Relatório

Meta I

Importação dos sinais de áudio

O código em MATLAB inicia-se com a importação de sinais de áudio correspondentes a um participante (pasta 55), que repete cada um dos dígitos de 0 a 9, 50 vezes (500 sinais de áudio totais). Isto é conseguido através do uso da função "audioread", para ler os arquivos de formato ".wav". Todos estes sinais são armazenadas numa célula "signals", previamente pré-alocada para conseguir armazenar a quantidade necessária de sinais. Este armazenamento consegue-se através de um loop iterativo, que percorre todos os dígitos (a começar em 0) e respetivas repetições (a começar em 0) para construir o nome do arquivo e importar o sinal correspondente ao mesmo. O resultado final será então uma célula contendo todos os sinais de áudio em análise.

```
% Usar audioread para ler o arquivo de áudio
[signal, sampling_frequency] = audioread(file_path);
```

Fig. 1 – Função de leitura de áudio

Reprodução e representação gráfica

Após o passo anterior, é escolhido um exemplo de sinal de áudio (dígito 9, repetição 25), que é reproduzido utilizando a função "audioplayer". De seguida, e recorrendo a uma função criada, "plot_graph.m", representam-se os sinais graficamente (para todos os dígitos, repetição 1), em função do tempo (em segundos). Isto proporciona uma visão inicial da forma de onda do sinal e auxilia na compreensão visual do conteúdo do áudio.

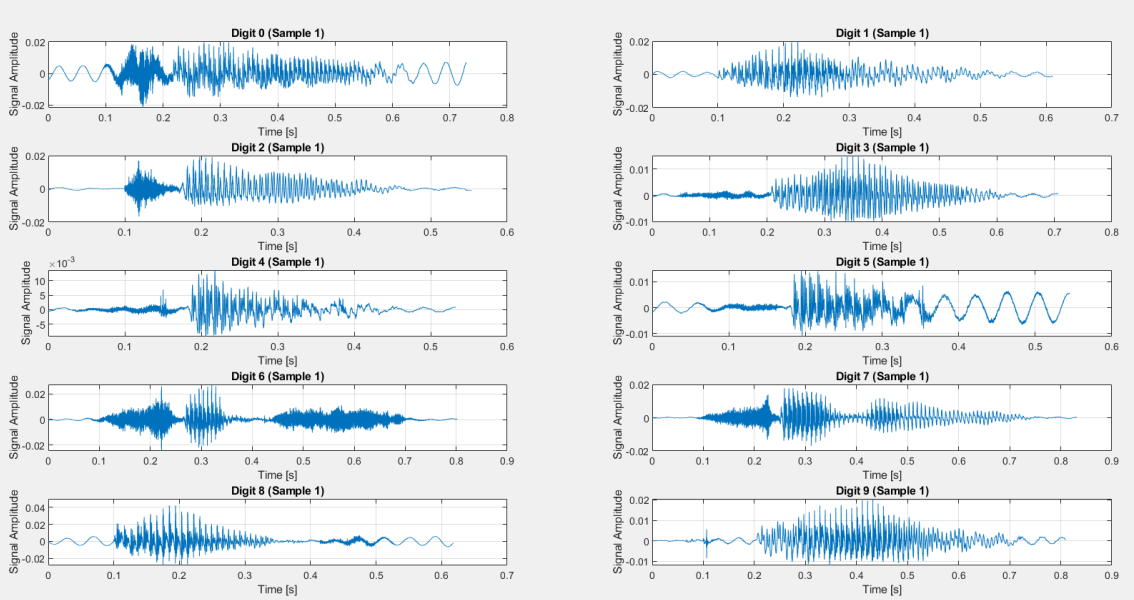

Fig. 2 – Sinais de áudio antes do pré-processamento

Pré-processamento e extração de características temporais

De forma a garantir que todos os sinais iniciem exatamente ao mesmo tempo, tenham a mesma duração e intervalo de amplitude, realiza-se um pré-processamento de sinal que recorre a três estratégias diferentes:

- Remoção do "silêncio" inicial (energia próxima a 0);

- Normalização da amplitude [-1 : 1] com base nas amplitudes máxima e média (correção de variações na gravação do som, como a distância do participante ao microfone);

- Adição de silêncio ao fim de todos os sinais cuja duração seja inferior à duração pretendida (duração máxima de todos os sinais após a remoção do silêncio inicial).

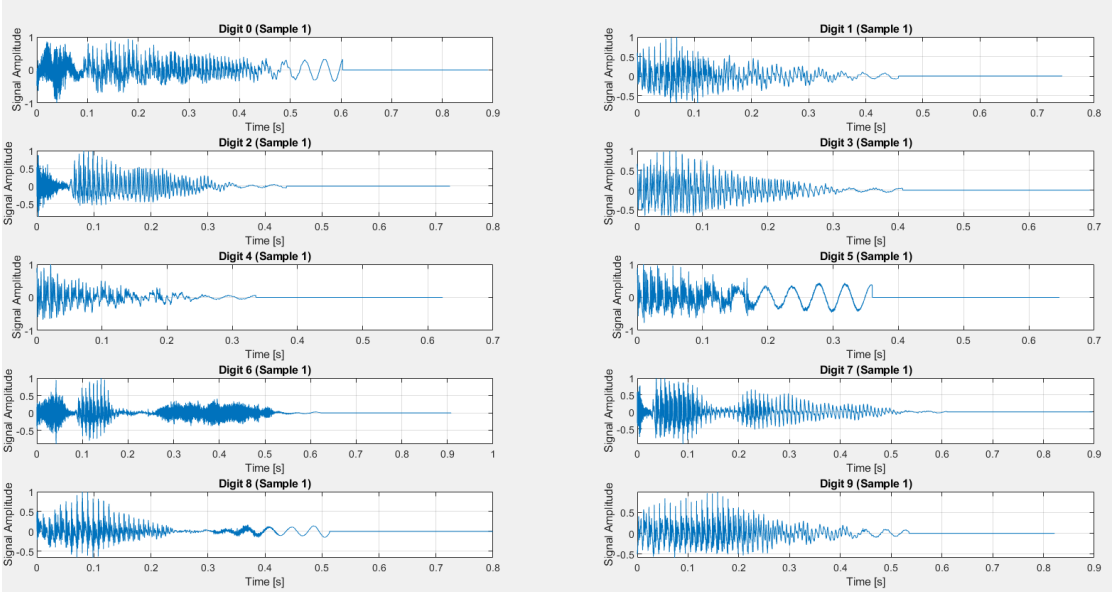

Fig. 3 – Sinais de áudio depois do pré-processamento

(1) Energia total

(2) Amplitude máxima

(3) Amplitude mínima

(4) Desvio padrão de amplitudes

(5) Razão de amplitudes máxima/média

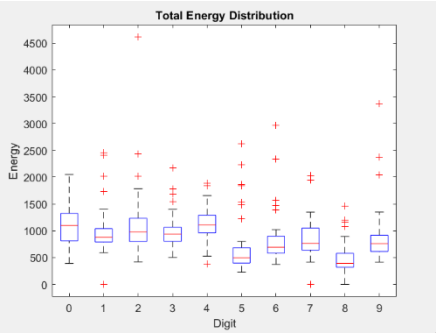

Fig. 4 – Energia total, para os dígitos de 0-9 (1)

