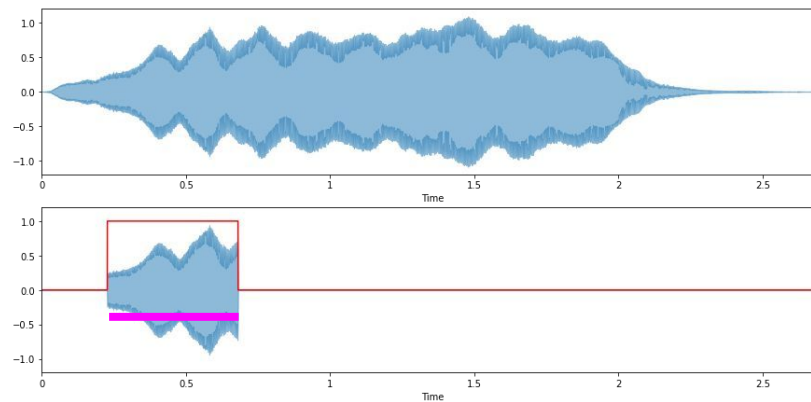
Onde:

$x(k)$  sinal original

$w(k)$  é a função de janelamento

$x_w(k)$  é o sinal janelado

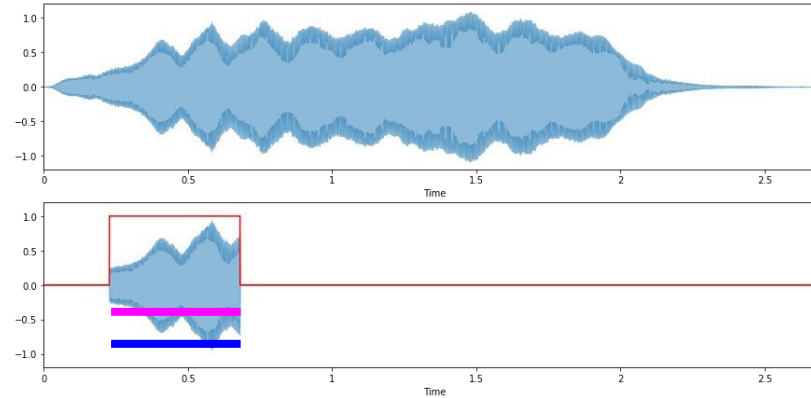
# Windowing (Janelamento)



window size

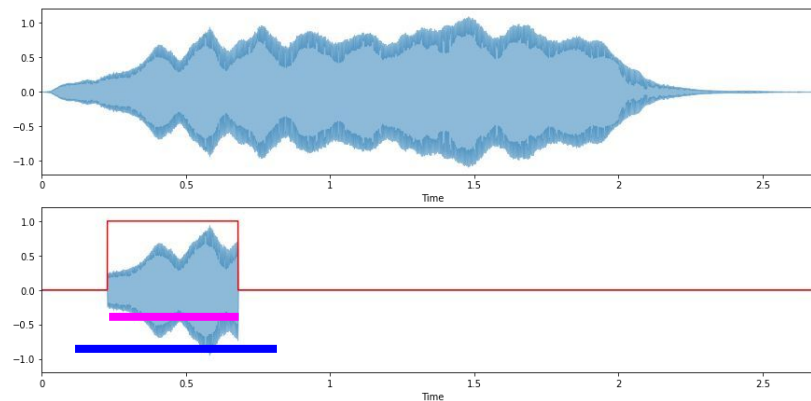
Ref: <https://github.com/musikalkemist/AudioSignalProcessingForML>

