

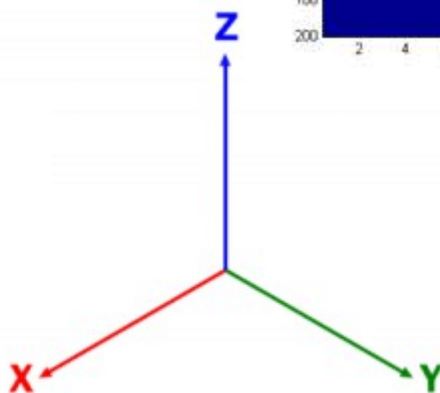
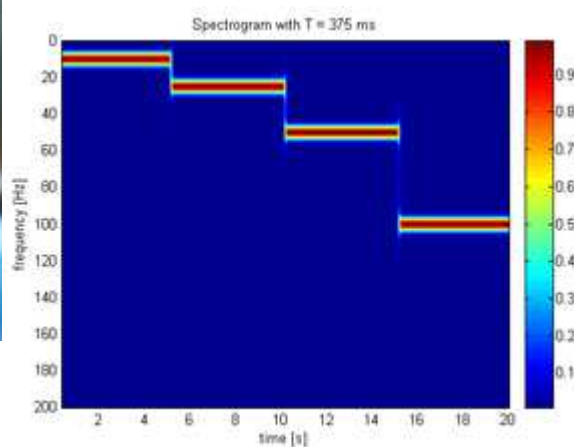


FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA INFORMÁTICA

Disciplina: *Análise e Transformação de Dados*

Relatório – Projeto ATD



Trabalho realizado pelos alunos:

Leonardo Moreira Pina (2019234318)

Luís Filipe Neto (2020215474)

João Fernandes Moura (2020235900)

- **Índice**

	Pág.
Introdução	3
Perguntas 1 e 2	3
Pergunta 3.1 e 3.2	6
Pergunta 3.3	7
Pergunta 3.4	8
Pergunta 3.5	9
Pergunta 4.1 e 4.2	10
Pergunta 4.3	11
Conclusão	11

- **Introdução**

O seguinte projeto foi realizado pelos alunos Leonardo Moreira Pina (nº 2019234318), Luís Filipe Neto (nº 2020215474) e João Fernandes Moura (2020235900) para a cadeira de Análise e Transformação de Dados com o objetivo de analisar um determinado conjunto de dados fornecido constituído por sinais recolhidos de acelerómetros de smartphones de a modo a tentar classificar e identificar o tipo de atividade realizada dentro de 12 tipos de movimento diferentes que se enquadram dentro de 3 categorias diferentes, estáticas , de transição e dinâmicas.

Este dataset recolhido para cada experiência contém informação dos 3 eixos espaciais (X, Y e Z) e uma frequência aproximada de 50 Hz, foi tratado na linguagem de programação MATLAB uma vez que a disciplina foi lecionada na mesma linguagem.

Neste relatório vamos abordar todas as perguntas enunciado fornecido bem como explicar de uma forma breve e sucinta a teoria por detrás de cada uma das respostas.

