

ATD

Mini-Projeto

Análise e Transformação de Dados (ATD)

Departamento de Engenharia Informática

2018-2019

David Jesus Vaz Cortesão Silva	2008109004	davidvazc@gmail.com
João Manuel Silva Antunes Neves de Carvalho	2001027493	joaomcarvalho@sapo.pt
Luís Pedro Morais Gonçalves	2013167905	lpgoncalves@outlook.pt

Índice

I - Importação e representação de sinais	2
II - Ponto 4 - Calcular a DFT de cada porção do sinal associado a uma atividade	3
4.1 Comparação dos diferentes tipos de janelas	3
4.2 Estatística do número de passos por minuto para as actividades dinâmicas	6
4.3 Características espectrais - diferenciar atividades estáticas e de transição de atividades dinâmicas	9
4.4 Características espectrais - diferenciar entre os diferentes tipos de atividades	12
4.5 Características espectrais - diferenciar entre atividades dinâmicas	13
III - Ponto 5 - Distribuições tempo-frequência - Short-time Fourier Transform (STFT)	14

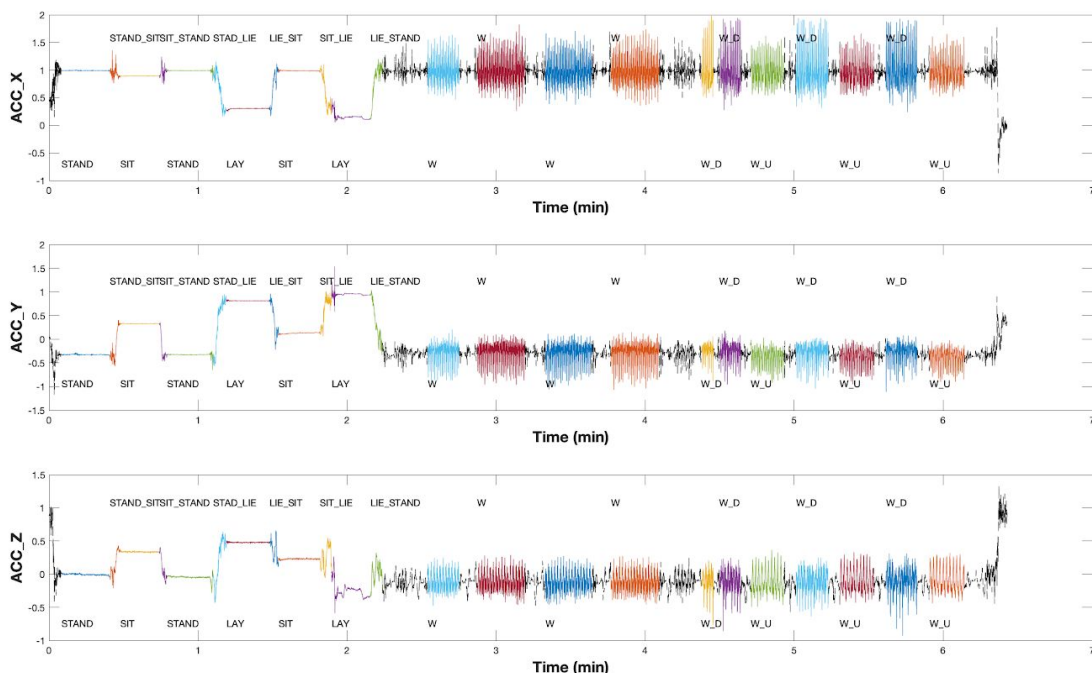
I - Importação e representação de sinais

Neste projeto os dados utilizados como caso de estudo correspondem ao primeiro grupo, nomeadamente “PL1: *acc_exp01_user01.txt* a *acc_exp10_user05.txt*”.

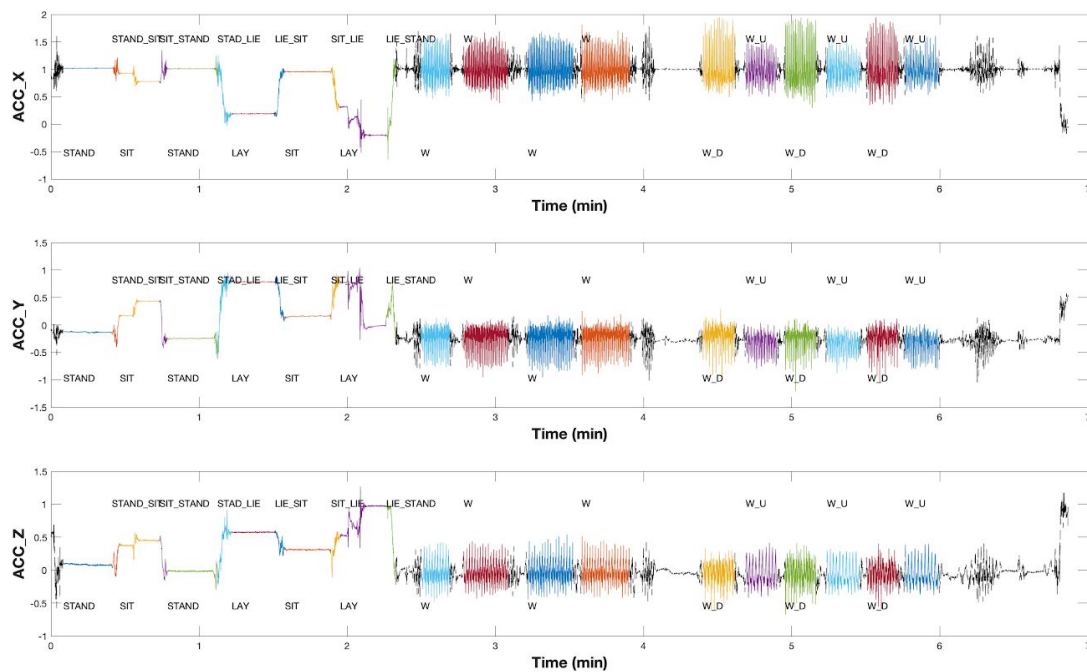
O ponto de partida de todo o trabalho passou por carregar a informação disponibilizada nos ficheiros, processá-la e apresentá-la graficamente.