# Windowing (Janelamento)



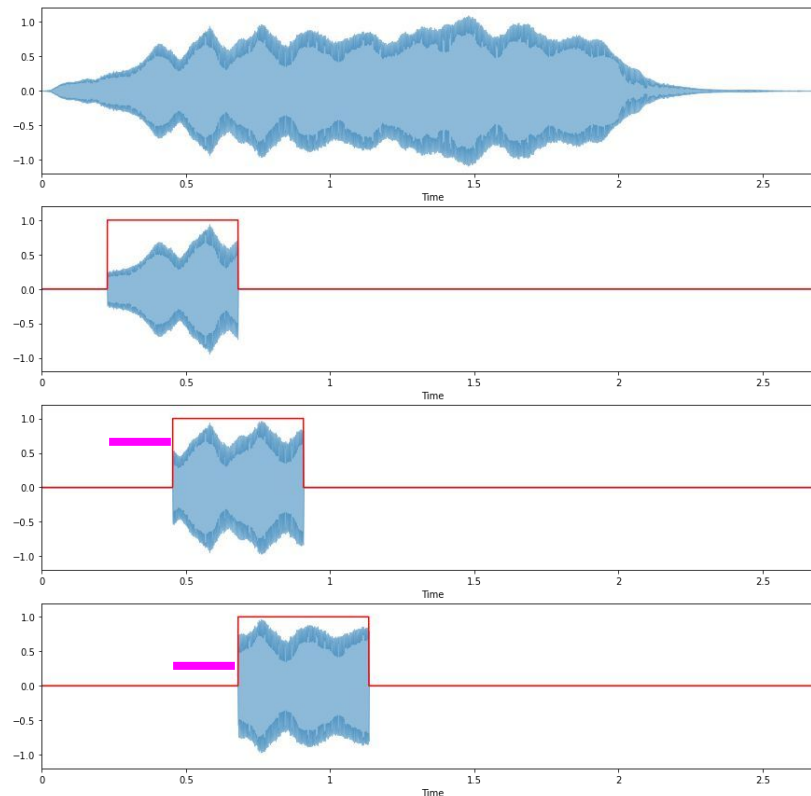
window size = frame size

# Windowing (Janelamento)



window size  $\neq$  frame size

# Overlapping (sobreposição) frames



hop size (H)  
Passo



# De DFT para STFT



**DFT** 
$$X(f) = \sum_{n=0}^{N-1} x(n) \cdot e^{-i \frac{2\pi f n}{N}}$$

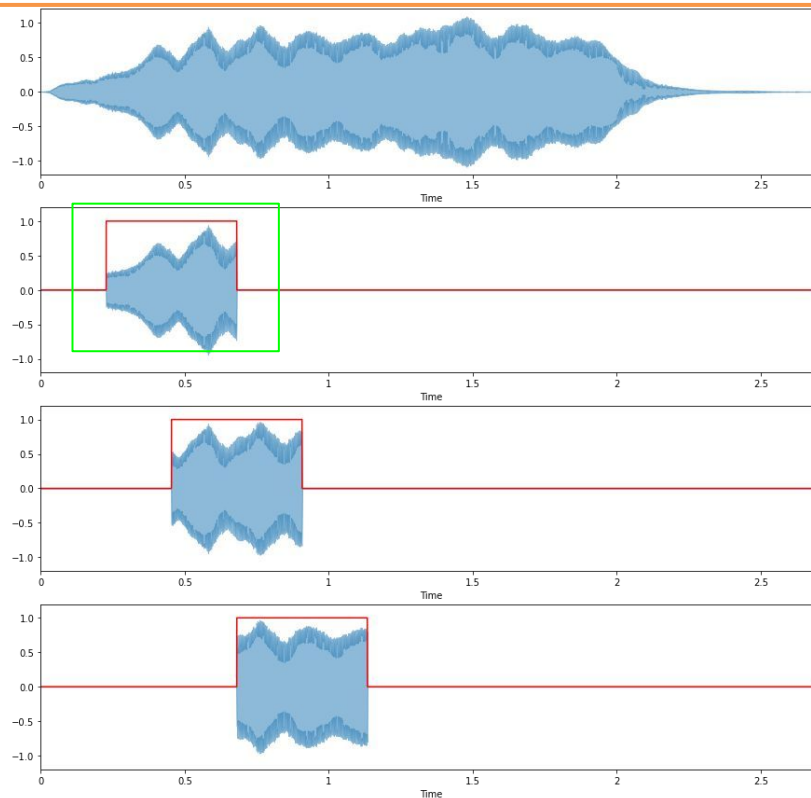
**STFT** 
$$X(\boxed{t}, f) = \sum_{n=0}^{N-1} x(n) \cdot \boxed{w(n - t)} \cdot e^{-i \frac{2\pi f n}{N}}$$

O processo de aplicar a janela ao sinal pode ser interpretado como uma convolução, pois estamos essencialmente "filtrando" o sinal através da janela antes de aplicar a transformada de Fourier

