1 2 9 0 FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE D COIMBRA

Identificação de dígitos através de características extraídas de sinais de áudio

Análise e Transformação de Dados

Relatório

PL1

Daniel Coelho Pereira 2021237092

- ## Estruturas de Dados

Para a implementação deste projeto, foi desde logo definida uma estrutura de dados chamada *arrayAudios*. Cada linha da estrutura representa um arquivo de áudio. As colunas armazenam as informações seguintes:

Coluna 1: Diretório do arquivo
Coluna 2: Nome do arquivo
Coluna 3: Participante
Coluna 4: Dígito
Coluna 5: Número da repetição
Coluna 6: Taxa de amostragem
Coluna 7: Sinal de áudio
Coluna 8: Energia total
Coluna 9: Amplitude Máxima
Coluna 10: Amplitude Mínima
Coluna 11: Razão de Amplitudes
Coluna 12: Desvio Padrão
Coluna 13: Coeficientes de Fourier
Coluna 14: Máxima Amplitude
Coluna 15: Spectral Edge Frequency
Coluna 16: Entropia Espectral
Coluna 17: Razão Média-Baixa
Coluna 18: Energia na Faixa 600-900 Hz
Coluna 19: Assimetria Espectral
Coluna 20: Dígito Classificado
Coluna 21: Taxa de Amostragem Original
Coluna 22: Sinal de áudio Original
Coluna 23: Centroide
Coluna 24: Largura de banda espectral
Coluna 25: Rolloff
Coluna 26: Fluxo espectral médio
Coluna 27: Planicidade Espectral
Coluna 28(...): DWT

Meta 1

- ## Primeira representação gráfica dos sinais

No ponto 3, foi realizada a primeira representação gráfica dos sinais importados. Para isso foi desenvolvida a função *plotSinais* que permite visualizar um exemplo de cada um dos 10 algarismos. Nesta visualização inicial, verifica-se que os sinais ainda não passaram por qualquer tipo de processamento. O silêncio inicial ainda é visível, e há variações nas amplitudes e nas durações dos sinais, dificultando a visualização das características reais dos sinais.

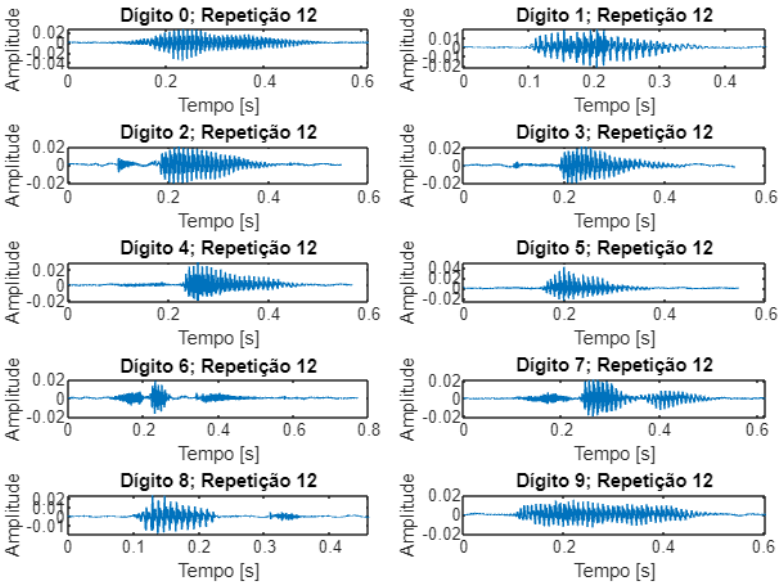


Figura 1 - Gráficos dos Áudios Originais

- ## Pré-processamento

Para garantir uma melhor análise e diferenciação entre os dígitos, foi aplicada uma etapa de pré-processamento. Esta etapa envolveu:

- Remoção do silêncio inicial
- Normalização da amplitude
- Ajuste da duração dos sinais

- ## Segunda representação gráfica dos sinais

Após esse pré-processamento, os sinais foram novamente representados graficamente através da função *plotSinais*. Nesta nova visualização, nota-se que os sinais possuem inícios alinhados, amplitudes uniformizadas e

comprimentos ajustados, facilitando a extração de características e a comparação entre os diferentes dígitos.

