

# RELATÓRIO PROJETO

## Análise e Transformação de Dados

2021



Realizado por:

Daniel Ferreira Veiga 2019216891

Guilherme Sobreiro Almeida 2019224555

Telmo Filipe Queirós da Silva Correia 2019224775

## Introdução

O objetivo deste projeto visa a análise e o tratamento de dados usando sinais recolhidos de um acelerómetro triaxial de um smartphone com a finalidade de determinar qual a atividade realizada pela pessoa que o usa.

Assim, para a realização deste projeto iremos ter em conta 12 movimentos distintos, sejam eles estáticos de transição ou dinâmicos. Para isto iremos recorrer a ferramentas dadas nas aulas como as Transformadas de Fourier Discretas e para o último ponto do trabalho as Transformadas de Fourier em Janelas.

## Desenvolvimento

### 1.

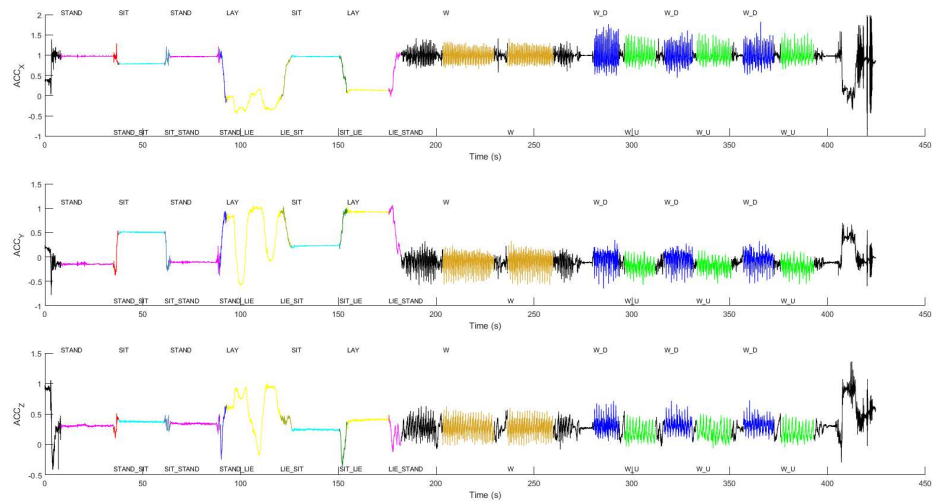
Para importar os sinais fornecidos no dataset, fez-se recurso das funções `importExps.m` e `importLabel.m`, estas acedem respetivamente as diretorias “data” (onde se encontram os ficheiros de texto com as experiências) e “label” (onde se encontra o ficheiro de texto `labels` que identifica temporalmente as atividades em cada uma das experiências). Depois destas recolherem a informação relevante para as análises futuras, guardam a mesma como parâmetros de uma tabela “data” que nos permite organizar melhor a informação.

Salienta-se alguma informação útil sobre o dataset utilizado, as experiências alvo de teste vão da 50 a 57, tendo estas sido recolhidas por acelerómetros triaxiais e têm uma frequência de amostragem de 50Hz.

### 2.

Depois dos sinais importados, fazemos um tratamento de informação relativamente as atividades e dividimos os sinais pelos seus eixos XYZ para se conseguir analisar a informação correspondente a cada atividade. Consequentemente fazemos um *subplot* da mesma para cada eixo, com a sua atividade correspondente identificada. Este consiste na representação gráfica de um sinal colorido a preto quando a sua informação não pertence ao conjunto de atividades alvo de estudo. Sendo este colorido a cores definidas na função `defineColor.m` e é acrescentada uma legenda a atividade a que corresponde a informação.