Neste seguimento, começamos por definir um vetor, (“acc_file”), que contém todos os nomes dos ficheiro que pretendemos carregar. Posteriormente, este vetor é percorrido através de um ciclo *for* e para cada entrada é feito o respectivo *importfile*. Além disso, neste ponto são também definidas, individualmente, variáveis para as *activities* (“W”, “STAND”, “LAY”, etc), *colours*, (“y”, “m”, “c”) e os *sensors* (“ACC\X”, “ACC\Y”, “ACC\Z”).

Por fim, são definidos para cada um dos *n_plots*, as características dos eixos, tipo de letra, legendas, etc, e finalmente desenhados cada um dos gráficos, como apresentado nas figuras seguintes.



Visualização dos dados relativos ao ficheiro *acc_exp01_user01.txt*



Visualização dos dados relativos ao ficheiro acc_exp02_user01.txt

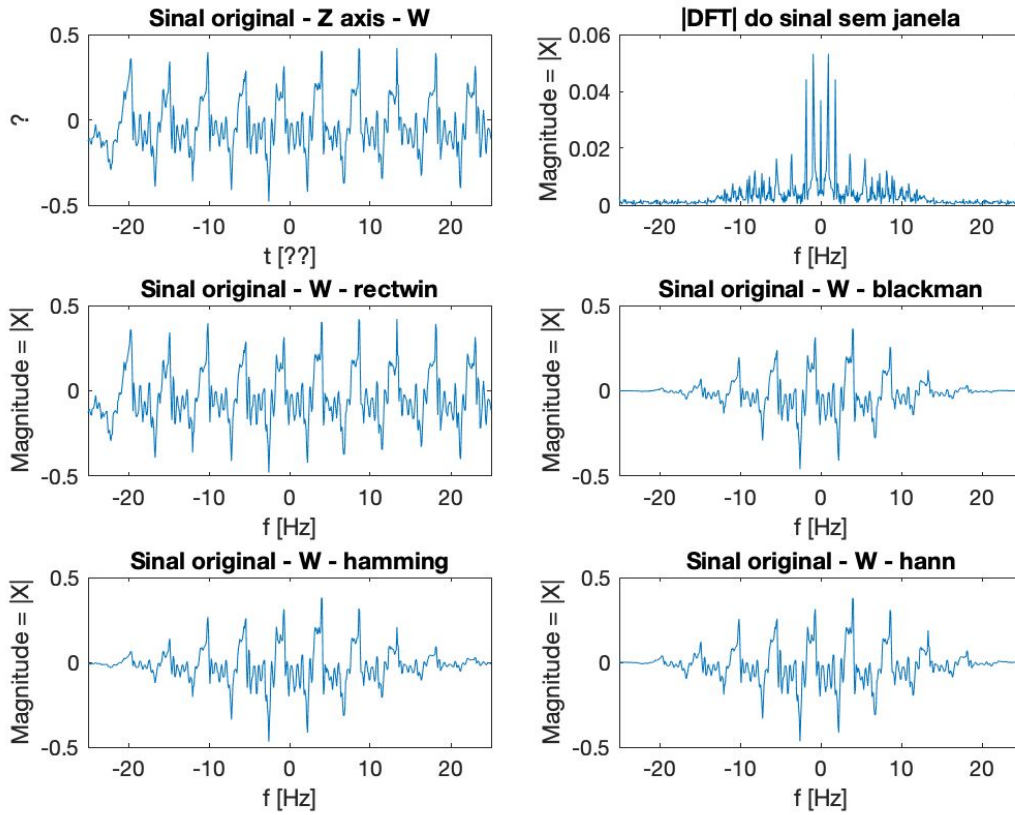
II - Ponto 4 - Calcular a DFT de cada porção do sinal associado a uma atividade

4.1 Comparação dos diferentes tipos de janelas

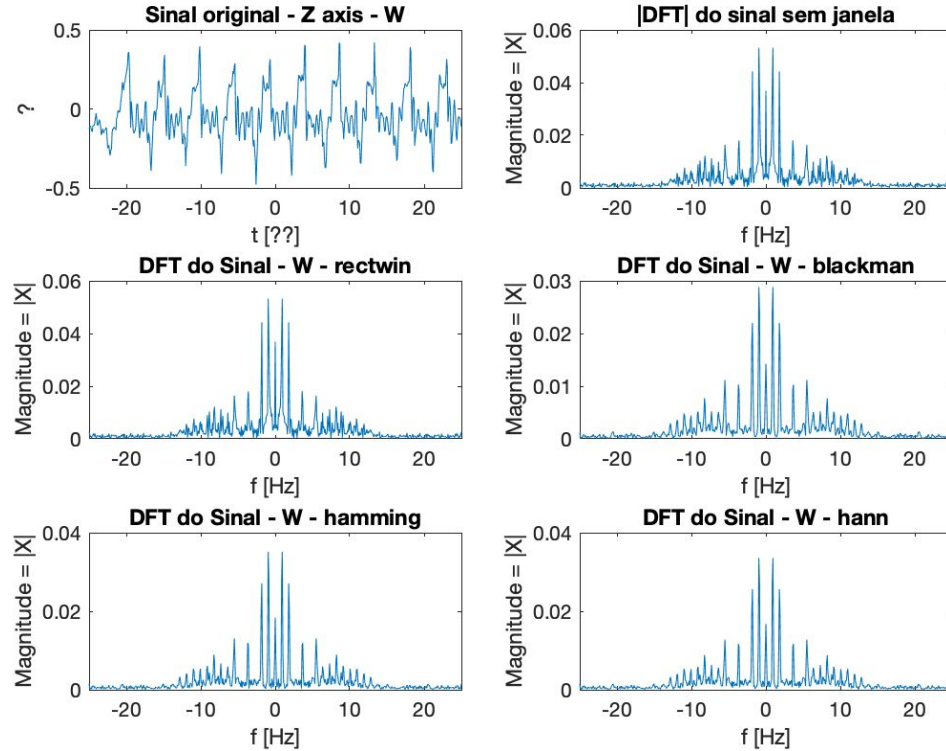
Foram testadas vários tipos de janelas. Embora saibamos que as janelas tipicamente têm por objetivo a suavização das partes laterais do sinal, dando maior relevância à parte central, cada uma delas apresenta um comportamento algo distinto.