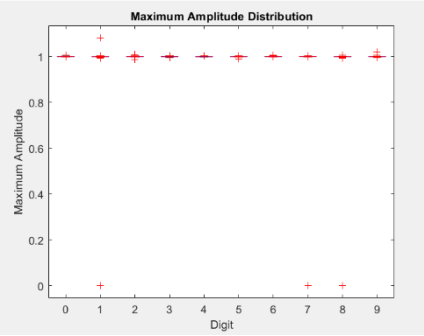

Fig. 5 – Amplitude máxima, para os dígitos de 0-9 (2)

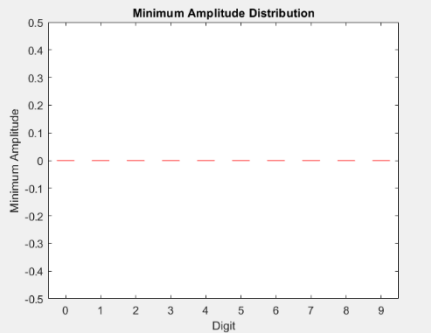

Fig. 6 – Amplitude mínima, para os dígitos de 0-9 (3)

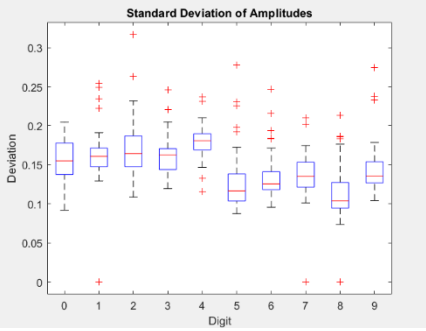

Fig. 7 – Desvio padrão de amplitudes, para os dígitos de 0-9 (4)

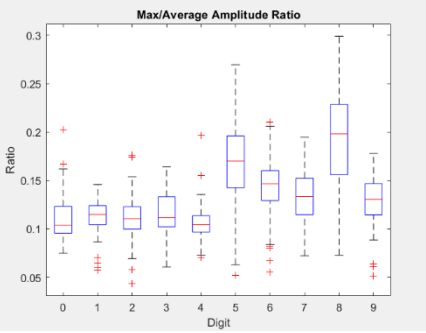

Fig. 8 – Razão de amplitudes máxima/média, para os dígitos de 0-9 (5)

Com base na análise dos gráficos das cinco características temporais calculadas, é possível ver claramente que nem a amplitude máxima nem a amplitude mínima de cada um dos dígitos de 0 a 9 é relevante para a sua distinção, dado que todos têm como amplitude máxima o valor 1 e como amplitude mínima o valor 0, havendo apenas ligeiras discrepâncias nos dígitos 1, 7 e 8, com alguns valores fora do normal. No entanto, estes são desprezáveis.

Já nas restantes três características, energia total, desvio padrão de amplitudes e razão de amplitudes máxima/média, há uma nítida separação de dígitos por dois conjuntos diferentes. O primeiro conjunto seria constituído pelos dígitos 0, 1, 2, 3, 4 e o segundo por 5, 6, 7, 8, 9. Ao analisar os gráficos de energia total e de desvio padrão, observa-se que o primeiro tanto tem valores de energia ligeiramente maiores como um desvio de amplitudes ligeiramente superior.

Isto significa que é característico destes dígitos serem pronunciados de forma mais enfática ou com maior clareza, o que resulta numa maior energia acústica. Além disso, são estes os dígitos que podem ter uma maior variação na forma como são pronunciados, possivelmente devido a características fonéticas próprias, que acabam por afetar os valores de amplitude detetados.

Ao analisar o último gráfico com maior detalhe, pode-se ainda decompor o último conjunto em dois: dígitos 5, 8 e dígitos 6, 7, 9. Os valores da razão entre as amplitudes máxima e média dos dois primeiros dígitos são facilmente diferenciáveis dos outros valores, tendo também máximos e mínimos mais fora da norma. A grande distinção visível na razão amplitude máxima/média pode dever-se ao facto de serem palavras cuja primeira sílaba é mais aberta, resultando numa maior amplitude, mas em que a segunda sílaba é fechada, resultando numa baixa amplitude, nesta parte da pronúncia, o que contribui para diminuir a média geral da amplitude. Assim sendo, a amplitude máxima vinda da primeira parte da palavra é muito superior aos valores médios, o que explica a maior razão.

Principais características temporais discriminatórias

A análise de sinais de áudio, especialmente no contexto do reconhecimento de dígitos falados de 0 a 9, envolve a extração e utilização de características temporais que possam diferenciar eficazmente cada dígito. Como referido anteriormente, as três características temporais escolhidas foram a energia total, razão de amplitudes máxima/média e desvio padrão das amplitudes, devido às suas capacidades distintivas e representativas das variações nos sinais de áudio, ainda que apenas sejam capazes de fazer esta distinção por grupos e não por dígitos, individualmente. Estas características são extraídas e analisadas de forma mais precisa após aplicar técnicas de processamento de sinal, como a normalização e a remoção de silêncios, permitindo que o sistema de reconhecimento seja efetivo. Vamos explorar cada uma delas detalhadamente:

1) Energia: a energia de um sinal de áudio é uma medida que reflete a intensidade total do som ao longo do tempo. É calculada como a soma dos quadrados das amplitudes do sinal em cada ponto de amostra, o que proporciona uma indicação da "força" ou "potência" do sinal. Em contexto de reconhecimento, dígitos com uma pronúncia naturalmente mais forte tendem a apresentar maior energia. Portanto, a energia pode ajudar a diferenciar entre os dígitos falados suavemente e aqueles falados com mais intensidade.

2) Razão de amplitudes máxima/média: a razão entre a amplitude máxima e a média das amplitudes de um sinal de áudio fornece *insights* sobre a distribuição e o pico das intensidades ao longo do tempo. A amplitude máxima é o valor mais alto de amplitude encontrado no sinal, indicando os momentos de maior intensidade sonora. A média das amplitudes representa a intensidade média do sinal ao longo da sua duração. A razão entre estes dois valores pode ser útil para identificar padrões:

   Dígitos com razão alta: podem indicar variações entre momentos de menor e maior intensidades vocais, refletindo uma fala com contrastes de pronúncia.

   Dígitos com razão baixa: podem representar uma fala mais uniforme ou constante em intensidade.

3) Desvio Padrão das Amplitudes: o desvio padrão das amplitudes mede a variação ou dispersão das amplitudes ao redor da média. Um desvio padrão alto indica que as amplitudes do sinal variam amplamente, sugerindo uma pronúncia com variações significativas na intensidade, o que pode ser característico de certos estilos de fala ou de certos dígitos. Por outro lado, um desvio padrão baixo indica que o sinal é mais uniforme, com menos variação na intensidade ao longo do tempo.