Como se segue no exemplo do gráfico relativo à experiência 50 apresentado abaixo



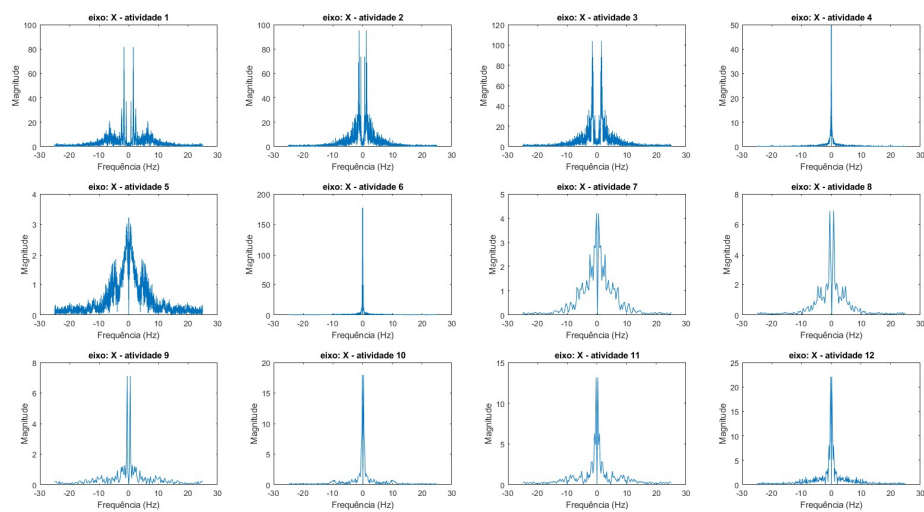
### 3.1

Para o cálculo da DFT de cada fragmento usou a função “DFT.m” aplicada a um ciclo obrigando-o a recolher todas as DFT de cada atividade em cada experiência, sendo esta depois guardada num parâmetro da tabela “data”.

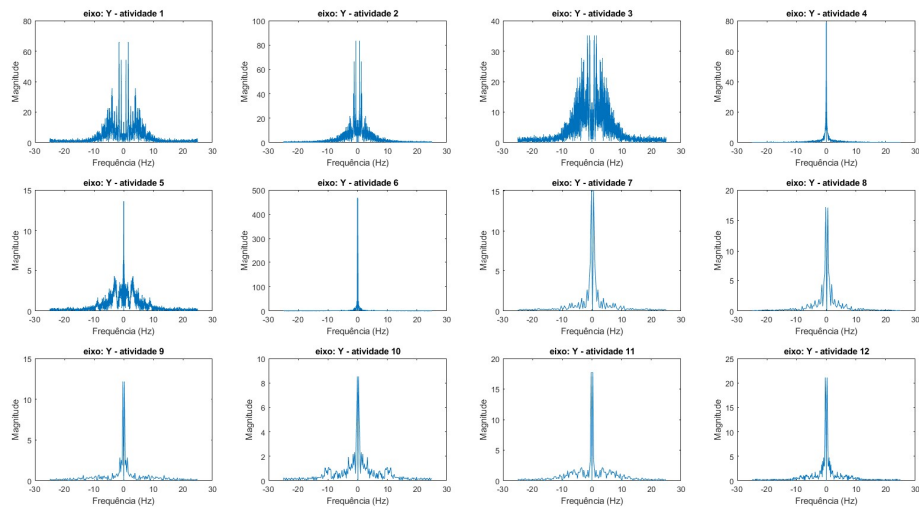
Procede-se á divisão do sinal pelos três eixos, XYZ, para serem representada graficamente a DTF de cada atividade.

As três imagens seguintes são o exemplo gráfico relativo á experiência 50.