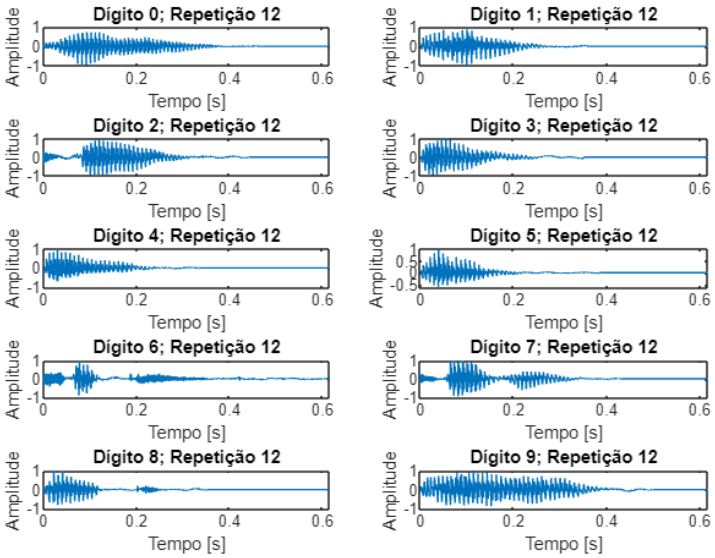


Figura 2 - Gráficos dos Áudios com Pré-processamento

- ## Extração e análise das características temporais

No ponto 7 foram calculadas, com ajuda da função *calcularFeatures*, as seguintes características temporais:

- Energia total
- Amplitude máxima
- Amplitude mínima
- Razão de amplitudes
- Desvio Padrão

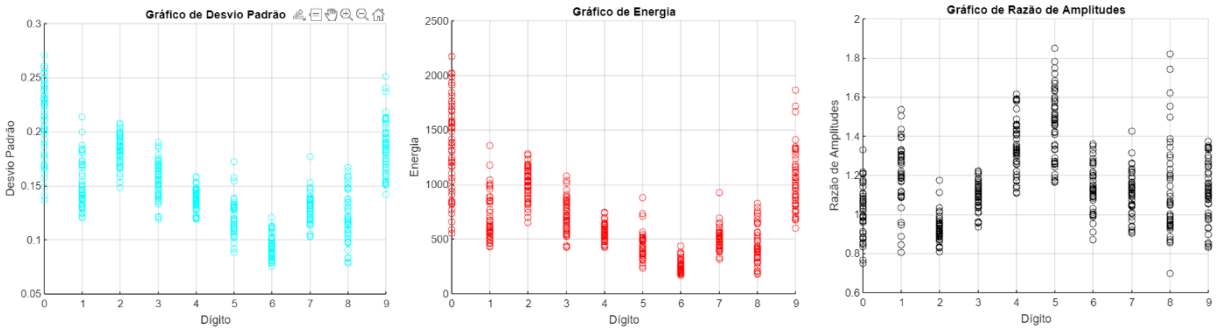


Figura 3 – Representação gráfica das características extraídas

Após a extração destas características, foram gerados gráficos para visualizar e comparar os valores obtidos para cada dígito. Essas representações permitem concluir que a **energia total**, a **razão de**

amplitudes e o **desvio padrão** são as características que apresentam maior discriminação entre os dígitos.