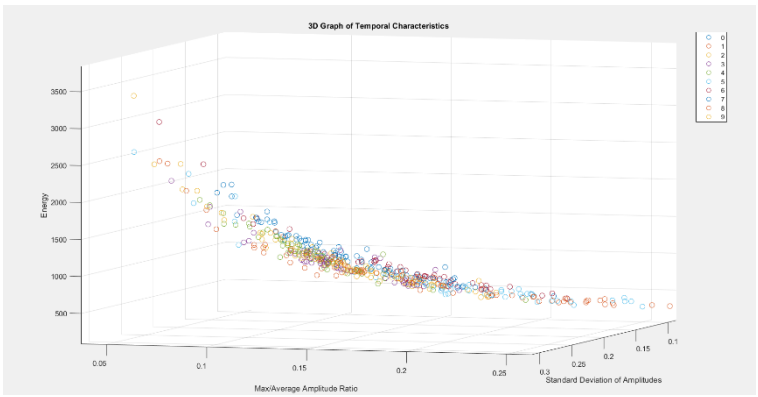

Fig. 9 - *3D Scatter plot* para "Energy vs. Max/Average Amplitude Ratio vs. Standard Deviation of Amplitudes"

META II

Cálculo do espetro de amplitude mediano, primeiro e terceiro quartis

Para calcular o espectro de amplitude mediano normalizado, o primeiro quartil (Q25) e o terceiro quartil (Q75), para os dígitos de 0 a 9, utilizando diferentes tipos de janela (retangular, *Hamming*, *Hann* e *Blackman*), são necessárias várias etapas. Primeiramente, é necessário aplicar a Transformada Rápida de Fourier (FFT) aos sinais normalizados de cada dígito, modificando-os conforme o tipo de janela escolhido. Como é explicado no tópico seguinte, cada janela tem um efeito específico na forma como o sinal é tratado antes da aplicação da FFT, influenciando diretamente o espectro de frequências resultante.

Processo de cálculo:

1) Aplicação da FFT: para cada repetição de cada dígito, a FFT é aplicada ao sinal, que pode ser modificado previamente pela aplicação de uma das janelas mencionadas. Isto resulta num conjunto de amplitudes para as frequências positivas do espetro.

2) Normalização: as amplitudes obtidas pela FFT são normalizadas pelo número de amostras, convertendo-as no espetro de amplitude normalizado.

3) Cálculo dos quartis: para cada conjunto de amplitudes (correspondente a cada dígito e janela), calculam-se os valores medianos (função "median"), o primeiro quartil (25%) e o terceiro quartil (75%) (função "quantile"). Esses valores representam, respetivamente, a mediana do espetro de amplitude, o valor abaixo do qual estão 25% das amplitudes e o valor abaixo do qual estão 75% das amplitudes.

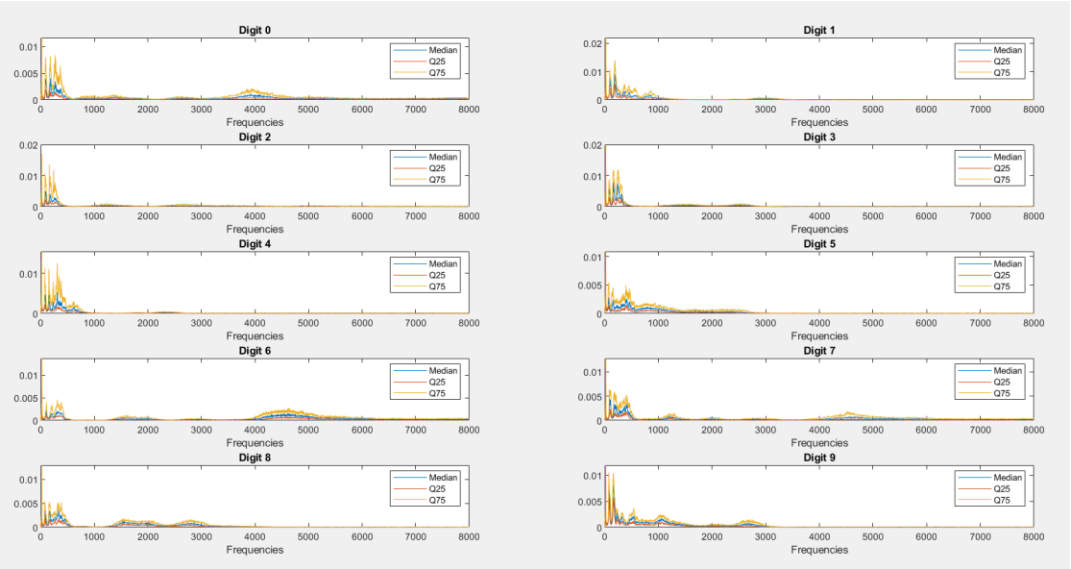

Fig. 10 - Espectro de amplitude mediano normalizado, primeiro quartil e terceiro quartil, para os dígitos de 0-9, em janela retangular. Nota: para efeitos de uma melhor visualização não é apresentado o eixo de frequências completo.

Comparação de três tipos de janelas diferentes

A janela retangular não aplica nenhuma ponderação ao sinal, mantendo-o inalterado. É o método mais simples e direto, mas pode introduzir descontinuidades nas bordas do sinal, tendo picos espectrais mais amplos, o que resulta num espetro com maior vazamento espectral.

Por outro lado, as janelas de *Hamming*, *Hann* e de *Blackman* são projetadas para minimizar esse vazamento espectral. Possuem transições mais suaves para zero nas extremidades, o que ajuda a concentrar a energia do sinal mais precisamente em torno das frequências de interesse. Como resultado, os picos espectrais tendem a ser mais estreitos e bem definidos, o que melhora a capacidade de distinguir entre frequências próximas no espectro. Além disso, estas janelas apresentam lóbulos laterais de menor amplitude em comparação com a janela retangular. Isso significa que as frequências fora do pico principal são mais atenuadas, o que é benéfico para a análise de sinais onde as componentes de frequência principais são de maior interesse e as componentes secundárias devem ser suprimidas.

Em resumo, ao escolher entre as janelas retangular, de *Hamming*, de *Hann* e de *Blackman* para análise espectral, deve-se considerar o equilíbrio entre a resolução espectral e a atenuação do vazamento espectral. Estes efeitos podem ser visualizados nos espetros apresentados, pela atenuação dos valores das amplitudes (para cada frequência) e pelo estreitamento das regiões de pico. Além disso, verifica-se que esta atenuação vai sendo cada vez maior em cada uma das janelas, pela ordem em que foram mencionadas anteriormente.