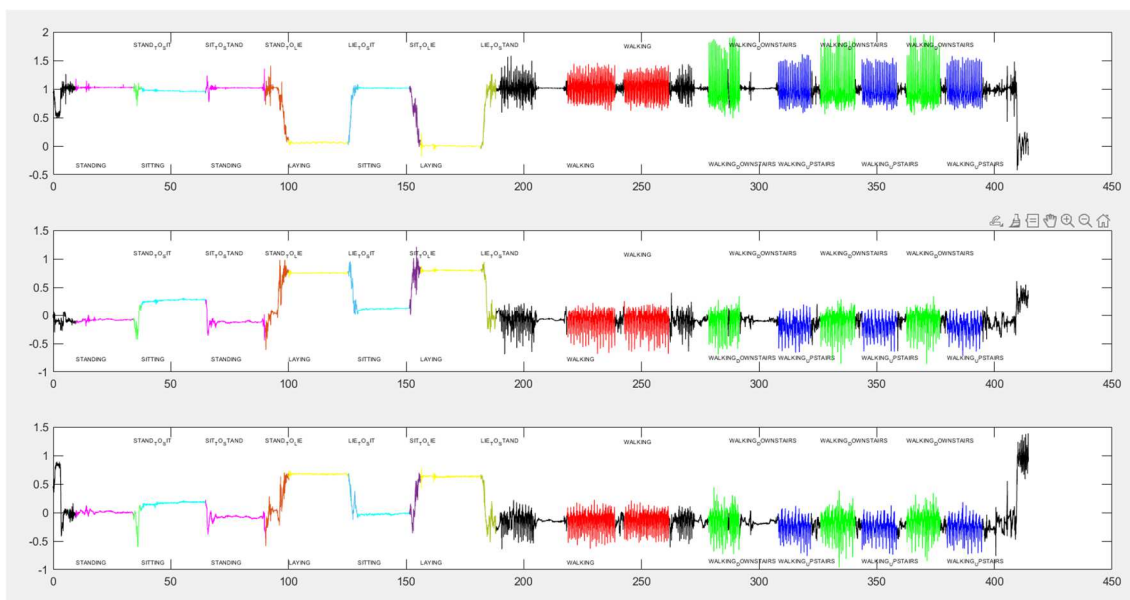
Como a turma na qual nos encontramos inscrito é a turma PL6 iremos analisar os ficheiros de input, de acc_exp42_user21.txt a acc_exp49_user24.txt