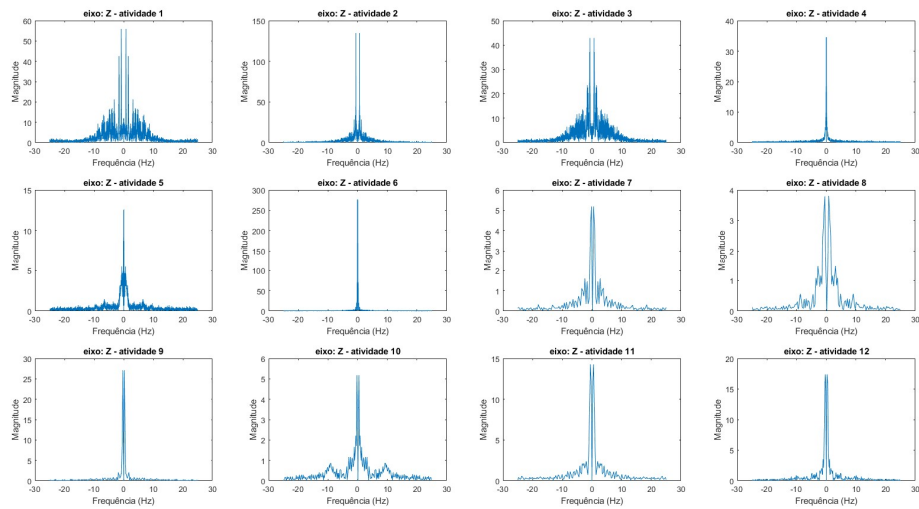
Para o eixo xx



Para o eixo yy



Para o eixo zz



Depois de uma análise dos resultados é possível ver que os sinais contêm secções com componentes de frequência diferentes o que torna o tipo de janela utilizada para apresentar a DFT menos vantajosa. Pois por vezes ocorrem variações bruscas nos extremos. O que faz com que a sua fiabilidade seja afetada, e também permite constituir descontinuidades no sinal, uma vez que a DFT assume o sinal como periódico.

### 3.2.

Para identificar as características mais relevantes de cada atividade optou-se por analisar os intervalos de frequência para as atividades em questão. Estes foram obtidos através do uso das funções `calculate_limits_freq.m` e `find_prevalent_frequency.m`.

Estas funções guardam e calculam as frequências mais relevantes para cada atividade obtidas a partir da DFT, depois são convertidas num intervalo de valores, que esta representado abaixo.

```
Atividades Dinamicas
Limites para a atividade 1 no eixo X : [1.5146; 1.7813], no eixo Y : [1.5146; 1.7813], no eixo Z : [0.72853; 1.7813]
Limites para a atividade 2 no eixo X : [1.1886; 1.7324], no eixo Y : [0.59429; 1.7324], no eixo Z : [0.59429; 1.7324]
Limites para a atividade 3 no eixo X : [1.4706; 1.9004], no eixo Y : [0.71706; 3.3403], no eixo Z : [0.69573; 3.8156]
Atividades Estaticas
Limites para a atividade 4 no eixo X : [0.016469; 0.70309], no eixo Y : [0.024343; 0.24205], no eixo Z : [0.024343; 0.2651]
Limites para a atividade 5 no eixo X : [0.028813; 3.7487], no eixo Y : [0.0085005; 0.057059], no eixo Z : [0.0097238; 0.29012]
Limites para a atividade 6 no eixo X : [0.026251; 0.074019], no eixo Y : [0.018505; 0.061627], no eixo Z : [0.018505; 0.061627]
Atividades de Transicao
Limites para a atividade 7 no eixo X : [0.1497; 3.125], no eixo Y : [0.13369; 0.4491], no eixo Z : [0.13369; 0.66667]
Limites para a atividade 8 no eixo X : [0.13812; 1], no eixo Y : [0.13812; 0.5], no eixo Z : [0.13812; 0.5814]
Limites para a atividade 9 no eixo X : [0.22124; 0.75], no eixo Y : [0.22321; 1.4444], no eixo Z : [0.11111; 0.4717]
Limites para a atividade 10 no eixo X : [0; 0.33784], no eixo Y : [0; 0.4902], no eixo Z : [0; 0.4902]
Limites para a atividade 11 no eixo X : [0; 1.2048], no eixo Y : [0; 0.91463], no eixo Z : [0; 0.60241]
Limites para a atividade 12 no eixo X : [0.10288; 0.40323], no eixo Y : [0.10288; 0.60241], no eixo Z : [0.10288; 1.6129]
```