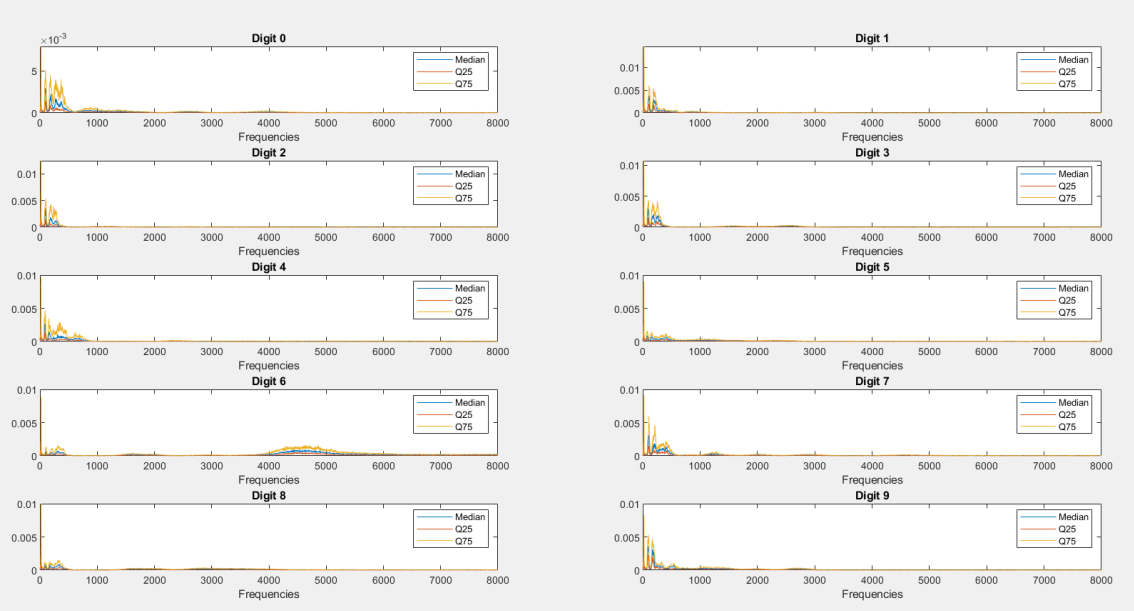

Fig. 11 - Espectro de amplitude mediano normalizado, primeiro quartil e terceiro quartil, para os dígitos de 0-9, em janela de *Hamming*. Nota: para efeitos de uma melhor visualização não é apresentado o eixo de frequências completo.

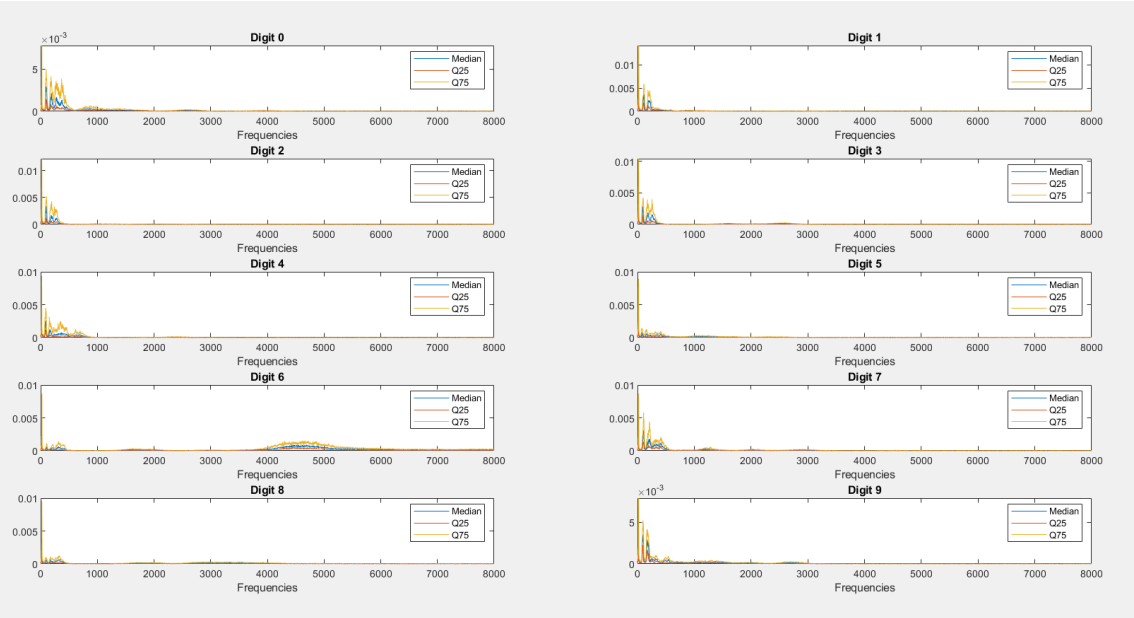

Fig. 12 - Espectro de amplitude mediano normalizado, primeiro quartil e terceiro quartil, para os dígitos de 0-9, em janela de *Hann*. Nota: para efeitos de uma melhor visualização não é apresentado o eixo de frequências completo.

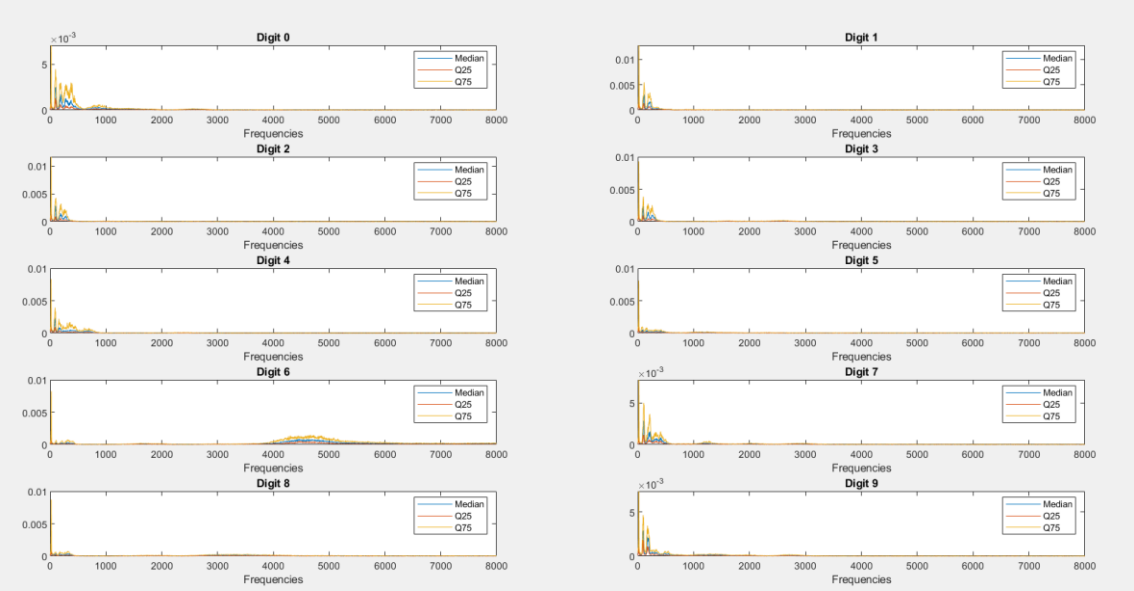

Fig. 13 - Espectro de amplitude mediano normalizado, primeiro quartil e terceiro quartil, para os dígitos de 0-9, em janela de *Blackman*. Nota: para efeitos de uma melhor visualização não é apresentado o eixo de frequências completo.