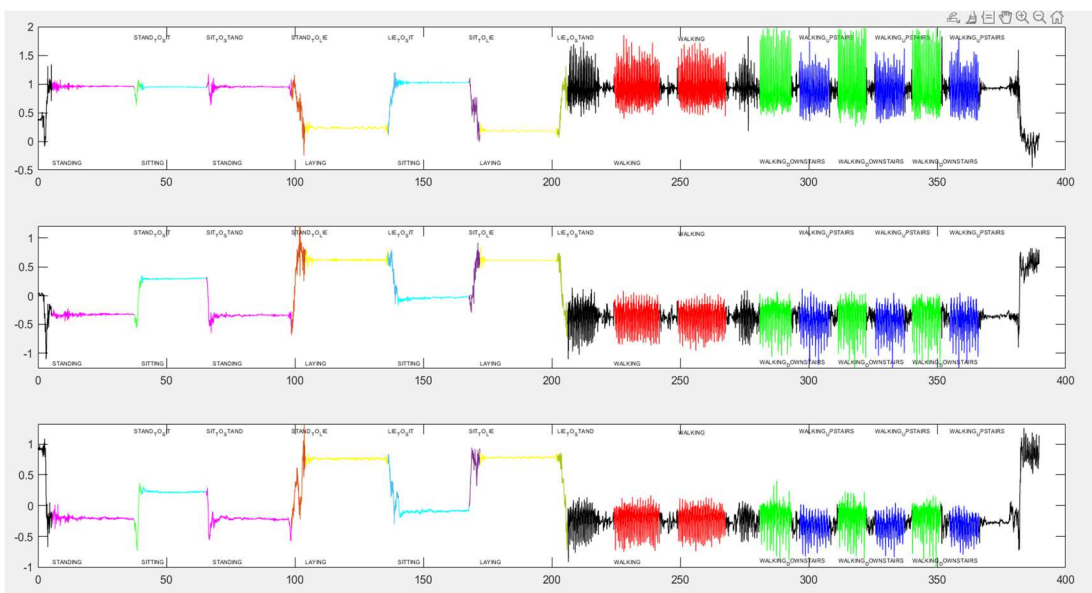
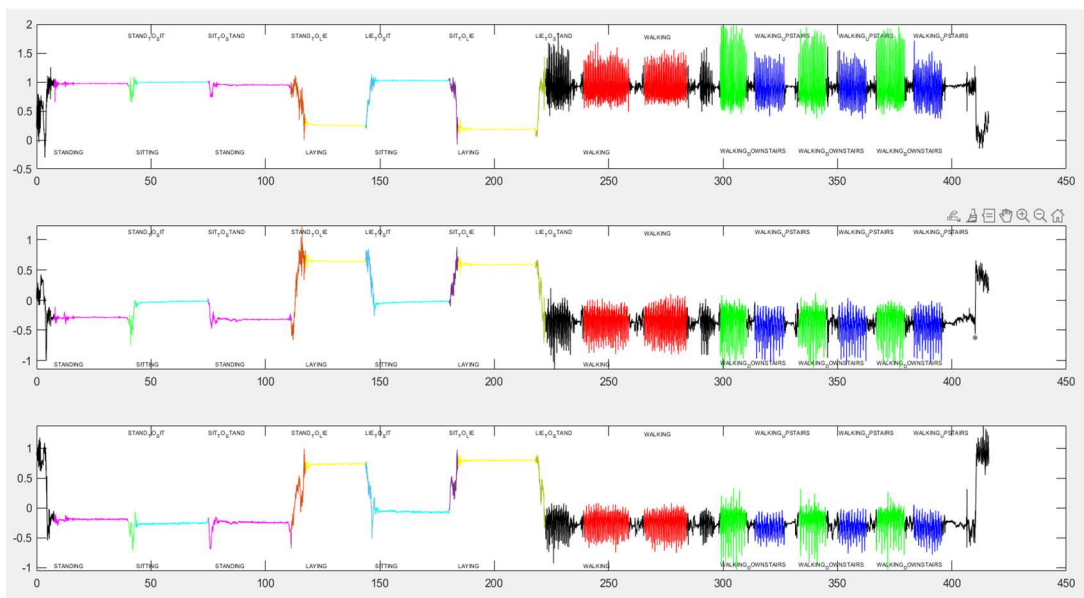
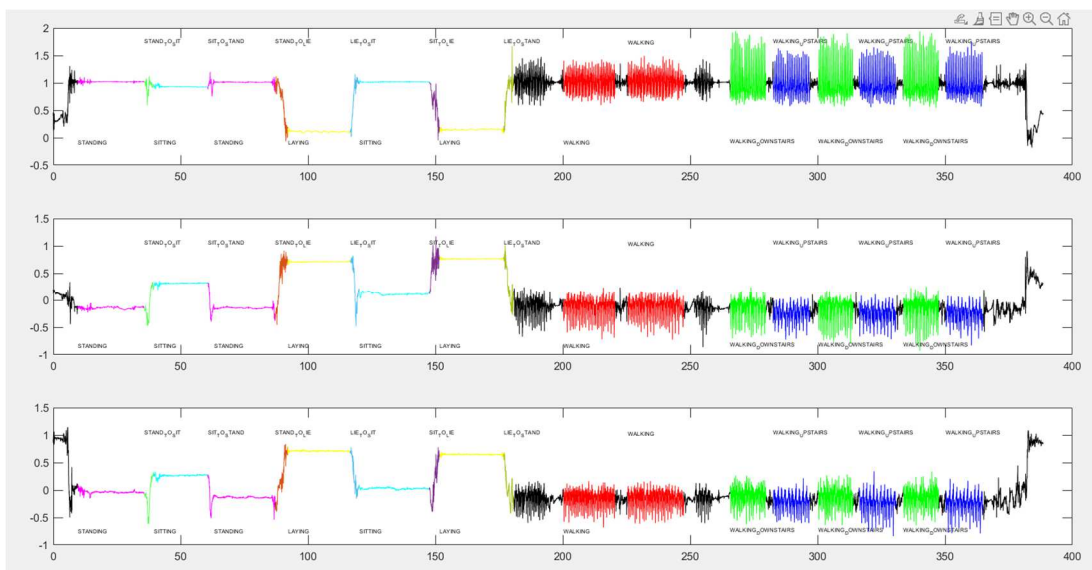
Posto isto, é ainda de referir que a linguagem utilizada neste trabalho para responder a todas as perguntas postas foi o MATLAB.