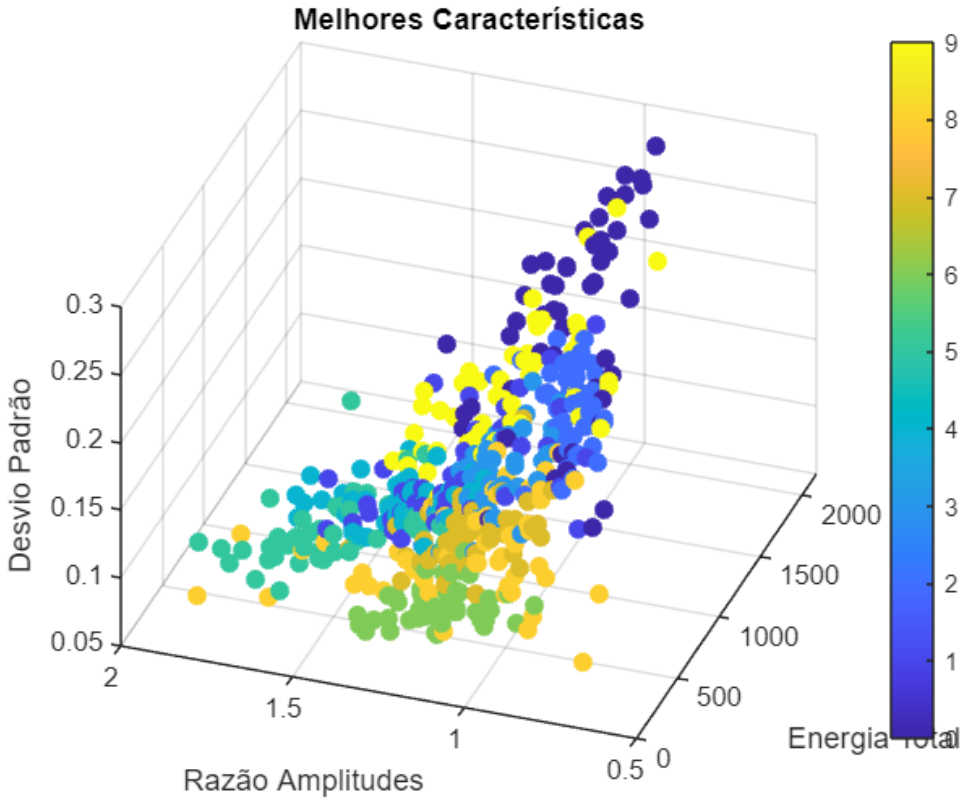
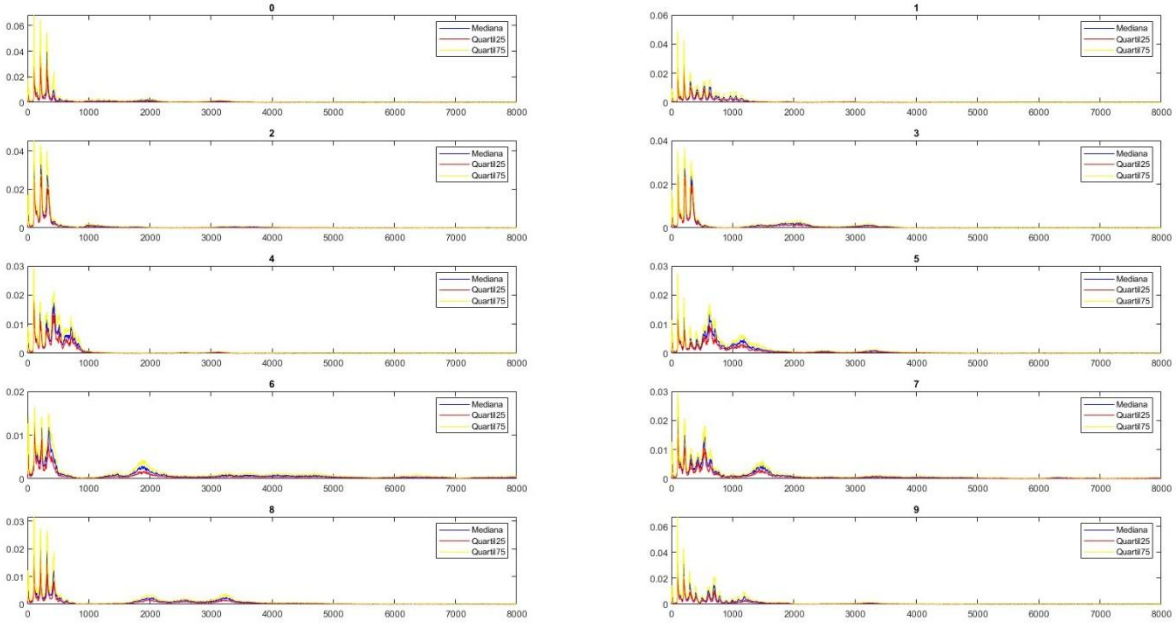


Figura 4 - Gráfico 3D das melhores features

Meta 2

- ## Espectro de amplitude mediano



De modo geral, os espectros de amplitude mediana dos dígitos tendem a concentrar mais energia nas faixas de baixa frequência (0-1000 Hz), o que é característico da fala humana.

Dígitos como 0, 1, 2 e 5 apresentam espectros mais estáveis e simples, com pouca variação entre as amostras, como se observa na proximidade dos quartis.

Por outro lado, os dígitos 3, 6 e 7 mostram maior complexidade espectral, com distribuição de energia em frequências médias e maior dispersão entre o primeiro e o terceiro quartil.

- ## Extração e análise das características espectrais

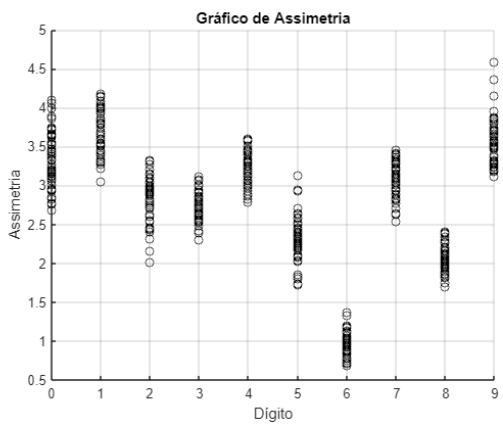
No processo de análise dos sinais de áudio, diversas características espectrais foram calculadas para fornecer informações detalhadas sobre a distribuição da energia nas diferentes frequências.