Extração de características espectrais

(1) Máximos espectrais (amplitude)

(2) Máximos espectrais (posição)

(3) Médias espectrais

(4) Spectral Edge Frequency

(5) Coeficientes de MFCC

(6) Variância espectral

Para encontrar os máximos espectrais (de amplitude e posição), calculamos primeiro a FFT do sinal de voz normalizado. Após isto, obtém-se o espectro de amplitudes normalizadas de cada componente da frequência. O máximo espectral é então identificado como a maior amplitude no espectro e a sua posição (índice) indica a frequência predominante no sinal.

A média espectral é calculada tomando a média das amplitudes normalizadas do espectro de amplitudes. Isto dá uma ideia geral do nível de energia distribuído ao longo do espectro de frequência do sinal.

A *Spectral Edge Frequency* (SEF) é uma medida que indica a frequência abaixo da qual uma certa percentagem da energia total do sinal está concentrada. Inicialmente, calcula-se o espectro de potência do sinal e normaliza-se para que a soma de todas as potências seja igual a 1. De seguida, calcula-se a energia cumulativa ao longo do espectro e identificamos o índice onde essa energia cumulativa atinge 95% do total. A frequência correspondente a esse índice é a SEF. Uma interpolação pode ser usada para obter uma estimativa mais precisa da SEF, especialmente se o índice identificado não corresponder exatamente a 95% da energia total.

Os MFCCs são calculados aplicando uma série de passos que incluem a aplicação de um banco de filtros Mel ao espectro de potência do sinal, seguido pela tomada do logaritmo das energias dos filtros e, finalmente, aplicando a Transformada Discreta do Cosseno (DCT) aos logaritmos das energias.

A variância espectral é calculada a partir do espectro de amplitudes do sinal e fornece uma medida da dispersão das amplitudes espectrais em torno da média espectral. Isto pode ajudar a identificar sinais com ampla distribuição de energia, em diferentes frequências em comparação com sinais mais concentrados.

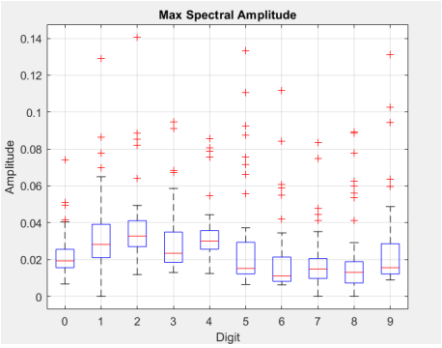

Fig. 14 – Máximos espectrais (amplitude), para os dígitos de 0-9 (1)

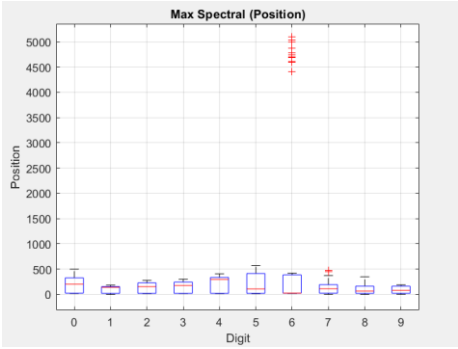

Fig. 15 – Máximos espectrais (posição), para os dígitos de 0-9 (2)

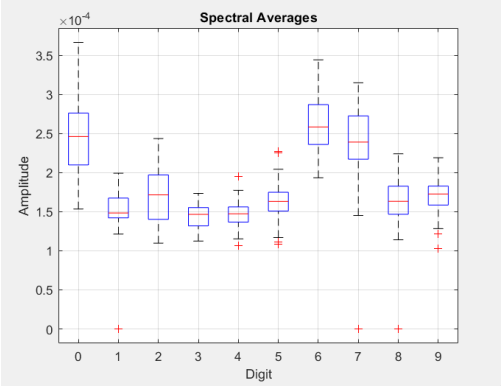

Fig. 16 –Médias espectrais, para os dígitos de 0-9 (3)

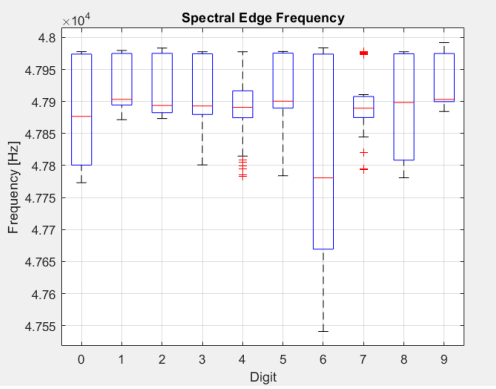

Fig. 17 – *Spectral Edge Frequency*, para os dígitos de 0-9 (4)

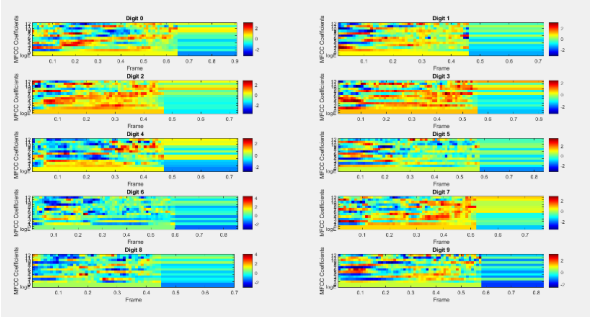

Fig. 18 – Coeficientes de MFCC (*Mel-Frequency Cepstral Coefficients*) , para os dígitos de 0-9 (5)

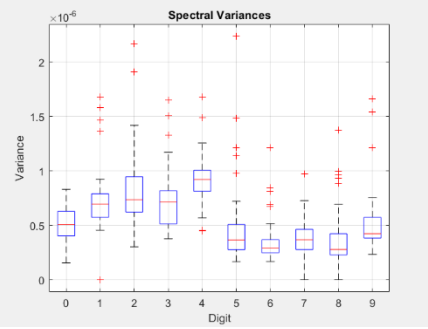

Fig. 19 – Variância espectral, para os dígitos de 0-9 (6)

De acordo com a análise dos diferentes gráficos, e tal como na meta anterior, é possível ver que muitos dividem grupos de dígitos nitidamente.

As características que melhor os distinguem, neste caso, são as médias espectrais, a *spectral edge frequency* e a variância espectral.

No caso das médias espectrais podemos definir dois conjuntos diferentes: o primeiro composto por 0, 6, 7 e o segundo por 1, 2, 3, 4, 5, 8, 9.

Já na *spectral edge frequency*, é possível delimitar quatro diferentes: o conjunto 0, 8, o conjunto 6, o conjunto 4, 7 e, finalmente, o conjunto 1, 2, 3, 5, 9.