### 3.3.

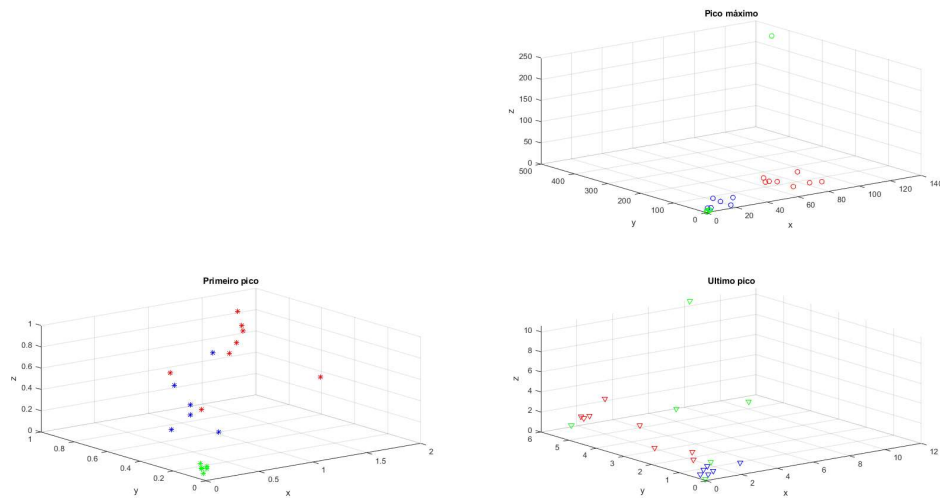
Para esta questão, recorreremos à DFT do sinal, para as atividades dinâmicas, onde com recurso à função `find_prevalent_frequency.m` obtém-se as frequências mais relevantes. Desta forma ao multiplicarmos a frequência obtida por 60, o valor resultante permite-nos chegar a uma aproximação de passos por minuto. Dado que a frequência é um valor muito próximo dos passos por segundo realizados por um individuo. Realizou-se o processo para as experiências do mesmo utilizador, uma vez que era pedido a média e desvio padrão para cada utilizador recorreu-se respetivamente as funções `mean` e `std` para obter os valores pretendidos.

Os valores apresentados abaixo correspondem a média e desvio padrão das atividades dinâmicas para o utilizador 25, que é composto pelas experiências 50 e 51.

		Medias	Desvios
Atividade 1	(W)	92.7469	2.6484
Atividade 2	(W_U)	88.5411	8.5964
Atividade 3	(W_D)	91.4275	4.5144

É de notar que os valores variam consideravelmente de indivíduo para indivíduo. Esta diferença depende de muitos fatores, nomeadamente, o peso, o sexo, a altura, a idade e a hora do dia em que foi realizada a experiência.

### 3.4.



Para este exercício resolvemos calcular 3 picos relevantes e analisá-los, sendo estes o primeiro pico, o pico máximo e o último pico de todos os eixos para cada atividade.

Exemplifica-se através da imagem acima que analisa a informação para a experiência 50. De forma a distinguir os diferentes tipo de atividades usaram se cores, sendo as atividades dinâmicas representadas pela cor vermelha, as estáticas pela cor verde e as de transição pela cor azul.

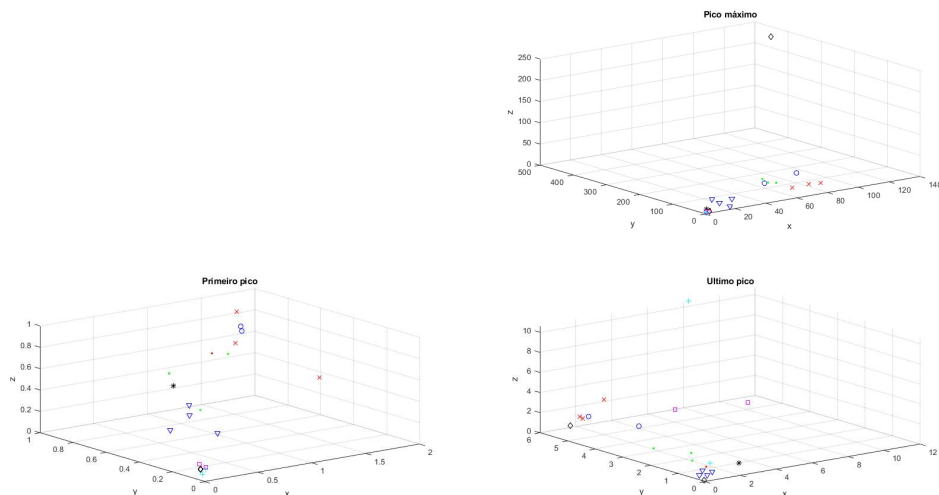
Ao observar os 3 gráficos, concluímos que os gráficos resultantes do pico máximo e do primeiro pico são os que apresentam separações mais visíveis, tendo, no entanto, alguns outliers, visíveis que afetam a performance da sensibilidade e especificidade.