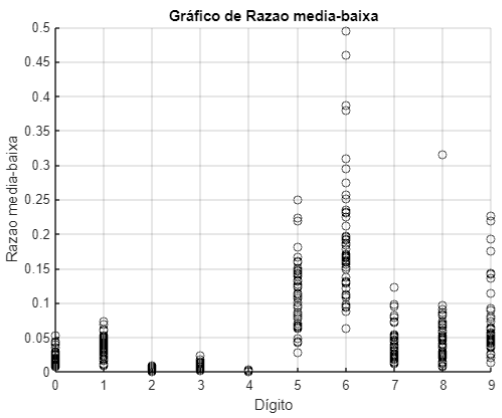
As seguintes características foram extraídas e armazenadas nas respetivas colunas do array arrayAudios:

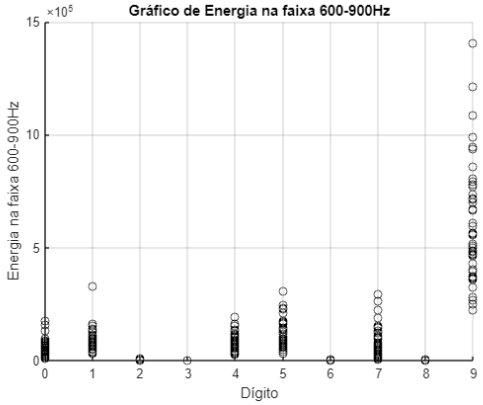
Máxima Amplitude - Coluna 14
Spectral Edge Frequency - Coluna 15
Entropia Espectral - Coluna 16
Razão Média-Baixa - Coluna 17
Energia na Faixa 600-900Hz - Coluna 18
Assimetria Espectral - Coluna 19

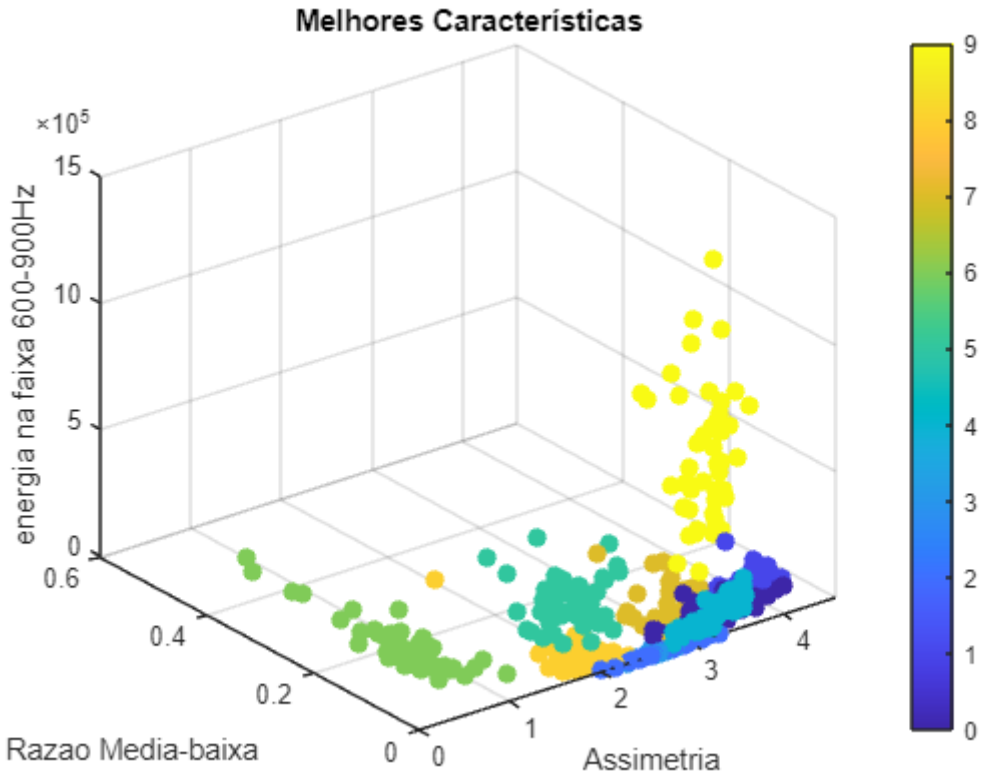
As características espectrais que permitiram uma melhor discriminação dos dígitos foram as seguintes:

Assimetria Espectral
Razão Média-Baixa
Energia na Faixa 600-900Hz









Meta 3

- ## Melhores features espectrais e temporais

Energia Total - Posição 8
Razão Amplitudes - Posição 11
Desvio Padrão - Posição 12
Assimetria Espectral - Posição 19
Razão Média-Baixa - Posição 17
Energia na faixa 600-900Hz - Posição 18