Por fim, na variância espectral, apesar desta divisão não ser tão visível, existem os conjuntos 0, 1, 2, 3, 4 e 5, 6, 7, 8, 9.

Nos restantes gráficos, principalmente nos máximos espectrais de amplitude e posição, as diferenças entre os vários dígitos são pouco visíveis e não seriam boas escolhas (nos máximos espectrais de posição obtêm-se ainda piores resultados), dado que o objetivo é criar critérios para uma distinção clara.

Os coeficientes de MFCC poderiam também ser uma boa escolha, pois são visíveis as diferenças de coloração. Por exemplo, os dígitos 5, 6 e 8 poderiam ser agrupados, devido ao facto de apresentarem na sua maioria coeficientes com valor mais reduzido. Para além destes, os dígitos 2, 3 e 7 constituem também um grupo, em que os coeficientes são maioritariamente elevados. Finalmente, os dígitos 0, 1, 4 e 9 apresentam coeficientes mais dispersos, com uma mistura de valores.

Principais características espectrais discriminatórias

As três características espectrais mais eficazes para a diferenciação entre dígitos falados de 0 a 9, considerando a sua popularidade e eficácia em análises de áudio, são:

1) Médias espectrais: as médias espectrais são uma medida importante de distribuição de energia ao longo das frequências num sinal de áudio. Estas representam o centro de gravidade do espectro de frequência, indicando onde a maior parte da energia está concentrada. No contexto de reconhecimento de fala, as médias espectrais são úteis para distinguir diferentes fonemas ou dígitos com base nas características de frequência. Diferenças nas médias espectrais entre dígitos podem refletir variações na pronúncia ou no timbre da voz.

2) *Spectral Edge Frequency* (SEF): a *Spectral Edge Frequency (SEF)* é definida como a frequência abaixo da qual uma certa percentagem (geralmente 95%) da energia total do sinal está contida. Essa medida é particularmente útil para caracterizar a largura de banda efetiva de um sinal e para identificar a presença de componentes de alta frequência. Em análises de áudio, pode ajudar a distinguir entre diferentes tipos de sons ou qualidades de voz, baseando-se na distribuição de energia espectral.

3) Variâncias espectrais: a variância espectral é uma medida da dispersão ou da distribuição da energia espectral ao longo das frequências. Fornece informações sobre a uniformidade ou a irregularidade do espetro de frequência de um sinal. No contexto de análise de áudio, a variância espectral reflete a diversidade e a complexidade das características espectrais presentes no sinal. Variações acentuadas na variância espectral entre dígitos diferentes podem indicar diferenças de pronúncia, entoação ou qualidade vocal.

Concluindo, as três principais características espectrais – médias espectrais, *spectral edge frequency* (SEF) e as variâncias espectrais – capturam aspetos distintos da distribuição de energia ao longo das frequências, da largura de banda efetiva e da uniformidade do espetro de frequência dos sinais de voz. Juntas, essas características permitem uma análise abrangente e discriminativa dos sinais.

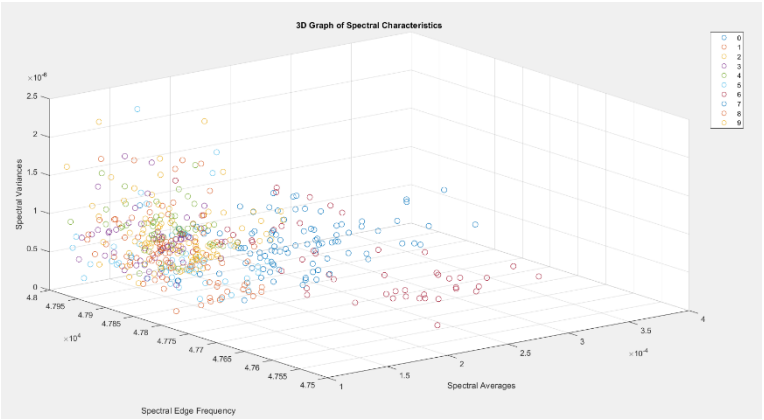

Fig. 20 – *3D Scatter plot* para "Spectral Variances vs. *Spectral Edge Frequency* vs. Spectral Averages"

Meta III

Cálculo da STFT (*Short-time Fourier Transform*)

Ao calcular a STFT, para cada dígito, podem ser utilizadas diferentes parametrizações (tamanho da janela, sobreposição entre janelas, número de pontos para cálculo da FFT, …). Ao efetuar alterações a estes parâmetros, tanto a resolução temporal como a resolução espectral são afetadas, podendo impactar a capacidade de identificação dos dígitos.

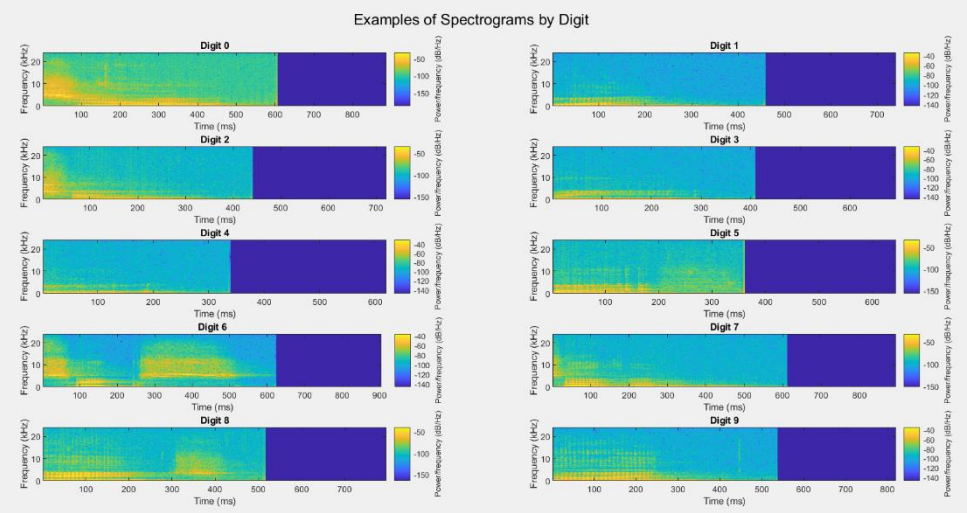

Fig. 21 – Espectrogramas, para os dígitos de 0-9

1) Tamanho da janela: o tamanho da janela determina a resolução temporal do espectrograma. Janelas maiores oferecem melhor resolução de frequência (espectral), pois são capazes de distinguir melhor frequências próximas. No entanto, janelas maiores sacrificam a resolução temporal, tornando o espectrograma menos sensível a mudanças rápidas no sinal ao longo do tempo. Para identificar dígitos, é importante encontrar um equilíbrio entre ambas as resoluções. Janelas menores (por exemplo, 256) podem ser adequadas para capturar mudanças rápidas no sinal, enquanto janelas maiores (por exemplo, 512) podem ser preferíveis se a precisão nas frequências específicas for mais importante.