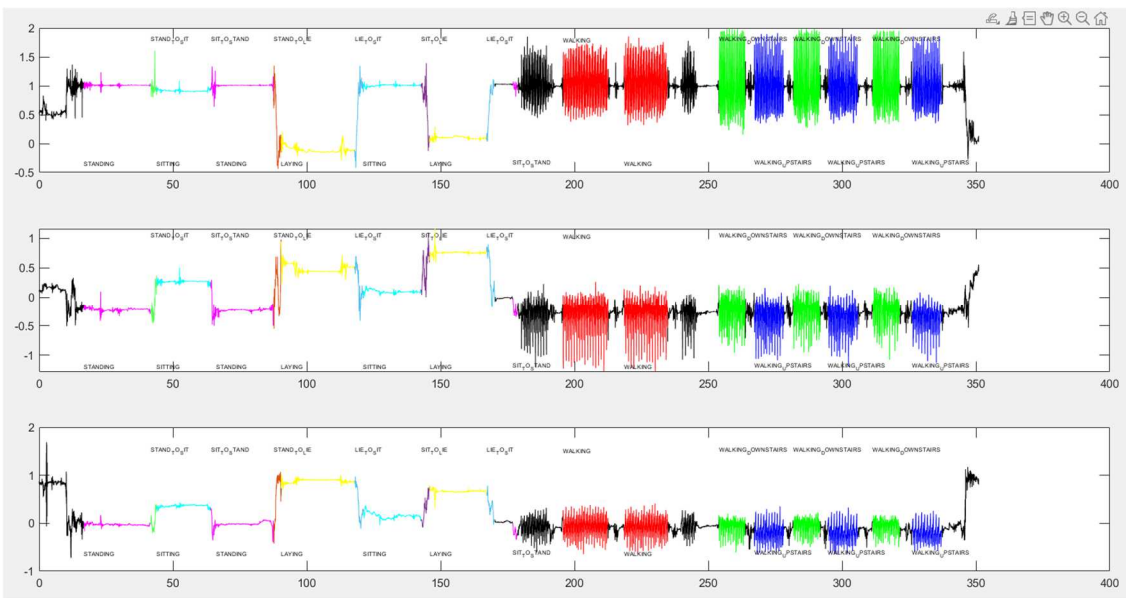
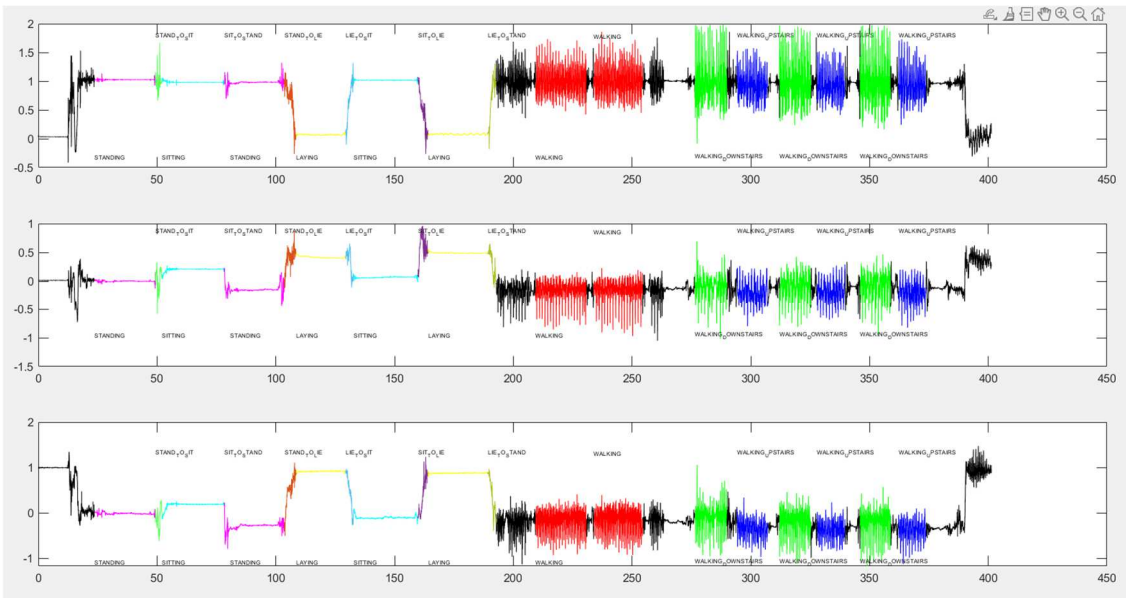
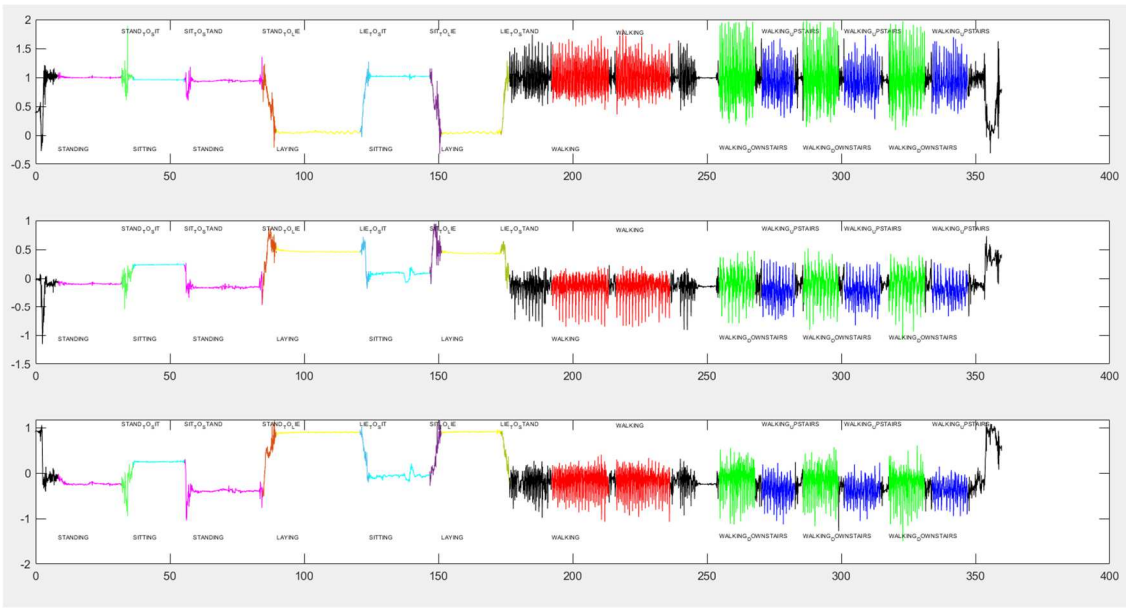
- **Pergunta 1 e 2**