Foram analisadas as seguintes janelas:

- rectwin
- blackman
- hamming
- hann



Visualização do sinal original e do efeito da aplicação das janelas sobre o sinal.

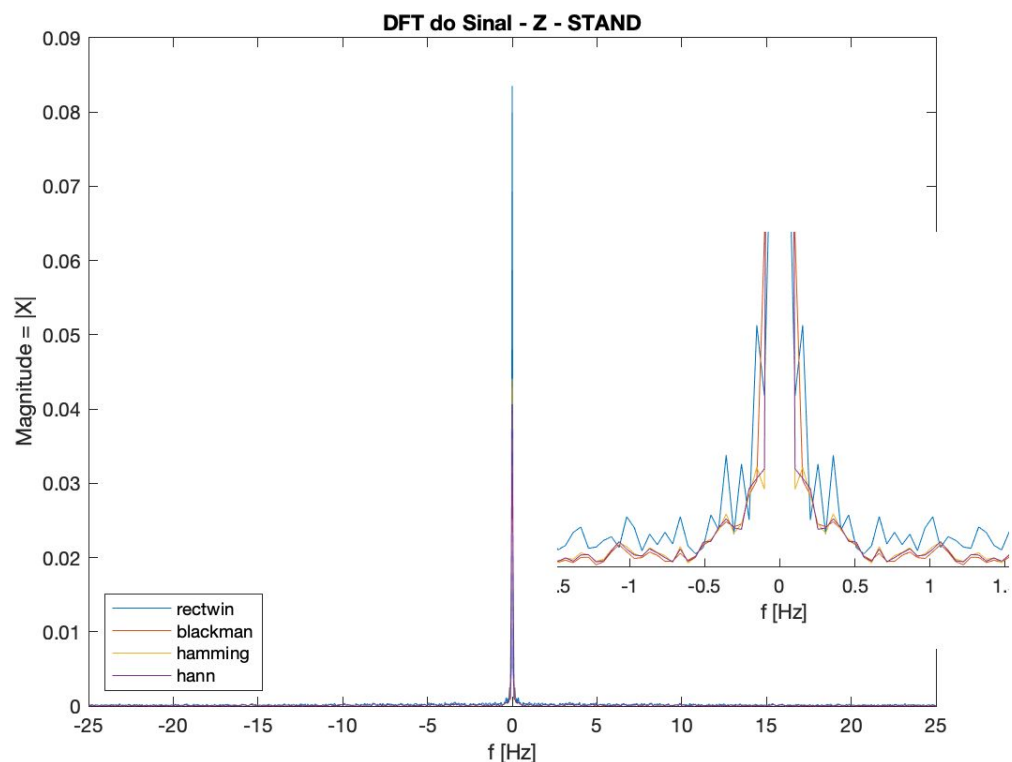


Visualização do sinal original (Walking) e do efeito de aplicação das janelas antes do cálculo da DFT

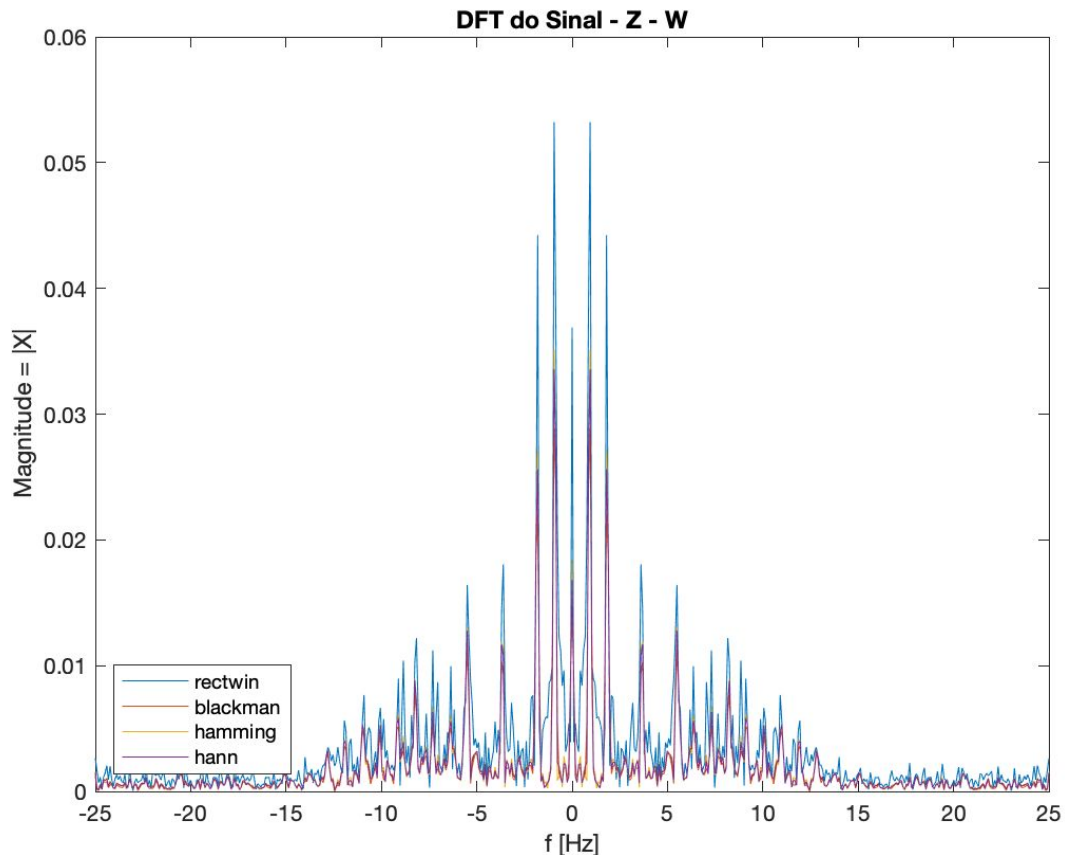
A **rectwin** não tem qualquer efeito neste caso, dado ser aplicada à totalidade do sinal. Pode ser útil caso se pretenda remover uma parte do sinal, mantendo o restante igual à representação original.