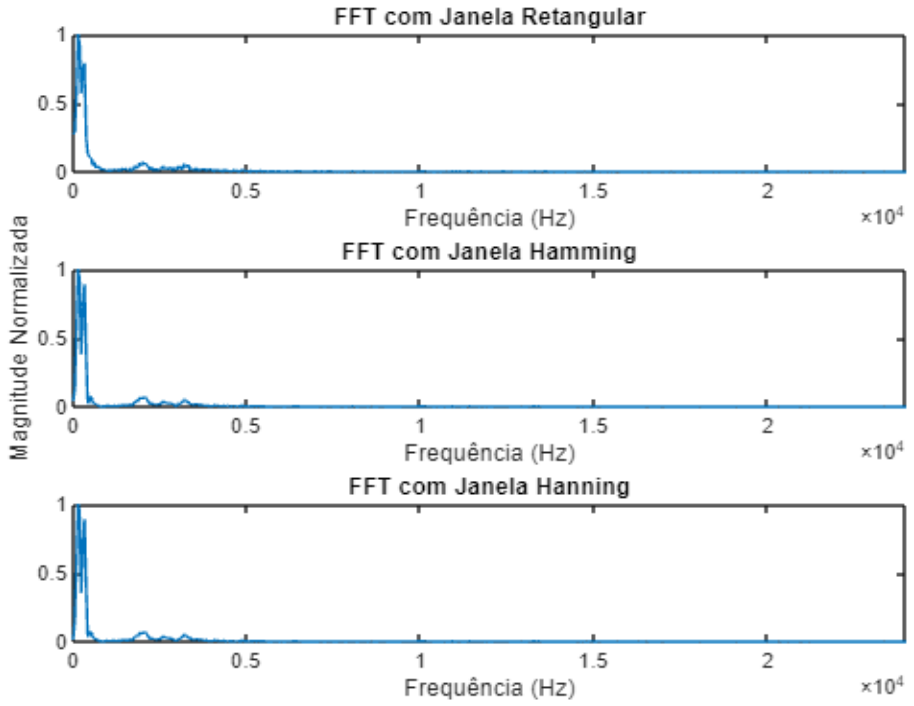
- ## Comparação dos dígitos

Os dígitos com maior taxa de acerto são os dígitos **2**, **6** e **9**, todos com 45 acertos e apenas 5 erros.

Já os dígitos **1, 0** e **3** têm um número significativo de erros, o dígito **1** tem 21 erros e o **0** e **3** têm 15 erros.

Os dígitos **4**, **5**, **7** e **8** têm um número razoável de acertos, com uma média de acertos entre 37 e 46, o que indica um desempenho mais equilibrado, mas com alguns erros.

- **Comparar três tipos de janela diferentes → Retangular, Hamming e Hanning**



Janela Retangular:

- O espectro tem picos muito agudos e existe ***muito leakage*** (vê-se energia espalhada fora dos principais picos). Isso acontece porque a **Janela Retangular** não suaviza o sinal.

Janela Hamming:

- O espectro está mais "limpo", ou seja, o **leakage é menor**. A **Janela Hamming** suaviza melhor o sinal nas extremidades, reduzindo bastante o espalhamento da energia e deixando os picos principais mais destacados.

Janela Hanning:

- Se aproximarmos o gráfico, conseguimos perceber que **existe ainda menos leakage do que na Janela Hamming**. As transições entre picos são ainda mais suaves, mas esta suavização extra **provoca uma ligeira redução da amplitude dos picos principais** em comparação com a **Janela Hamming**.

Assim, a escolha da janela influencia diretamente a qualidade da análise em frequência. Enquanto a **Janela Retangular** oferece maior detalhe mas com muito leakage, as **janelas Hamming e Hanning** conseguem reduzir significativamente o leakage, proporcionando espectros mais limpos