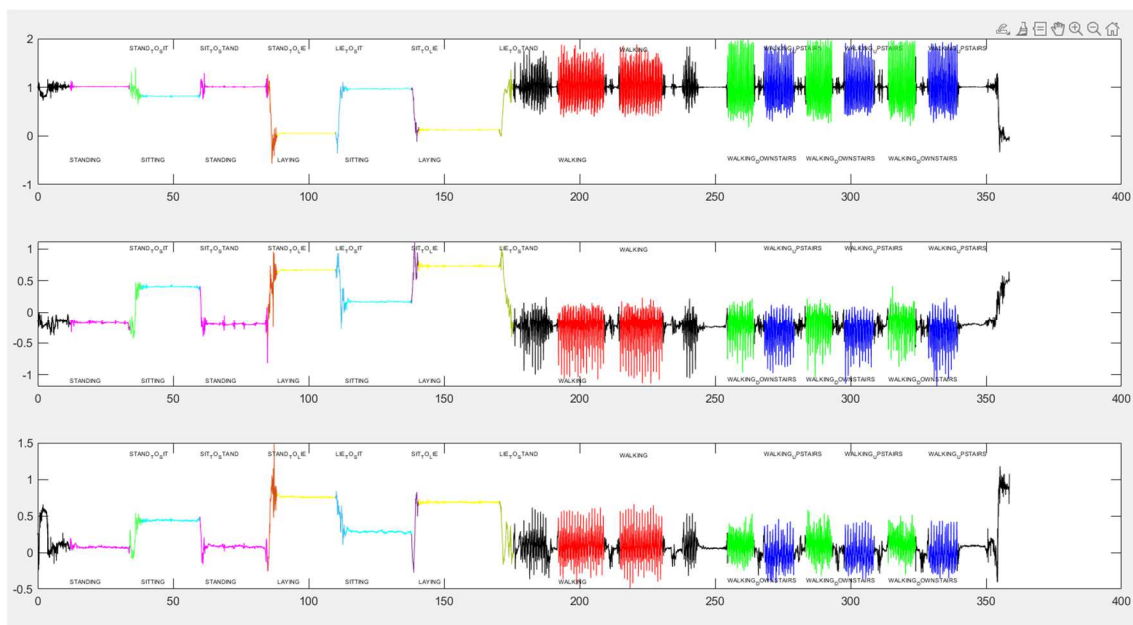
Assim, para isto, tendo já selecionado os ficheiros de input, fez-se o load dos ficheiros das experiências e do ficheiro labels.txt composto por 5 colunas de dados em que a 1ª coluna representa o número da experiência, a 2ª coluna o número do utilizador, a 3ª coluna o número da atividade e a 4ª e 5ª colunas o início e fim da atividade. Para isto utilizou-se a função load que lê a matriz e as células do ficheiro labels.txt.