2) Sobreposição entre janelas: a sobreposição entre as janelas afeta diretamente a suavidade da resolução temporal do espectrograma. Um valor alto de sobreposição (por exemplo, 0.5), resulta numa melhor resolução temporal, pois há mais informações sobrepostas entre as janelas consecutivas. Isso é útil para calcular mudanças rápidas ao longo do tempo. No entanto, uma maior sobreposição aumenta o processamento computacional necessário e pode introduzir redundâncias nos dados. Um valor de sobreposição menor pode ser preferível se a resolução temporal for menos crítica e a eficiência computacional for um fator importante.

3) Número de pontos para cálculo da FFT: o número de pontos utilizados na transformada de Fourier de curto tempo (FFT) determina a resolução espectral do espectrograma. Mais pontos na FFT resultam numa melhor resolução espectral, permitindo distinguir frequências com mais precisão. Apesar disso, um número maior de pontos na FFT também implica um processamento mais intensivo e um espectrograma mais detalhado, o que pode ser demais, dependendo do objetivo da análise.

Objetivo do trabalho:

Para identificar que dígito está a ser falado, é importante ter uma boa resolução temporal, de forma a capturar as características dinâmicas da fala (como transições rápidas entre fonemas). Além disto, é necessária uma resolução espectral razoável, principalmente para distinguir dígitos que possam ter sobreposição em certas faixas de frequência.

O equilíbrio ideal entre resolução temporal e espectral deve ser encontrado para se obter o melhor resultado. Para tal, devem ser experimentadas diferentes combinações, até atingir a mais adequada.

Análise do gráfico:

Em cada um dos dígitos, há uma secção final sem alterações e que corresponde ao silêncio acrescentado na etapa de pré-processamento.

Os dígitos que apresentam uma maior relação potência (dB) / frequência (Hz) são 1, 3 e 4. Já o dígito em que esta relação é a mais inferior é o dígito 0. Todos os restantes possuem valores mais diversos, situados em toda a gama de valores.

Extração de características de potência em diferentes janelas tempo-frequência

(1) Energia Espectral Média por Banda de Frequência (0-500 Hz)

(2) Energia Espectral Média por Banda de Frequência (500-1500 Hz)

(3) Energia Espectral Média por Banda de Frequência (1500-4000 Hz)

(4) Frequência Dominante (Pico) no Espectrograma

(5) Centro de Gravidade Espectral

Após o cálculo dos espectrogramas, para cada dígito, são extraídas várias características de potência em diferentes janelas tempo-frequência.

A energia espectral média por banda de frequência divide o espectrograma em bandas de frequência específicas e calcula a média da potência em cada banda.

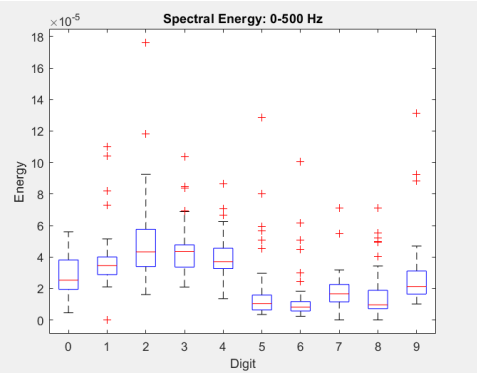

Fig. 22 – Energia espectral média por banda de frequência (0-500 Hz), para os dígitos de 0-9 (1)

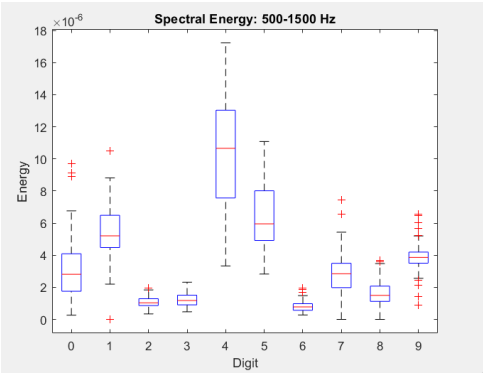

Fig. 23 – Energia espectral média por banda de frequência (500-1500 Hz), para os dígitos de 0-9 (2)

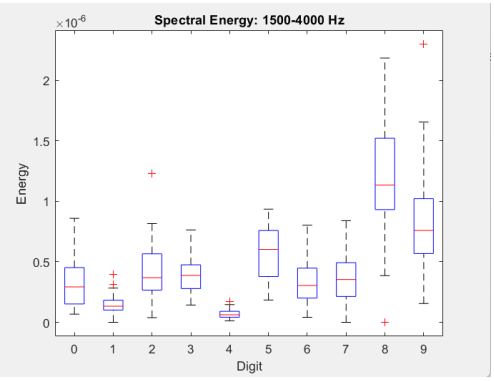

Fig. 24 – Energia espectral média por banda de frequência (1500-4000 Hz), para os dígitos de 0-9 (3)

A frequência dominante (pico) no espectrograma encontra a frequência com a maior potência (pico) em cada espectrograma.

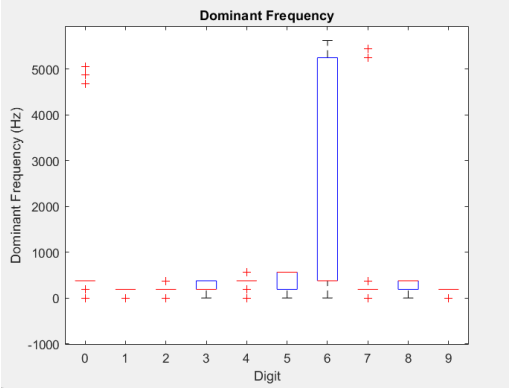

Fig. 25 –Frequência dominante (pico) no espectrograma, para os dígitos de 0-9 (4)

O centro de gravidade espectral calcula o centro de gravidade espectral como a média ponderada das frequências no espectro.

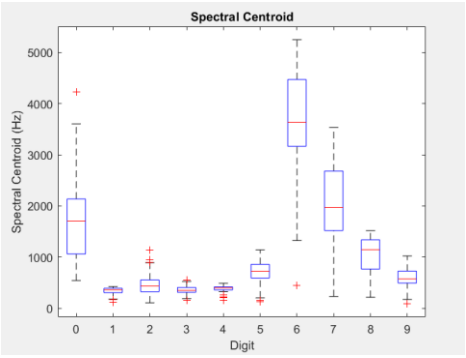

Fig. 26 –Centro de gravidade espectral, para os dígitos de 0-9 (5)

Na primeira *feature* calculada, a distinção entre dígitos não é bem conseguida, uma vez que apenas se observam duas subdivisões, um grupo constituído pelos dígitos de 0 a 4 e outro pelos restantes. Dentro desses grupos, todos os valores são muito próximos, o que dificultaria a criação de condições de seleção.

Na segunda, com bandas de frequência entre 500 e 1500 Hz, é possível distinguir o grupo com os dígitos 2, 3, 6 e 8, outro grupo com o 1 e 5, outro com o 0, 7 e 9 e o dígito 4 isolado dos restantes.