Meta 4

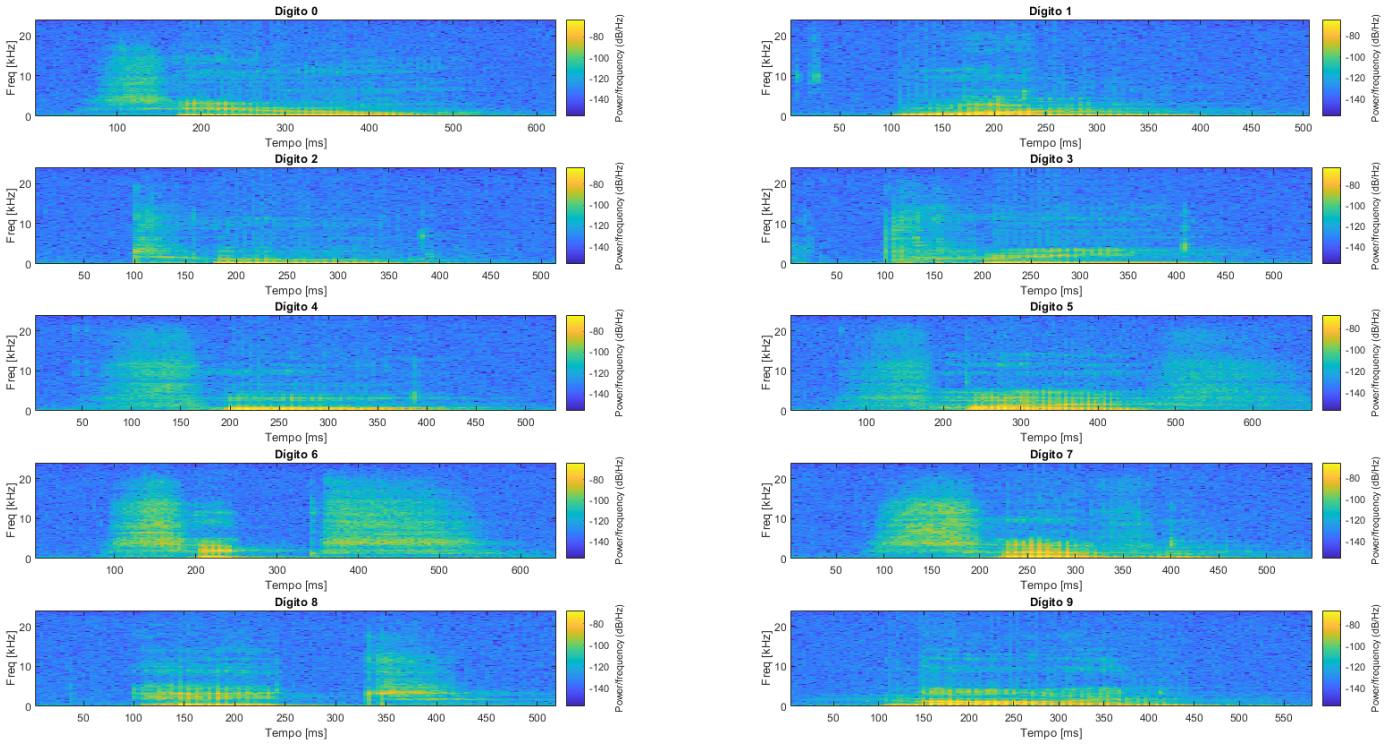
- **STFT**

Para o cálculo da STFT, foram escolhidos os seguintes valores:

Tamanho da janela: 400

Sobreposição: round(0.5 * tamanho da janela)

Número de pontos: 1024



Dígitos como 1 e 2 apresentam **traços curtos**, com energia concentrada em pequenos intervalos de tempo.

Dígitos como 0 e 6 mostram **zonas de energia mais largas e contínuas**, com cores quentes ao longo de mais tempo.

Dígitos como 5, 7 e 8 têm energia distribuída de forma mais variada no tempo e em diferentes faixas de frequência.

- ## Características tempo-frequência

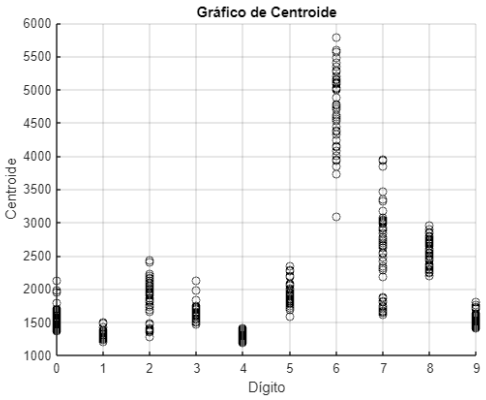
No processo de análise dos sinais de áudio, diversas características tempo-frequência foram calculadas com o intuito de diferenciar os dígitos.

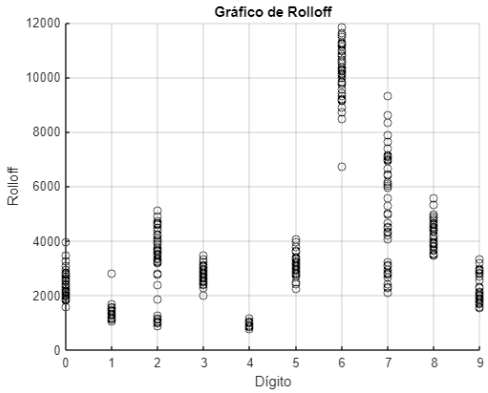
As seguintes características foram extraídas e armazenadas nas respetivas colunas do array arrayAudios:

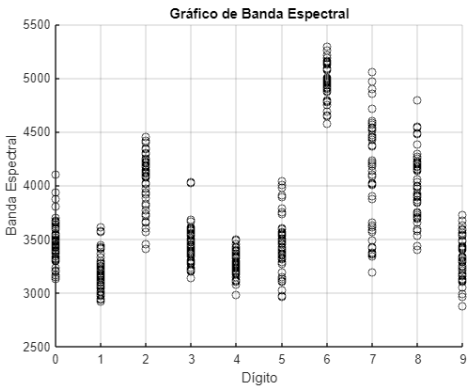
Centroide - Coluna 23
Largura de banda espectral - Coluna 24
Rolloff 85% - Coluna 25
Fluxo espectral médio - Coluna 26
Planicidade espectral - Coluna 27