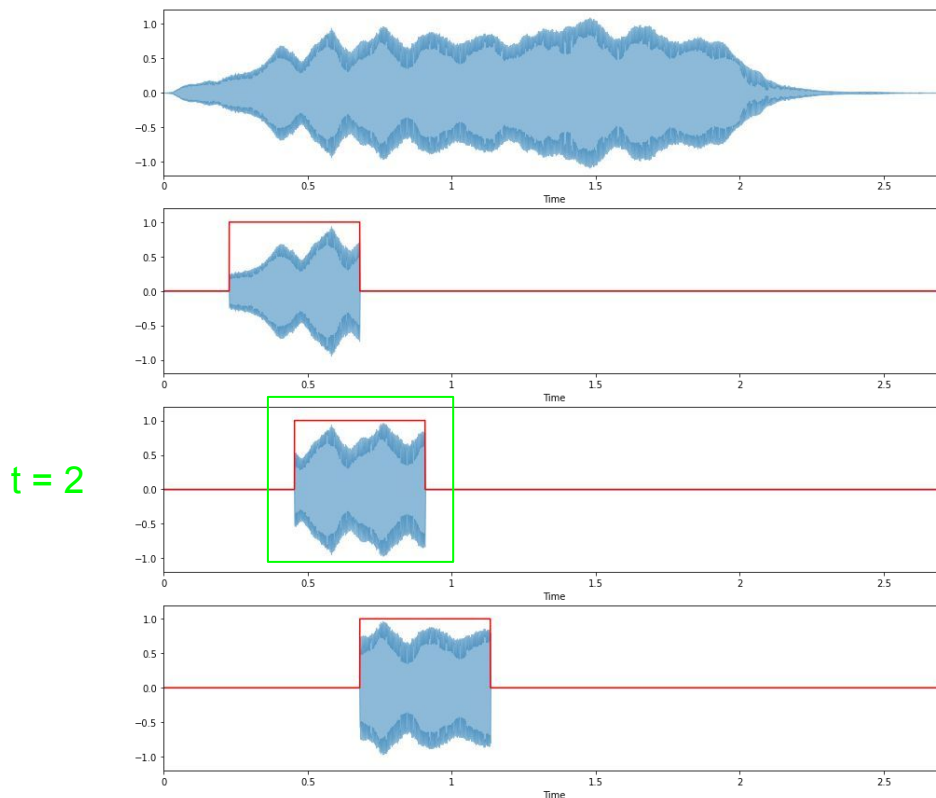
# De DFT para STFT



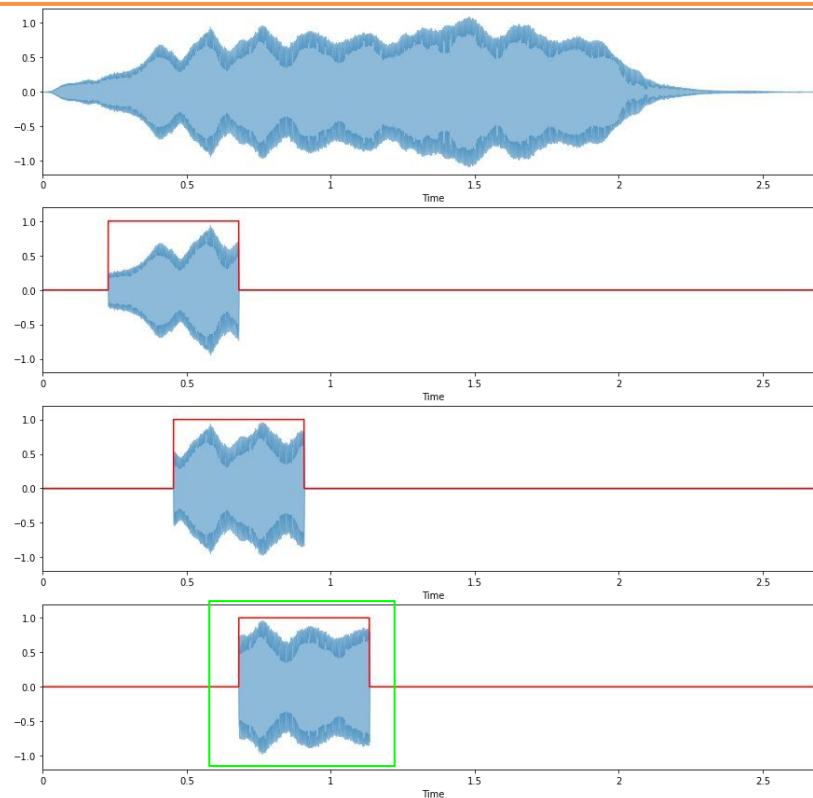
$t = 1$



# De DFT para STFT



# De DFT para STFT



# Resumindo...

---



- **DFT:** transforma uma sequência finita de valores no tempo em uma sequência finita de valores no domínio da frequência
- **STFT:** analisa um sinal não estacionário ao aplicar a FFT em segmentos (janelas) curtos do sinal

# Semelhanças e Diferenças

---



- **Semelhanças:** Ambas transformadas utilizam somas de exponenciais complexas para decompor o sinal em componentes de frequência
- **Diferenças:** A DFT transforma o sinal todo de uma vez, assumindo que ele é estacionário, enquanto a STFT aplica a FFT a segmentos curtos do sinal para capturar mudanças temporais na frequência, sendo útil para sinais não estacionários

# STFT Parâmetros



- Frame size

**512, 1024, 2048, 4096, 8192**

$$512 = 2^9$$

$$1024 = 2^{10}$$

$$2048 = 2^{11}$$

$$4096 = 2^{12}$$

$$8192 = 2^{13}$$

Potências  
de 2

Os valores de tamanho de janela (**frame size**) na STFT são escolhidos com base em considerações de resolução tempo e frequência, bem como eficiência computacional

# Time / frequency trade off

---



frame  
size



Frequência  
Resolução

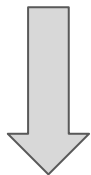


Resolução  
Temporal

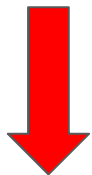


# Time / frequency trade off

---



frame  
size



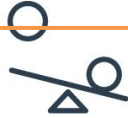
Frequência  
Resolução



Resolução  
Temporal

# Time / frequency trade off

---



O **"trade-off"** envolve avaliar esses pontos de perda e ganho para encontrar um equilíbrio que atenda melhor às necessidades específicas da análise de sinais de áudio em uma aplicação particular