As restantes janelas - **Blackman**, **Hamming** e **Hann** - têm um comportamento semelhante, isto é, reduzem os lobos laterais e dão ênfase ao lobo central. A grande diferença entre a Hamming e a Hann reside no facto de nesta os lobos laterais alcançarem o valor zero no eixo dos X.

No tratamento dos dados, ao longo dos pontos seguintes do trabalho, optámos por utilizar a janela de Hamming.



Visualização do efeito das diferentes janelas sobre um sinal STAND no eixo Z, com área ampliada sobreposta.



Visualização do efeito das diferentes janelas sobre um sinal WALK sobre o eixo Z.

4.2 Estatística do número de passos por minuto para as actividades dinâmicas

No desenvolvimento desta questão procuramos primeiramente tratar o sinal original. Neste ponto, no seguimento da questão 4.1, aplicamos a janela que a nosso ver apresentou melhores resultados nomeadamente a janela “**hamming**”.

A aplicação de uma janela sobre o sinal original permite suavizar as extremidade do mesmo dando especial relevância à composição central deste.

De seguida aplicamos sobre o módulo deste sinal (“*abs*”) a função “*fftshift*” que nos retorna a DFT do respectivo sinal unicamente para valores positivos.

A parte fulcral desta questão passou por determinar o segundo pico de maior frequência, ou seja, o primeiro pico diferente de zero presente na DFT. Para isso fizemos uso da função

“*findpeaks*” que nos devolve dois parâmetros, nomeadamente, a amplitude do sinal (que guardamos em “pks”) e respectivo índice (que guardamos em “locs”) dos picos mais relevantes da função DFT.

A relevância destes varia de acordo com o limiar “*MinPeakProminence*” definido a 8 após algumas tentativas. Como seria de esperar, o pico mais relevante do sinal é o pico central e, como tal, temos de determinar a posição deste para posteriormente o descartar e aceder ao pico mais próximo. Acontece que este ponto nem sempre está exatamente na posição zero e portanto tivemos que realizar uma pesquisa nas zonas fronteiriças de zero.

Após esta etapa, e mantendo o índice do ponto central guardado na variável “index”, sabemos que o ponto que realmente nos interessa está na posição “index+1”. Desta feita, obtendo a frequência deste ponto e multiplicando por 60 (segundos) obtemos o número de passos por minuto. É importante referir que as frequências do sinal estão contidas no vetor *f* previamente definido.

Para cada atividade, guardamos o número de passos no vetor *saveSteps*, para que no final fosse possível determinar a média e o desvio padrão (“*mean*”, “*std*”) representados em baixo para cada eixo estudado (x,y,x).

Para o eixo do X's

	<i>Média</i>	<i>Desvio Padrão</i>
WALKING	97.4372	20.1759
WALKING_UPSTAIRS	79.9060	23.6199
WALKING_DOWNSTAIRS	98.7355	11.1867

Para o eixo dos Y's:

	<i>Média</i>	<i>Desvio Padrão</i>
WALKING	88.0782	23.6243
WALKING_UPSTAIRS	86.1672	21.1695
WALKING_DOWNSTAIRS	99.7121	17.6043

Para o eixo dos Z's:

	<i>Média</i>	<i>Desvio Padrão</i>
WALKING	53.9215	2.7120
WALKING_UPSTAIRS	46.8829	3.8673
WALKING_DOWNSTAIRS	54.4742	14.2951

Analisando os dados constata-se alguma consistência nos valores, nomeadamente há um acréscimo do número de passos na atividade WALKING_DOWNSTAIRS e o respectivo decréscimo na atividade WALKING_UPSTAIRS, para cada um dos 3 eixos.

Contudo, os valores obtidos no eixo dos z's não nos parecem ir de encontro ao teoricamente expectável, uma vez que seria neste eixo que os picos nas funções seriam mais perceptíveis, situação esta que não vai de encontro aos nossos resultados.

Na nossa perspectiva existem uma série de fatores que podem perturbar a experiência e como tal corromper os resultados obtidos. Exemplo disso é o facto do pico estudado seguir a tendência do ponto central, zero, e deste modo apresentar um valor para a frequência acima daquilo que seria suposto. Esta situação implica que haja situações em que o ponto de estudo mais promissor seja o 2º ou 3º de maior frequência e diferente de zero.