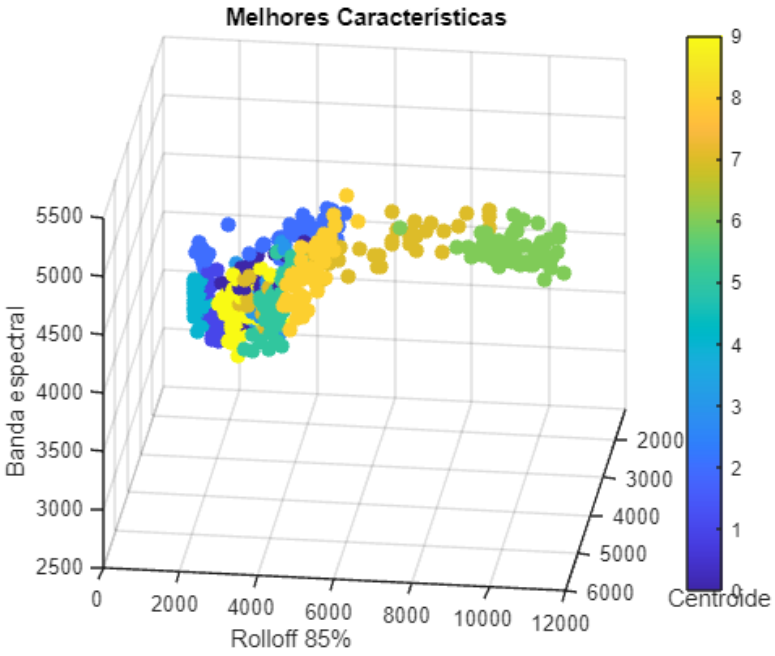
As características que permitiram uma melhor discriminação dos dígitos foram as seguintes:

Centroide - Coluna 23
Rolloff 85% - Coluna 25
Largura de banda espectral - Coluna 24







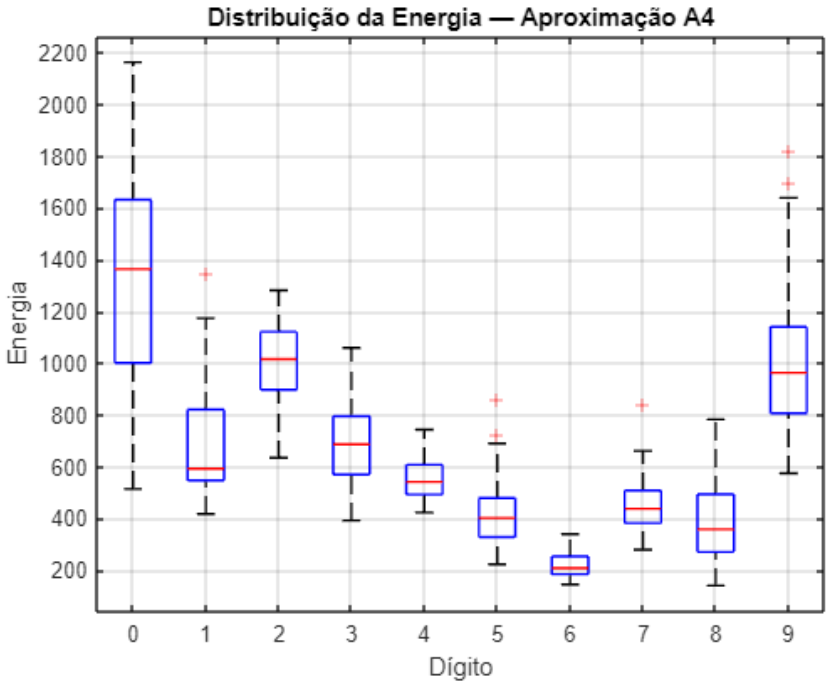


- ## Transformada de Wavelet Discreta

wavelet = 'db4'

nivel_decomposicao = 4

Distribuição da Energia — Aproximação



O boxplot de **aproximação A4** mostra, para cada dígito em inglês, quanta energia o sinal concentra nessa faixa muito grave (baixas frequências):

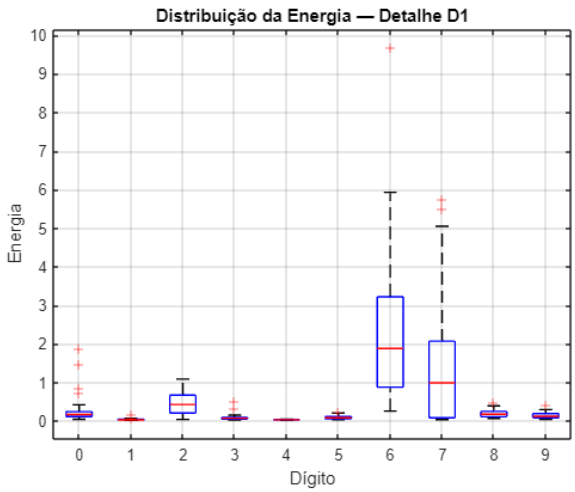
Dígito 0: energia mais alta (mediana ≈ 1400), graças à longa vogal.