# STFT Parâmetros

---



- Frame size
- Hop size

256, **512**, 1024, 2048, 4096

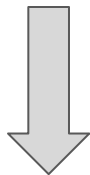
# STFT Parâmetros

---



- Frame size

- Hop size



Passo menor



Resolução  
Temporal



Custo  
Computacional

# STFT parameters

---

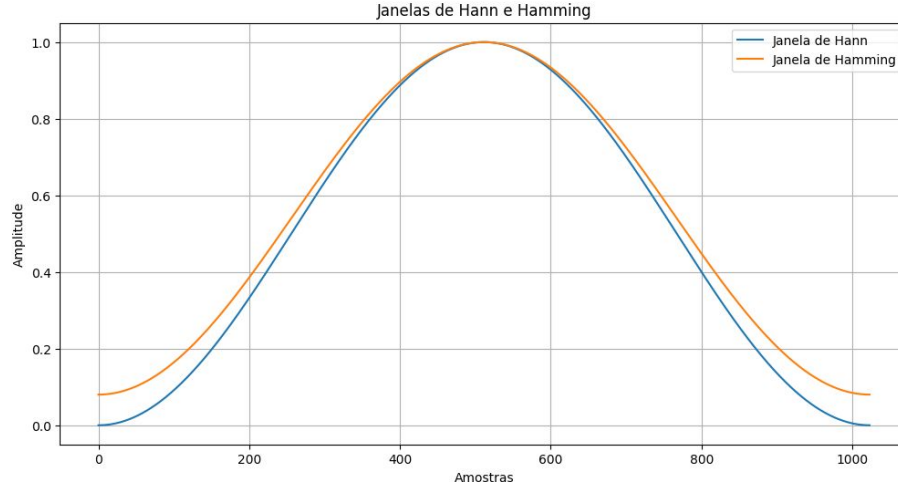


- Frame size
- Hop size
- Windowing function

# Hann e Hamming window



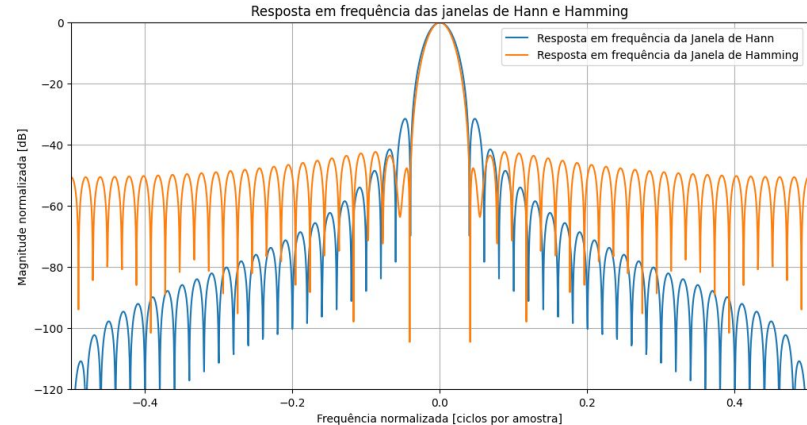
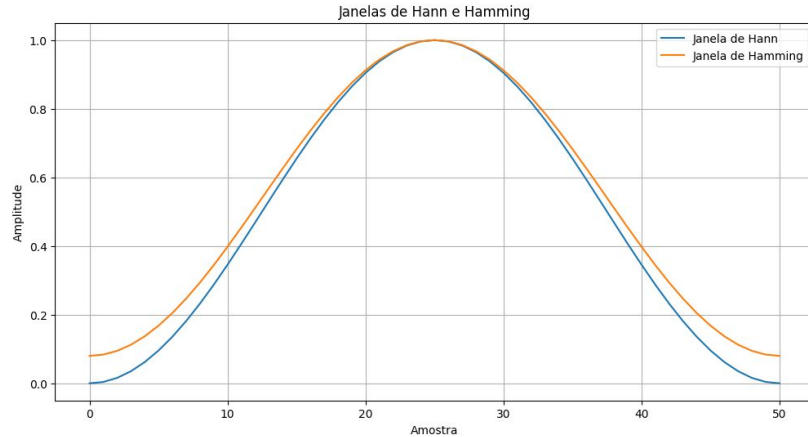
As funções de **Hann** e **Hamming** são normalmente usadas como uma função de janela em processamento de sinal digital para selecionar um subconjunto de uma série de amostras, a fim de realizar uma transformação de Fourier



# Hann e Hamming window



A janela de **Hann** toca zero em ambas as extremidades, eliminando toda a descontinuidade. A janela de **Hamming** não chega a zero e, portanto, ainda apresenta uma ligeira descontinuidade (vazamento espectral) no sinal



# Análise e Processamento de Sinais de Áudio

---



Parte 1: Transformada de Fourier de Tempo Curto (STFT)

**Parte 2: Geração de Espectrogramas com a Biblioteca Python Librosa**

Parte 3 - Casos de Uso



# Visualizando som

---

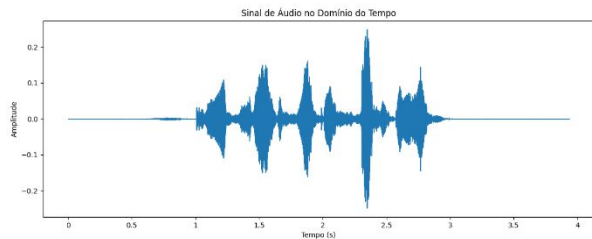


$$S(t, f) = |X(t, f)|^2$$

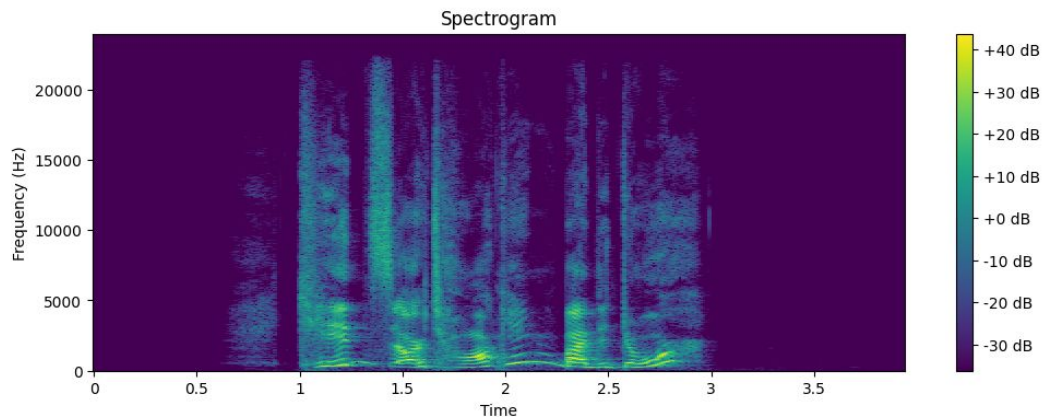
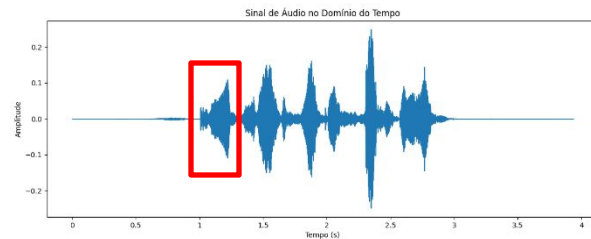
Nesta notação:

- $X(t, f)$  é a STFT do sinal no tempo  $t$  e na frequência  $f$ .
  - $S(t, f)$  representa o espectrograma, que é a magnitude ao quadrado da STFT
- $S(t, f)$  indica a energia espectral do sinal em função do tempo **t** e da frequência **f**

# Espectrograma



**STFT**



# Espectrograma

---



Bibliotecas Python para gerar espectrogramas:

## Matplotlib

- `matplotlib.pyplot.specgram`: Gera um espectrograma usando a STFT (Short-Time Fourier Transform) e plota usando Matplotlib.