Tendo todos os ficheiros lidos, procedeu-se à representação gráfica dos mesmos obtendo para cada uma das experiências três gráficos, correspondendo a cada uma das coordenadas dadas (x, y e z), em que numa primeira fase apenas estavam sem nenhuma distinção dos tipos de movimentos, sendo depois sobrepostos com a legenda de todos os movimentos e com cores distintas para diferenciar cada atividade tal como se encontra representado na figura abaixo.









- **Pergunta 3.1 e 3.2**

Na pergunta 3, foi pedido para calcular a DFT (Transformada de Fourier Discreta) do segmento do sinal de cada atividade, isto para todos os utilizadores, assim como identificar as características principais de cada uma das atividades existentes.

Assim, deve-se ainda indicar quais as características espectrais mais relevantes para cada atividade, ou seja, as frequências e amplitudes mais relevantes. Também é necessário o cálculo do número de picos, da magnitude máxima e do número de passos médio por utilizador.

Assim sendo, na realização destes exercícios foram criados 3 vetores, um em que serão guardados os valores mais relevantes da amplitude, um em que se guardam os valores mais relevantes da frequência e um para guardar os dados de todas as experiências.

De seguida, percorreu-se o vetor das experiências e para cada atividade dentro de cada experiência obteve-se os labels a partir da função obterLabels anteriormente referida e calculou-se a DFT para essa experiência, obtendo-se então as amplitudes e frequências mais relevantes.

Foi também retirada a tendência dos gráficos, já que esta mostra o comportamento da variável retratada pela série temporal a longo prazo.

É importante ainda referir que ambos os valores obtidos das frequências mais relevantes e das amplitudes mais relevantes foram armazenados em cells, isto para poderem ser usadas e acedidas mais facilmente no futuro.

- **Pergunta 3.3**

Atividade	Eixo X		Eixo Y		Eixo Z	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
Walking	114.8936	0	114.8936	0	114.8936	0
Walking up	103.4483	0	76.3547	0.4080	76.3547	0.6386
Walking down	98.0392	0	98.0392	0	165.0.327	0.9747

No que toca à pergunta 3, é pedido que para cada atividade dinâmica, sejam determinados o número médio de passos por minuto a nível de utilizador. Segue-se a seguir a tabela responsável por mostrar esses valores:

Assim, foram usados os valores mais relevantes da frequência já anteriormente calculados pela função DFT que devolve as frequências nos locais onde ocorrem os picos de magnitude. Desta forma, espera-se que ocorram picos nos passos.