4.3 Características espectrais - diferenciar atividades estáticas e de transição de atividades dinâmicas

Com o intuito de diferenciar actividades estáticas das actividades dinâmicas, procedemos a duas abordagens distintas. Em ambas foram analisados os primeiros picos de maior frequência diferente de zero, contudo, numa delas, a detecção foi feita sobre o sinal original, enquanto na outra foi feita sobre a DFT de cada sinal.

Na implementação das duas abordagens socorremo-nos do código já implementado para a pergunta 4.2, porém, com algumas alterações, dado que no código da alínea anterior apenas eram detectados os picos para atividades dinâmicas. Assim, tivemos que alterar o algoritmo para uma detecção sobre todas as atividades, embora partindo sempre do mesmo princípio, alterando apenas o ciclo for inicial.

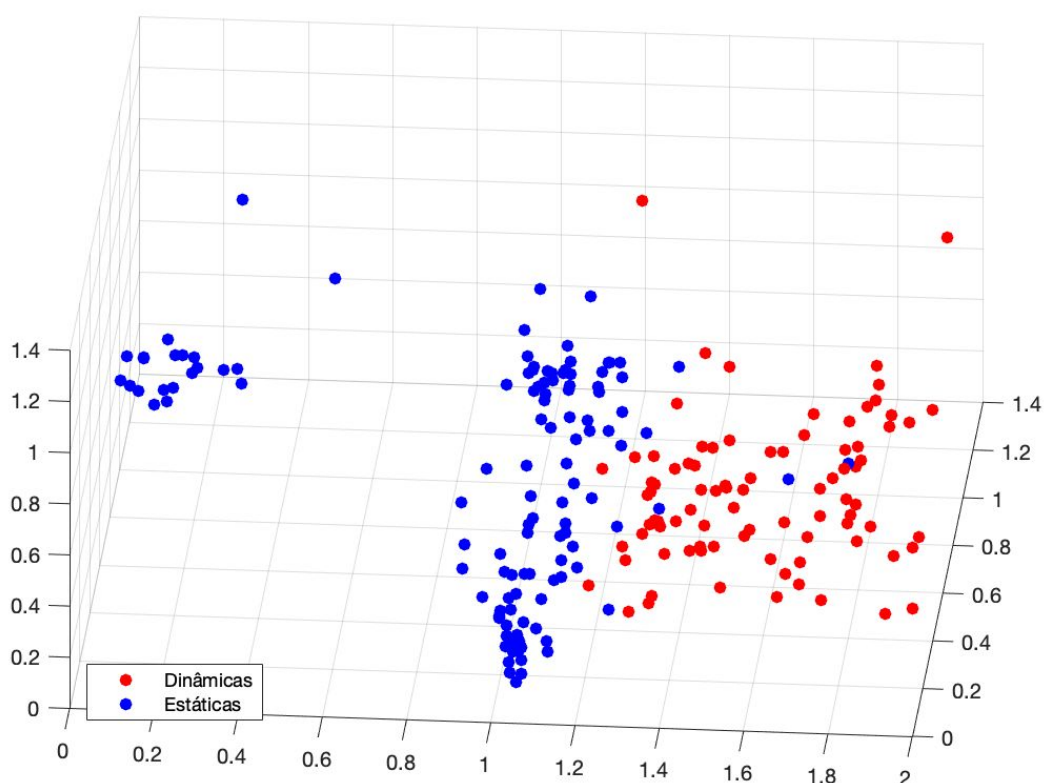
No decorrer desta questão foi também necessário manter informação acerca de todos os picos detectados para cada eixo, por forma a com esses dados construir posteriormente os gráficos para análise.

Para manter os dados guardados até ao momento do plot adicionamos 3 vetores “peaks_X”, “peaks_y” e “peaks_Z” sendo que cada um guarda os picos detectados em sobre cada um dos eixos, respectivamente, x, y e z.

Por fim, realizamos um plot tridimensional, sendo que cada dimensão está associada aos dados de um dos vetores. Para possibilitar uma melhor percepção dos tipos de atividade, associamos a cor vermelha às frequências relativas e a cor azul a atividades estáticas.

Actividades:

Estáticas e de transição	Dinâmicas
4 SITTING	1 WALKING
5 STANDING	2 WALKING_UPSTAIRS
6 LAYING	3 WALKING_DOWNSTAIRS
7 STAND_TO_SIT	
8 SIT_TO_STAND	
9 SIT_TO_LIE	
10 LIE_TO_SIT	
11 STAND_TO_LIE	
12 LIE_TO_STAND	