## Scipy

- `scipy.signal.spectrogram`: Calcula e retorna um espectrograma usando a STFT.
- `scipy.signal.stft`: Calcula a STFT (utilizada internamente pelo `spectrogram`) e retorna a matriz de valores.
- `scipy.signal.spectrogram`: Gera um espectrograma usando a STFT.

# Espectrograma

---



Bibliotecas Python para gerar espectrogramas:

## Librosa

- `librosa.stft`: Calcula a STFT de um sinal de áudio.
- `librosa.feature.melspectrogram`: Calcula um espectrograma mel (uma representação mais perceptualmente relevante de áudio).
- `librosa.amplitude_to_db(np.abs(D), ref=np.max)` converte os valores de magnitude da STFT para dB, utilizando o máximo valor como referência.

O uso de `amplitude_to_db` é adequado para visualizar a magnitude absoluta dos coeficientes da STFT

# Espectrograma

---

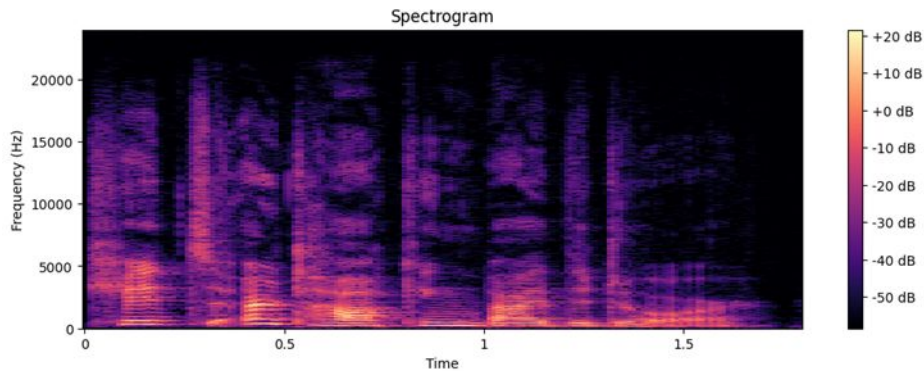


``power_to_db``

A função ``power_to_db`` converte valores de potência em decibéis (dB). É especialmente útil quando trabalhamos com espectrogramas baseados na potência dos coeficientes da STFT

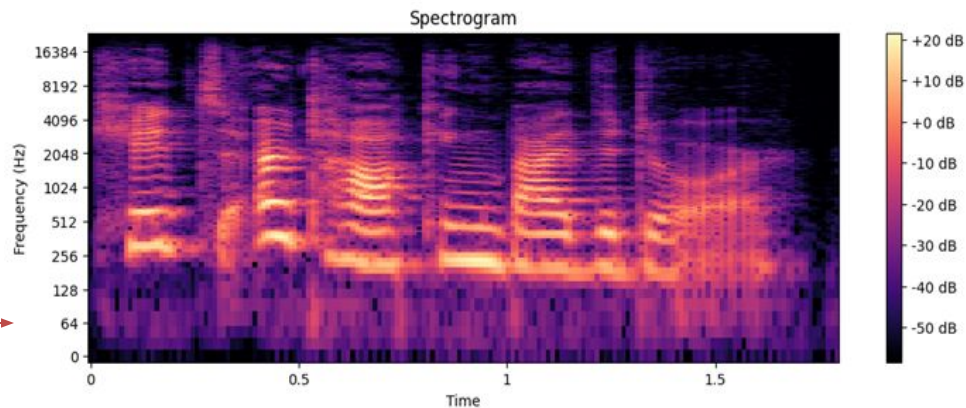
**`power_to_db`** fornece informações sobre a intensidade do sinal ao longo do tempo ou em uma região específica do espectro

# Espectrograma



**amplitude\_to\_db**

**power\_to\_db**



# Parte Prática – Notebook

---



Colab:

<https://colab.research.google.com/drive/1OE-qnVwzUbyaFTtu2mWp8d3T2ALx1Yye?usp=sharing>