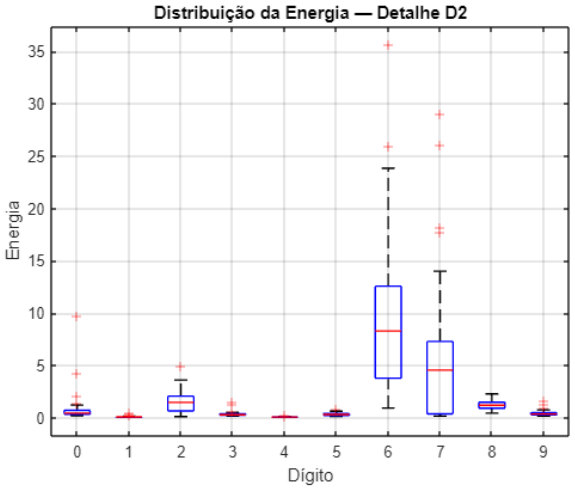
Dígitos 9 e 2: medianas elevadas, pois as suas vogais ainda concentram muita energia grave.

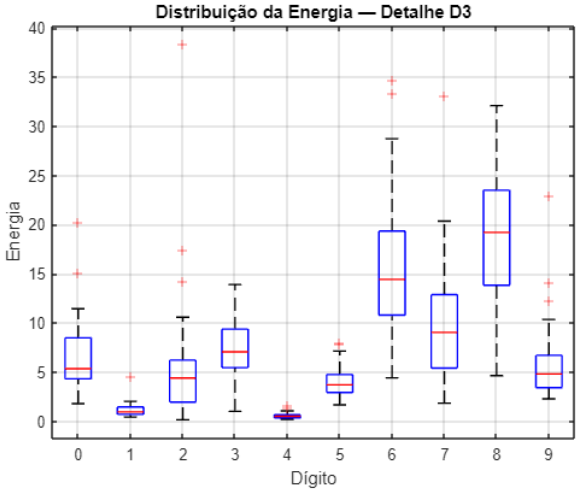
Dígitos 3, 4, 5, 7 e 8: medianas decrescentes de 600 (“three”) a 300 (“eight”), refletindo vogais progressivamente mais fechadas.

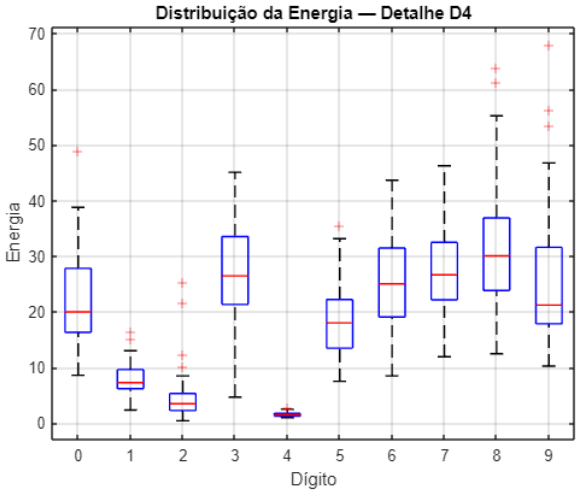
Dígito 6: mediana de energia baixa (200), pois a vogal breve e as consoantes geram pouca energia nas baixas frequências.

Distribuição da Energia — Detalhe









Os boxplots de **detalhe D1 a detalhe D4** mostram, para cada dígito em inglês, quanta energia o sinal concentra nessas faixas agudas (altas frequências).

No **detalhe D1**, sobressaem claramente os **dígitos 6 e 7**, devido ao som agudo acentuado no “s”. **Este nível isola com eficácia os sons de alta frequência.**

O **detalhe D2** segue o mesmo padrão que o **D1,** mas com valores maiores**.**

No **detalhe D3**, o algarismo **8** tem um aumento muito significativo devido ao facto de concentrar a sua energia nesta faixa.

D4 realça as vogais mais marcantes, com “eight”, “seven”, “six”, “nine” e “three” exibindo as maiores energias.

Comparação entre STFT e DWT

A **STFT** gera um espectrograma com resolução fixa tanto no tempo quanto na frequência, o que a torna muito útil para observar como o conteúdo espectral muda ao longo dos instantes e para extrair estatísticas locais como centroide, rolloff ou fluxo espectral.

No entanto, essa resolução fixa impõe um compromisso: janelas mais curtas favorecem a resolução temporal, mas perdem detalhes em frequências graves; janelas mais longas melhoram a resolução em frequência, mas desfavorecem eventos rápidos.

A **DWT**, por sua vez, **extrai características de energia em múltiplas escalas**, facilitando a separação entre componentes de alta e baixa frequência. Isso resulta em um conjunto de características simples, eficientes e complementares às da STFT, contribuindo para uma melhoria no desempenho da classificação dos dígitos.