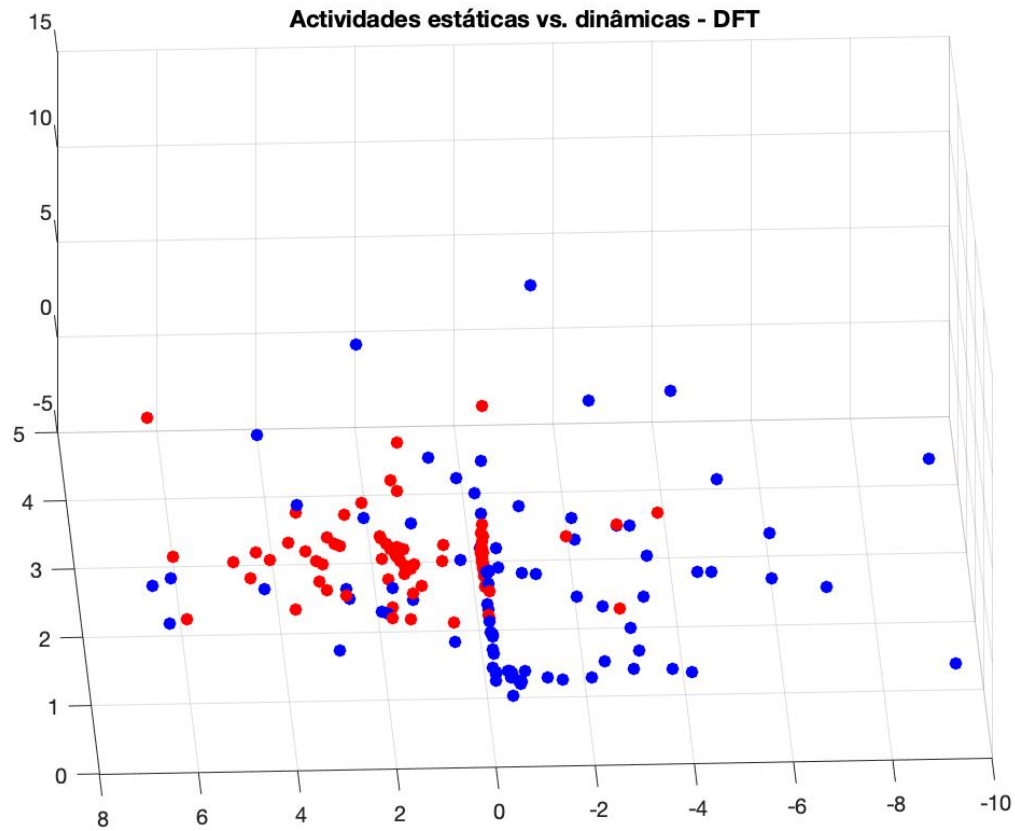


Representação de actividades estáticas e dinâmicas a partir do primeiro pico detectado no sinal raw

Este primeiro gráfico recorre às amplitudes máximas presentes em cada um dos ficheiros com o sinal original.

É possível, sem grande dificuldade, agrupar os dados em dois grupos distintos sendo o grupo a azul relativo a actividades estáticas e o grupo a vermelho relativo a actividades dinâmicas.

Constata-se que as actividades estáticas apresentam, como era de esperar, amplitudes máximas mais baixas e portanto mais próximas de zero, contrariamente às amplitudes de actividades dinâmicas.



Representação de actividades estáticas e dinâmicas a partir do primeiro pico diferente de zero na DFT.

Este gráfico representa os mesmos dados que o gráfico anterior. Contudo, neste último, o processamento dos dados fez-se sobre a DFT do sinal e não sobre o sinal original. Imediatamente depreendemos que a delineação de um segmento de reta que separe os pontos azuis dos pontos vermelhos é bastante mais difícil.

Relativamente à sensibilidade e especificidade, em nenhum dos casos seriam de 100%, uma vez que existem vários pontos considerados falsos positivos e, por sua vez, falsos negativos.

Contudo ambas seriam mais altas para a primeira abordagem, sem DFT.

$$SE = \frac{VP}{(VP + FN)} * 100 \quad ; \quad ESP = \frac{VN}{(VN + FN)} * 100$$

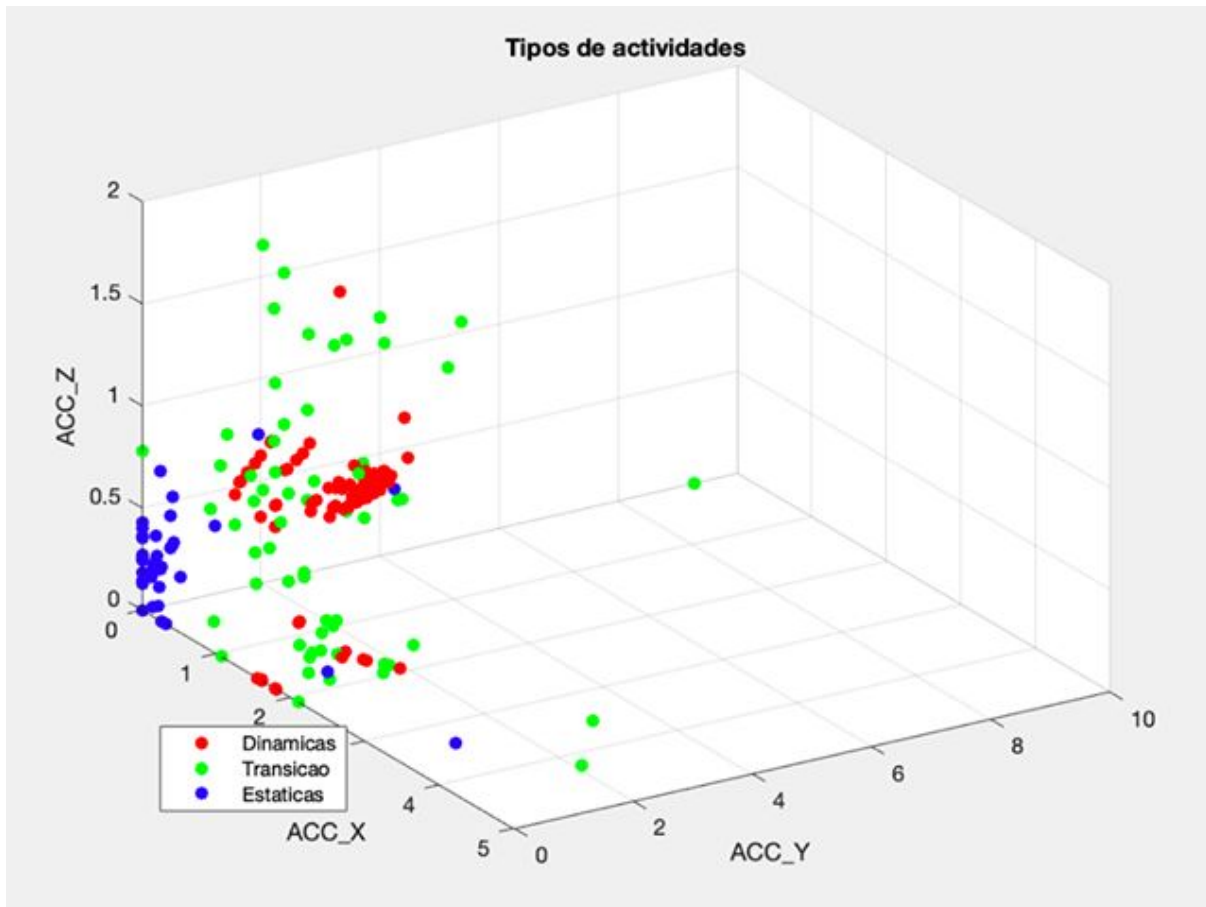
SE - Sensibilidade;
ESP - Especificidade;
VP - Verdadeiros Positivos;
FN- Falsos Negativos;
VN- Verdadeiros Negativos;