# Análise e Processamento de Sinais de Áudio

---



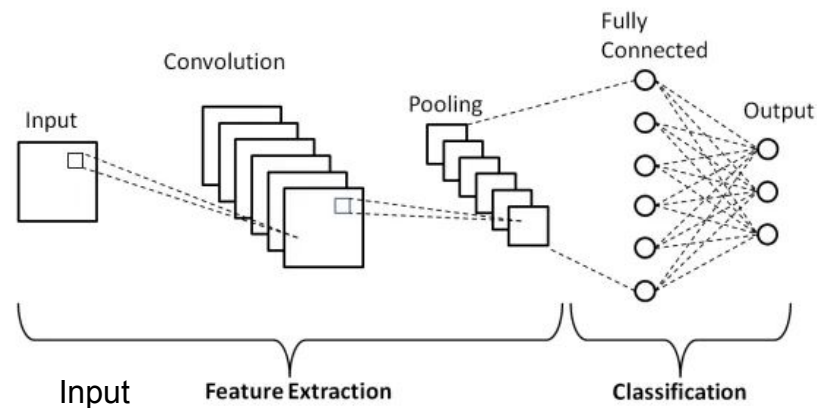
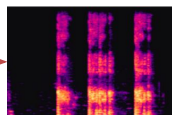
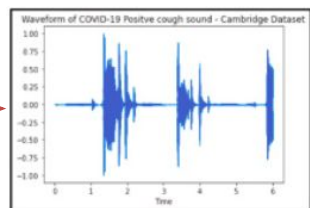
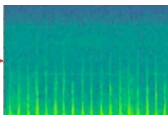
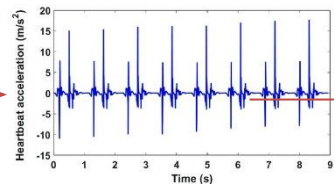
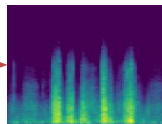
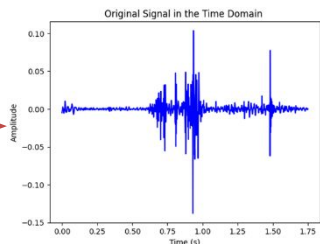
Parte 1: Transformada de Fourier de Tempo Curto (STFT)

Parte 2: Geração de Espectrogramas com a Biblioteca Python Librosa

**Parte 3 - Casos de Uso**



# Análise e Processamento de Sinais de Áudio



Ref: [The Normal Swallow | Otolaryngology — Head & Neck Surgery | Stanford Medicine](#)

Ref: [Treinamento: Aprenda como ouvir \(littmann.com.br\)](#) – Ref: [Cough sound analyzed to identify the severity of COVID-19 patients - Institute for Bioengineering of Catalonia \(ibecbarcelona.eu\)](#)

# Classificação sons da fala (Demência x Saudável)



XX Congresso Brasileiro de Informática em Saúde  
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

## Automatic Classification of Alzheimer's Disease Through Features Extracted from Speech Recordings

### Classificação Automática da Doença de Alzheimer Através de Características Extraídas de Gravações de Fala

### Clasificación Automática de la Enfermedad de Alzheimer Mediante Funciones Extraídas de Grabaciones de Voz

Dieine Estela Bernieri Schiavon<sup>1</sup>, Carla Diniz Lopes Becker<sup>2</sup>

<sup>1</sup> Master's Student, Federal University of Health Sciences of Porto Alegre – UFCSPA, Porto Alegre (RS), Brazil.

<sup>2</sup> Ph.D., Federal University of Health Sciences of Porto Alegre – UFCSPA, DECESA, Porto Alegre (RS), Brazil.

Corresponding author: Dieine Estela Bernieri Schiavon  
E-mail: dieineb@ufcspa.edu.br

#### Abstract

Alzheimer's disease is a progressive neurodegenerative pathology and is among the most common forms of dementia in older people. Changes in memory are common symptoms, and changes in speech and language can be signs of cognitive decline. Intelligent systems have the potential for use as diagnostic support tools. **Objective:** To propose a

### A Proposal for a Convolutional Neural Network for Classification of Alzheimer's Disease Through Features Extracted from Speech Recordings

Dieine Estela Bernieri Schiavon<sup>1</sup>, Carla Diniz Lopes Becker:  
<sup>1</sup> Federal University of Health Sciences of Porto Alegre, Brazil

ALZHEIMER'S ASSOCIATION  
AAC 23

#### INTRODUCTION

**Background:** Alzheimer's disease, a progressive neurodegenerative pathology, is among the most common forms of dementia in the elderly. Memory alterations are a frequent symptom, and speech and language alterations are also frequently mentioned, as can already be noticed in the initial phase of the disease. Deep machine learning is usual in extracting speech features to detect early cognitive decline. This study aims to propose a Convolutional Neural Network model for classifying Alzheimer's disease through features extracted from speech recordings.

#### RESULTS

□ **Proposed CNN:** model inspired by EfficientNet-B5 architecture. The details of the layers and their order in the proposed model are in Figure 1.

□ After defining the architecture, cross-validation was performed in several experiments until the model reached a satisfactory result. The best folder learning curves for this approach are in Figure 2.

#### CONCLUSIONS

□ This study proposes a Convolutional Neural Network to support diagnosing patients with Alzheimer's disease.

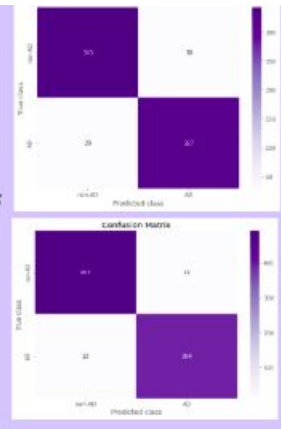
□ The authors point out important limitations regarding the number of data available for analysis. The experience and knowledge acquired with Convolutional Neural Networks state that performance will increase, avoiding overfitting, with a considerable increase in the number of samples available for training. Alternatively, pre-trained networks on large datasets helps to improve classification results when the

#### □ Metrics to the Proposed Model

Metrics	With Pauses	Without Pauses
Accuracy	91.91%	90.26%
Sensitivity	98.03%	85.71%
ROC Curve	0.90	0.91

#### □ Results of the pre-trained EfficientNet-B5 network

Metrics	With Pauses	Without Pauses
Accuracy	97.37%	95.39%
Sensitivity	97.04%	98.03%
Precision	97.04%	92.13%
Specificity	97.62%	93.27%
F-score	97.04%	94.09%
ROC Curve	0.97	0.96



# Classificação sons deglutição (Normal x Alterado)



São Paulo, 04 de Julho de 2024

Prezado(a) Colega **DIEINE ESTELA BERNIERI SCHIAVON DE ANDRADE**,

A Comissão Organizadora do VI Congresso Brasileiro de Disfagia, tem a grande satisfação de cumprimentá-lo(a) pela seleção do seu trabalho intitulado **EFICÁCIA DE ESPECTROGRAMAS E ESCALOGRAMAS NA IDENTIFICAÇÃO DE PADRÕES DE DEGLUTIÇÃO USANDO APRENDIZADO DE MÁQUINA** para apresentação no formato **ORAL TEMAS LIVRES** no evento. Estendemos nossos cumprimentos aos demais co-autores.