Na terceira, com bandas de frequência entre 1500 e 4000 Hz, pode ser definido o grupo 1 e 4, o grupo 0, 2, 3, 6 e 7, o grupo 5 e 9 e o dígito 8 isolado.

No gráfico de frequência dominante (pico) do espectrograma, os dígitos 3, 5 e 8 têm uma distribuição de valores bastante semelhante e o dígito 6 tem valores muito dispersos na gama, sendo difícil prever o seu valor na maioria dos casos. Por estes motivos, não é uma boa condição a ser escolhida.

Ao analisar o gráfico do centro de gravidade espectral, podemos subdividir os dígitos num grupo composto por 1, 2, 3 e 4, outro por 5 e 9, outro 0 e 7 e os dígitos 8 e 6 podem ser isolados dos restantes.

Principais características de potência discriminatórias

De acordo com as conclusões retiradas anteriormente, as três melhores características são então a energia espectral média por banda de frequência no intervalo 500-1500 Hz, a energia espectral média por banda de frequência no intervalo 1500-4000 Hz e o centro de gravidade espectral.

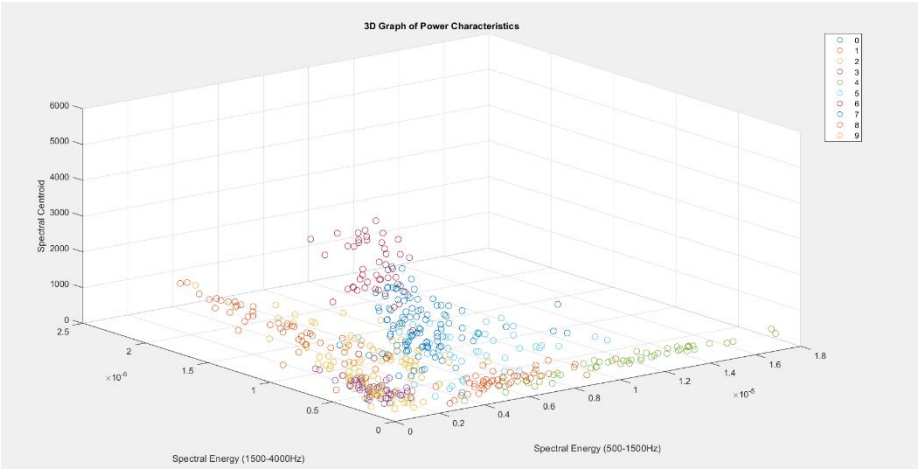

Fig. 27 – 3D *Scatter plot* para "Spectral Energy (500-1500Hz) vs. Spectral Energy (1500-4000Hz) vs. 'Spectral Centroid"

Neste gráfico 3D, é possível ver com clareza os agrupamentos de dígitos e confirmar que as regiões em que se situam seguem um padrão. Por exemplo, o dígito 3 (roxo), encontra-se maioritariamente situado numa pequena região, em que a *Spectral Energy* (1500-4000 Hz) está entre 0 e 1, a *Spectral Energy* (500-1500 Hz) está entre 0 e 0.2 e *Spectral Centroid* apresenta valores muito inferiores, relativamente aos restantes.

Definição de regras de decisão

Após o cálculo de todas as características necessárias, e respetiva avaliação gráfica da distribuição de valores entre os diferentes dígitos, cria-se uma estrutura de decisão, baseada em condições *if … elseif*, com thresholds que definem intervalos de valores nas quais é mais provável encontrar um determinado dígito. Para definir estes intervalos, foram utilizados os quartis 25 e 75 das várias características, que definem onde a maior parte dos valores se encontram, para cada dígito.

Percentagem de acertos

Através da aplicação da fórmula abaixo, foi obtida uma taxa de acertos de 22.20% (seriam necessário mais testes para otimizar este valor).

$$\text{Acertos} = \frac{\text{Dígitos identificados corretamente}}{\text{Total de dígitos}} x\ 100$$

```
Total de dígitos: 500
Previsões corretas: 111
Percentagem de acertos: 22.20%
>>
```

Fig. 28 – Percentagem final de acertos

Melhores características discriminatórias globais

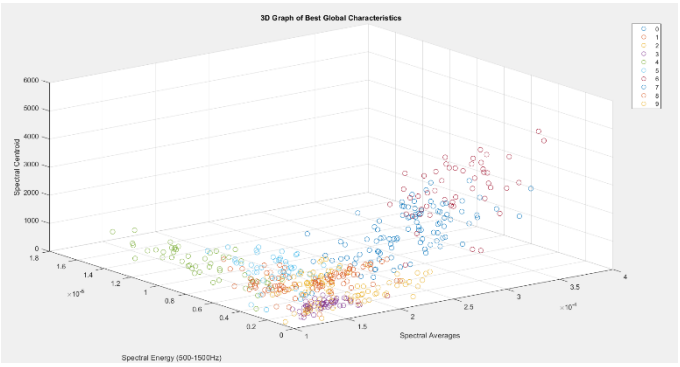

Fig. 29 – Opção 1: 3D *Scatter plot* para "Spectral Energy (500-1500Hz) vs. Spectral Averages vs. 'Spectral Centroid"

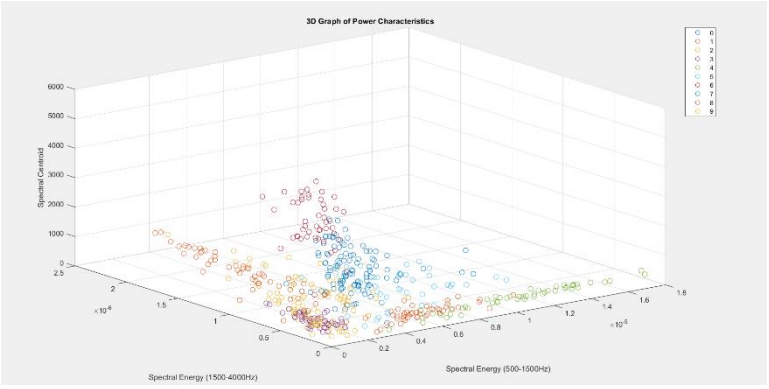

Fig. 30 – Opção 2: 3D *Scatter plot* para "Spectral Energy (500-1500Hz) vs. Spectral Energy (1500-4000Hz) vs. 'Spectral Centroid"

Ao analisar todos os gráficos de um ponto de vista global, há duas escolhas possíveis que se destacam das restantes:

1) Energia espectral média por banda de frequência (500-1500 Hz), médias espectrais e centro de gravidade espectral.

2) Energia espectral média por banda de frequência (500-1500 Hz), energia espectral média por banda de frequência (1500-4000 Hz), e centro de gravidade espectral.

Após fazer o *plot* 3D de ambas, confirma-se que a melhor escolha é a segunda opção, isto é, todas as características que foram selecionadas na meta anterior.

É nítida a melhoria dos agrupamentos ao selecionar a energia espectral média na banda de frequências 1500-4000 Hz como característica para o eixo y. Todos os pontos de uma determinada cor estão mais próximos entre si.