4.4 Características espectrais - diferenciar entre os diferentes tipos de atividades

Para obter as características necessárias e conseguir ver claramente as diferenças entre os três tipos de atividade, utilizámos a informação anteriormente obtida. Esta informação, depois de realizada a DFT, e obtidos os picos, foi armazenada em vetores.

Começámos por criar três matrizes vazias, de forma a poder realizar o gráfico com os três tipos de atividade, de seguida percorremos os dados, verificando a que atividade pertencem e guardando na matriz correspondente: ‘dinD’, responsável por armazenar os dados relativos às atividades dinâmicas, ‘dinT’ para as atividades de transição e por último, ‘dinE’ para as atividades estáticas.

Por fim, depois de obtida toda a informação, foi possível através da função ‘*scatter3*’, visualizar todos os picos de cada atividade num só gráfico.

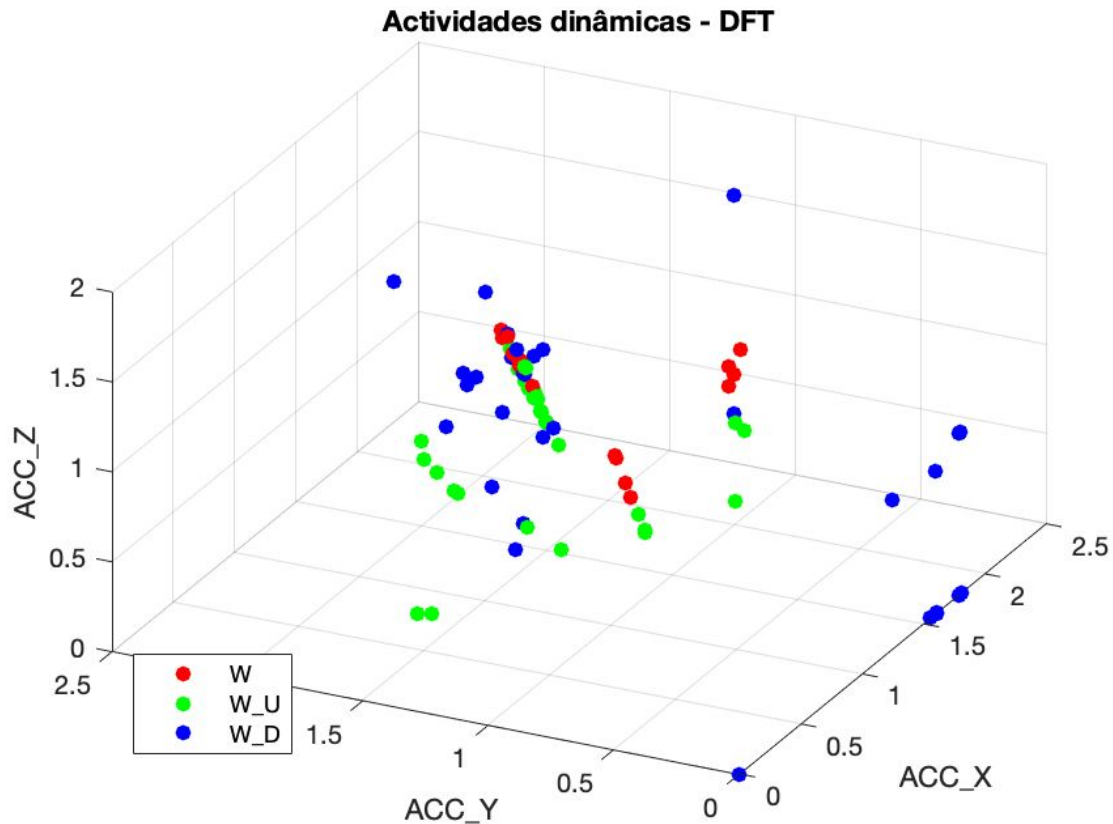


4.5 Características espectrais - diferenciar entre atividades dinâmicas

Nesta questão e no seguimento das anteriores procedemos à detecção do pico de frequências. Para todas as atividades dinâmicas, esta detecção segue o mesmo padrão da abordagem utilizada na pergunta 4.2, contudo guardaram-se, mais uma vez, as frequências obtidas em vetores, para posteriormente proceder ao plot dos dados sobre o respectivo eixo.