## EFICÁCIA DE ESPECTROGRAMAS E ESCALOGRAMAS NA IDENTIFICAÇÃO DE PADRÕES DE DEGLUTIÇÃO USANDO APRENDIZADO DE MÁQUINA

Dieine Estela Bernieri Schiavon, Danielli Pires Vieira, Carla Diniz Lopes Becker, Fernando Neves Hugo, Rafaela Soares Rech

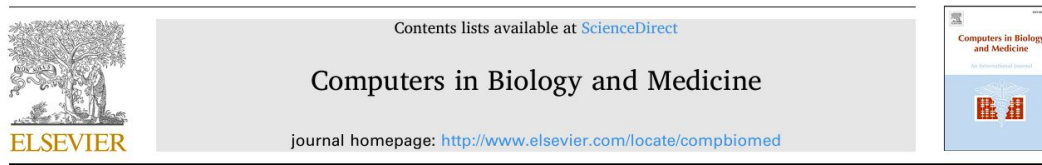
**INTRODUÇÃO:** A disfagia orofaríngea (DO) pode resultar em graves complicações de saúde, assim diagnosticar e reabilitar adequadamente é de suma importância. Recentes avanços em aprendizado de máquina oferecem novas oportunidades para desenvolver técnicas não invasivas para a identificação de DO. **OBJETIVO:** Verificar a acurácia de espectrogramas e escalogramas como ferramentas para identificar padrões de deglutição alterados utilizando técnicas de aprendizado de máquina. **MÉTODO:** Estudo transversal, que foi aprovado no Comitê de Ética em Pesquisa sob número 50257821.9.0000.5345. Sons de deglutição foram coletados de um grupo de 78 participantes, incluindo 50 indivíduos com diagnóstico de DO, comprovado por videofluoroscopia e 28 indivíduos saudáveis. No pré-processamento dos dados os espectrogramas e escalogramas foram gerados a partir dos sons de deglutição, captados por estetoscópio digital, utilizando linguagem de programação *Python*. Os dados pré-processados foram utilizados como entradas para algoritmos de aprendizado de máquina, mais especificamente, redes neurais convolucionais. O desempenho dos modelos foi avaliado utilizando métricas de acurácia, sensibilidade e especificidade. **RESULTADOS:** O modelo baseado em redes neurais convolucionais com utilização de **espectrogramas** obteve **acurácia de 90,54%, sensibilidade de 91,89% e especificidade de 89,19%**. O modelo com **escalogramas** obteve **acurácia de 87,84%, sensibilidade de 91,89% e especificidade de 83,78%**. A análise dos resultados mostrou que esses métodos capturam características relevantes dos sons de deglutição que são eficazmente utilizadas pelos algoritmos de aprendizado de máquina para a detecção de DO. **CONCLUSÕES:** A combinação de espectrogramas ou escalogramas com algoritmos de aprendizado de máquina mostrou-se promissora na identificação de padrões de deglutição normais e alterados.

**PALAVRAS-CHAVE:** Alterações de Deglutição; Apoio ao Diagnóstico; Aprendizado de Máquina; Escalograma; Espectrograma

# Classificação sons mosquitos



Computers in Biology and Medicine 129 (2021) 104152



## Detecting *Aedes aegypti* mosquitoes through audio classification with convolutional neural networks

Marcelo Schreiber Fernandes<sup>a,\*</sup>, Weverton Cordeiro<sup>a</sup>, Mariana Recamonde-Mendoza<sup>a,b,\*\*</sup>

<sup>a</sup> Institute of Informatics (INF), Universidade Federal Do Rio Grande Do Sul (UFRGS), Porto Alegre, Brazil

<sup>b</sup> Bioinformatics Core, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil

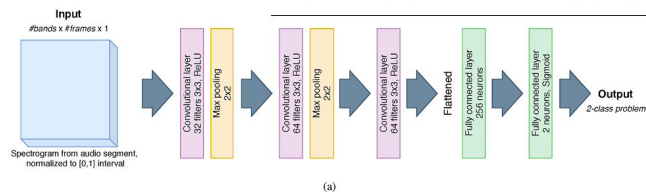


Table 5

Summary of performances obtained by the proposed strategies in the classification of *Aedes aegypti* based on its wingbeat sounds. For the the results for the general model (i.e., all mosquito species) and for *Aedes aegypti* class.