No gráfico que representa os picos máximos é possível encontrar uma separação visível entre as atividades dinâmicas das de transição. No entanto o mesmo não se verifica para as atividades estáticas que onde a informação se encontra toda dispersa e misturada com as outras duas atividades o que afeta significativamente a performance da sensibilidade e especificidade.

### 3.5.

À semelhança do exercício anterior neste é pedido o mesmo, mas desta vez sendo entre todas as atividades e não entre os seus grupos. Dado a semelhança dos exercícios apenas houve pequenas alterações na função anterior, que permitisse distinguir as diferentes atividades nos gráficos.

Obteve-se então o gráfico abaixo para a mesma experiência



Dada a análise dos gráficos do exercício anterior é espectável que tanto os gráficos que representam o primeiro pico e o pico máximo sejam os iram obter uma melhor performance da sensibilidade e especificidade que o gráfico que representa o último pico.

Fazendo uma análise mais detalhada é possível ver que no gráfico do primeiro pico existe alguma dispersão das frequências da mesma atividade o que faz com que a sua performance da sensibilidade e especificidade seja ligeiramente afetada.

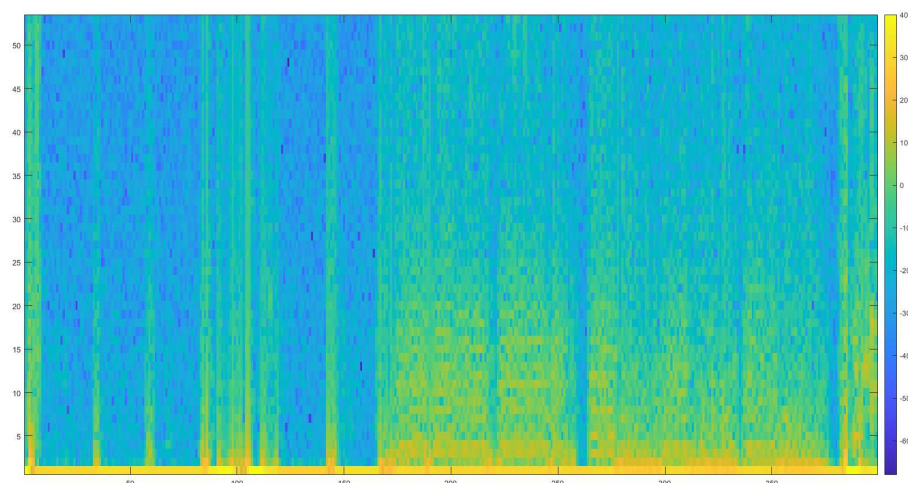
Para o gráfico do pico máximo é possível ver uma boa separação onde existe apenas dois outliers, no entanto existem atividades que apesar de terem todas as frequências juntas estão muito próximas umas das outras. Acontecendo o mesmo com o gráfico do ultimo pico. Desta forma a performance da sensibilidade e especificidade seja pouco afetada.

#### 4.

Neste exercício, aplicamos o Short-Time Fourier Transform (STFT) na experiência 50 para obtermos as distribuições para o sinal do acelerômetro no eixo Z.

Depois de fazermos uma função para a STFT, fomos testando vários valores para o tamanho da janela e para a sobreposição, de maneira a conseguirmos uma boa visibilidade do gráfico, também convertemos unidades para dB e usamos a escala logarítmica.

Obtivemos o seguinte espectrograma.



Como você pode ver na imagem, há muitas semelhanças entre as duas parcelas. No espectrograma, as áreas amarelas correspondem aos locais onde a densidade de pontos é maior no gráfico da função de desenvolvimento. A separação entre as várias atividades também é evidente no espectrograma. O momento em que um indivíduo começa a se envolver em atividade dinâmica é fácil de perceber (a prevalência do amarelo é maior neste momento), enquanto a distinção entre atividade estática e transicional é mais difícil de distinguir. Mesmo assim, os picos correspondentes à atividade transicional são notáveis porque são muito mais frequentes do que a atividade estática.