Atividades Dinâmicas

- 1 WALKING
- 2 WALKING_UPSTAIRS
- 3 WALKING_DOWNSTAIRS



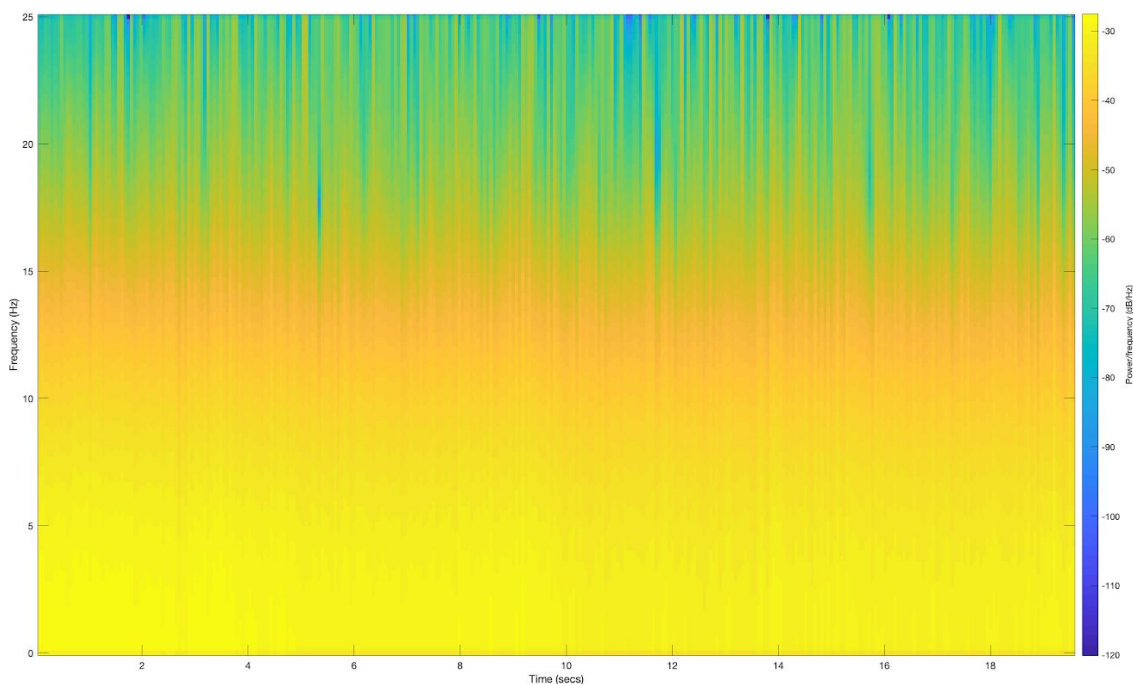
Visualização do primeiro pico diferente de zero, exclusivamente para actividades dinâmicas

Os resultados apresentados não permitem distinguir claramente os diferentes tipos de actividade. Pensamos que os resultados obtidos na análise de picos da DFT não deverão estar de acordo com o esperado, seja pela dificuldade de obter um limiar de detecção funcional para todos os casos (3º parâmetro da função findpeaks) ou mesmo pela influência do ponto máximo em zero sobre os máximos locais adjacentes.

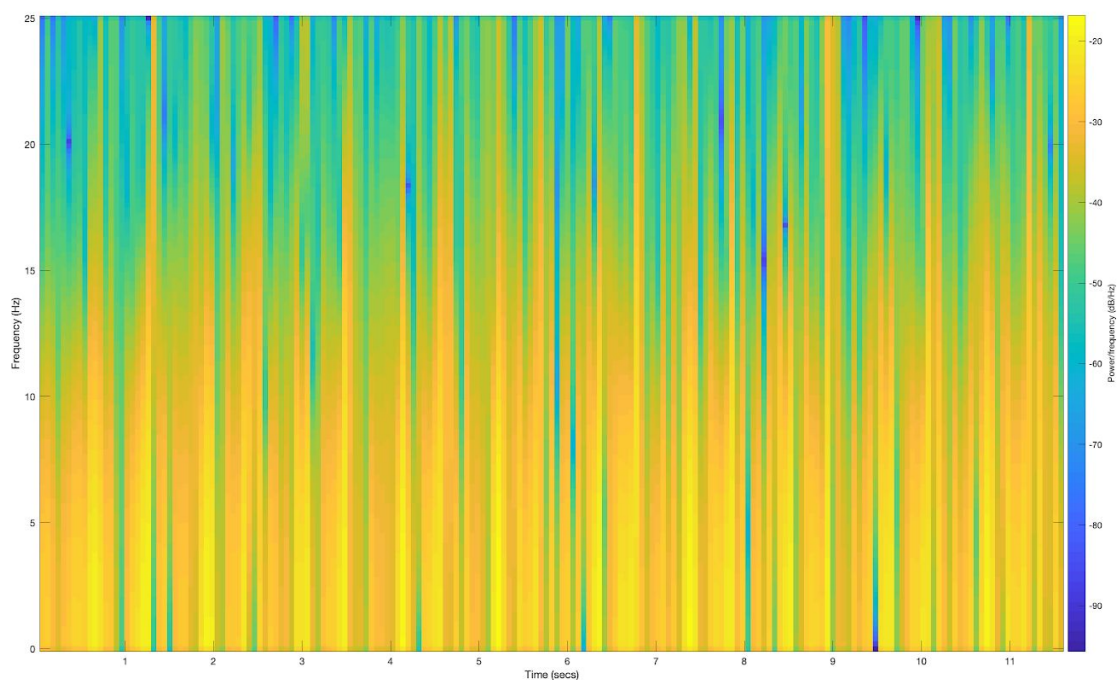
III - Ponto 5 - Distribuições tempo-frequência - Short-time Fourier Transform (STFT)

Para a análise das distribuições tempo-frequência escolhemos os dados do ficheiro *acc_exp01_user01.txt*. No código implementado aplicámos a função FFT sobre uma janela deslizante, com sobreposição parcial, necessária à análise em STFT.

Deixamos aqui alguns dos espectrogramas que gerámos, para uma actividade estática e outra dinâmica. É possível perceber uma maior variabilidade não apenas nas frequências mas também ao longo do tempo.



Actividade estática (STAND)



Actividade dinâmica (WALKING)

O nosso objectivo seria realizar uma análise do sinal não apenas em frequência, como nos pontos anteriores, mas também no tempo. No entanto, por manifesta falta de tempo, não nos foi possível avançar mais neste ponto.