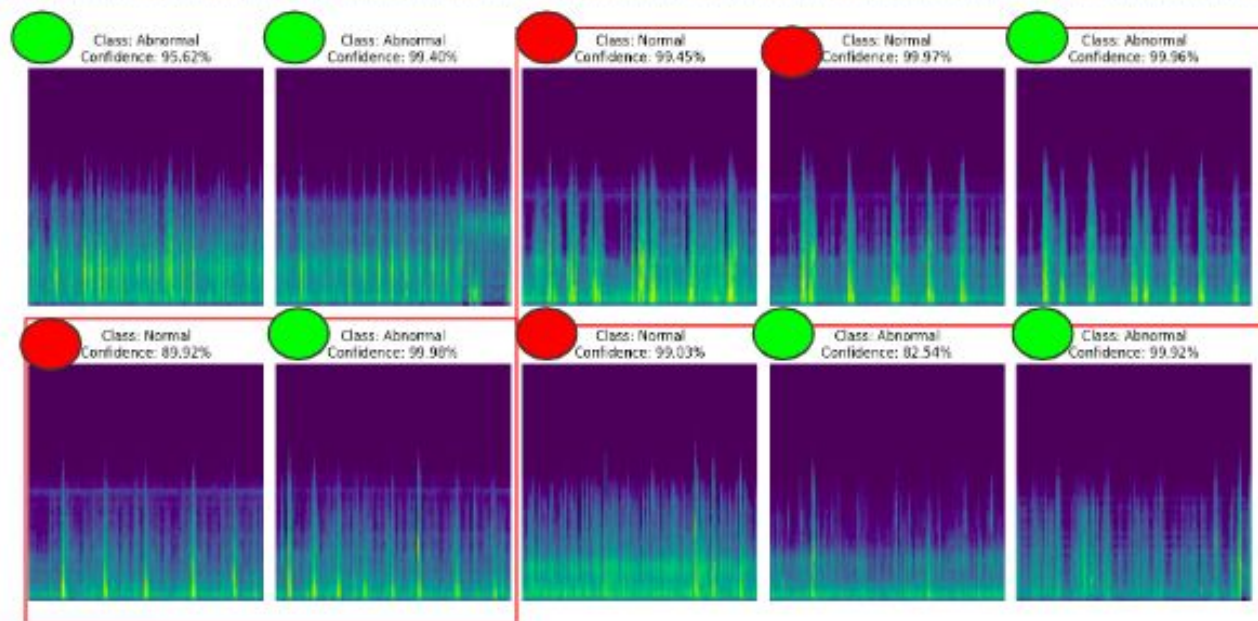
Model	Accuracy (%)		Precision (%)		Recall (%)	
	Avg.	Std. Dev.	Avg.	Std. Dev.	Avg.	Std. Dev.
Binary	97.65	0.55	90.63	4.33	88.49	6.68
Multiclass - <i>Aedes aegypti</i> only	78.12	2.09	89.52	3.69	90.23	3.83
Multiclass - all mosquitoes			79.41	13.00	76.22	16.49
Ensemble (90% voting threshold)	94.56	0.77	68.19	3.67	96.82	1.62



# Classificação Sinais de Deglutição



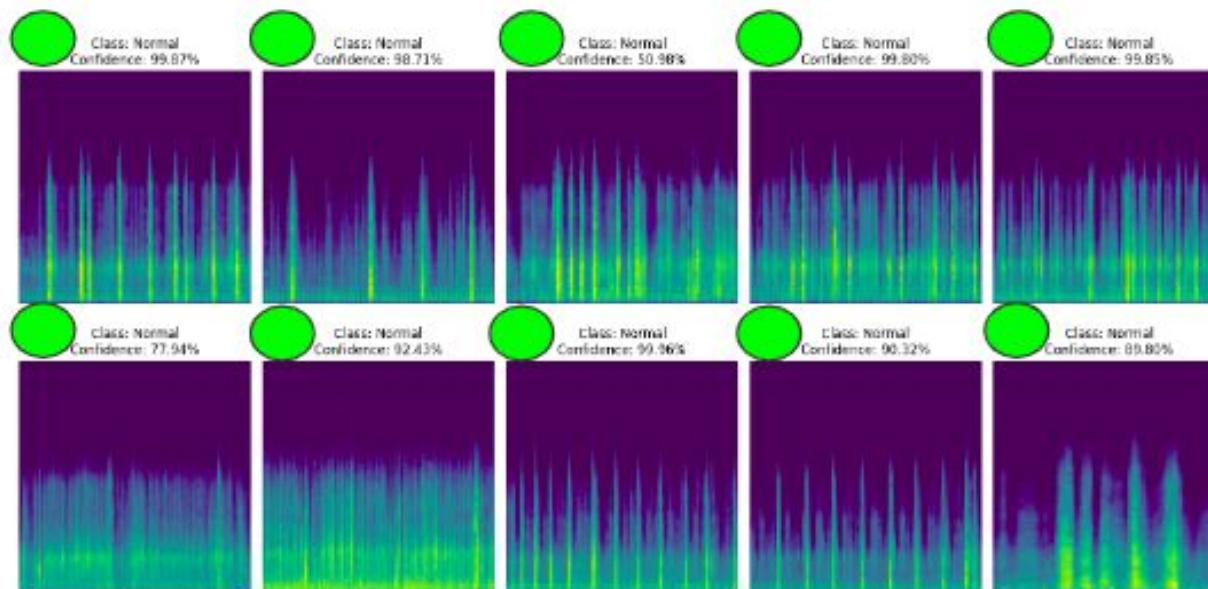
## Class: 1 (Abnormal)



# Classificação Sinais de Deglutição



## Class: 0 (Normal)



# Resumo da Aula

---



## Parte 1:

- **Introdução à STFT:** Decomposição de sinais no domínio do tempo e frequência.
- Aplicações: Análise de áudio, detecção de padrões para tarefas de classificação.

## Parte 2:

- **Introdução à Librosa:** Ferramenta para análise de áudio em Python.
- Passo a Passo: Demonstração prática no Colab sobre a geração de espectrogramas.

## Parte 3:

- **Deep Learning:** Espectrogramas como entrada para modelos de classificação de sinais biológicos, melhorando a precisão e a generalização

# OBRIGADA!