Os valores obtidos remetem para o desvio padrão e à média de passos por minuto.

Apesar de o esperado ser que o eixo Z fosse o melhor para calcular os passos uma vez que é o mais propício a que ocorram picos seja um passo, as médias obtidas ocorreram diversas vezes repetidas devido a valores de frequência idênticos destacando-se na atividade de walking em que todos os valores calculados são iguais para todos os eixos.

• **Pergunta 3.4**

A pergunta 3.4 tinha como objetivo identificar as características que permitissem distinguir as atividades estáticas, de transição e dinâmicas.

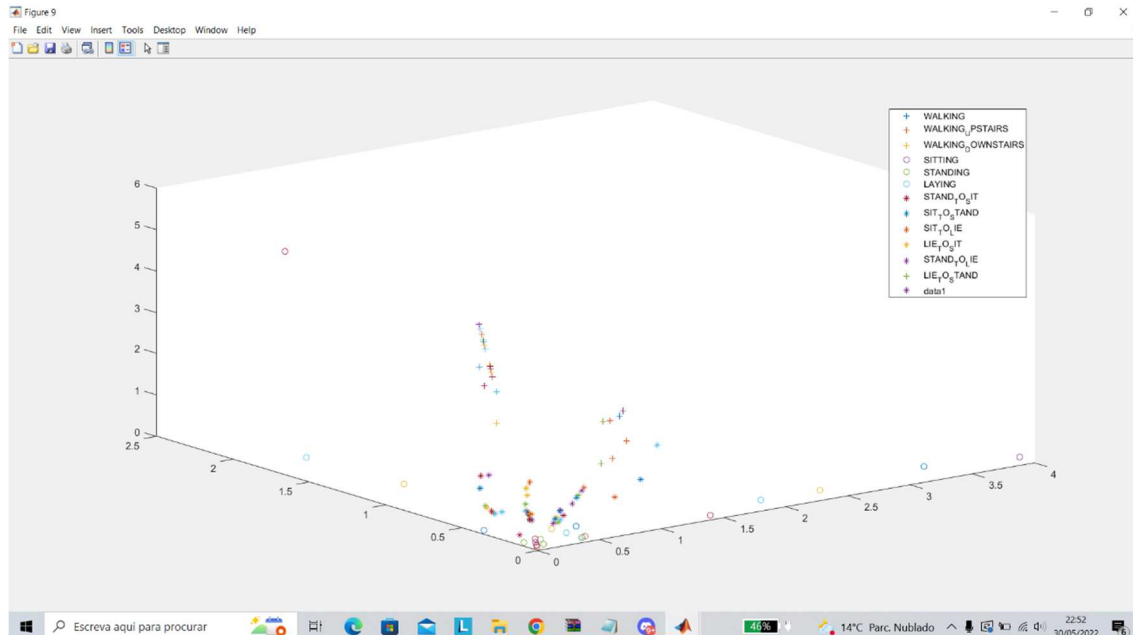
Assim sendo, uma das características espectrais principais que consideramos foram as frequências mais relevantes para podermos distinguir as diferentes atividades pelas suas categorias.

Deste modo, tal como é possível visualizar do gráfico abaixo, tal como era esperado, as atividades estáticas e de transição encontram-se mais perto do zero/origem (exceto alguns valores) enquanto as atividades dinâmicas apresentam-se com valores superiores de frequências mais relevantes.

Assim, analisando todos os resultados obtidos de acordo com a especificidade e a sensibilidade, conclui-se que para as atividades estáticas é possível distinguir os valores das frequências mais relevantes mesmo que estas

se encontrem muito perto de zero, porém, em muitos casos estes resultados não foram obtidos.

Para as atividades de transição, embora os valores de frequências se encontrem todos dentro do que era esperado, existe dificuldade em conseguir distinguir todos os tipos de atividades dentro da mesma categoria, revelando má especificidade (no gráfico as atividades aparecem quase sobrepostas).



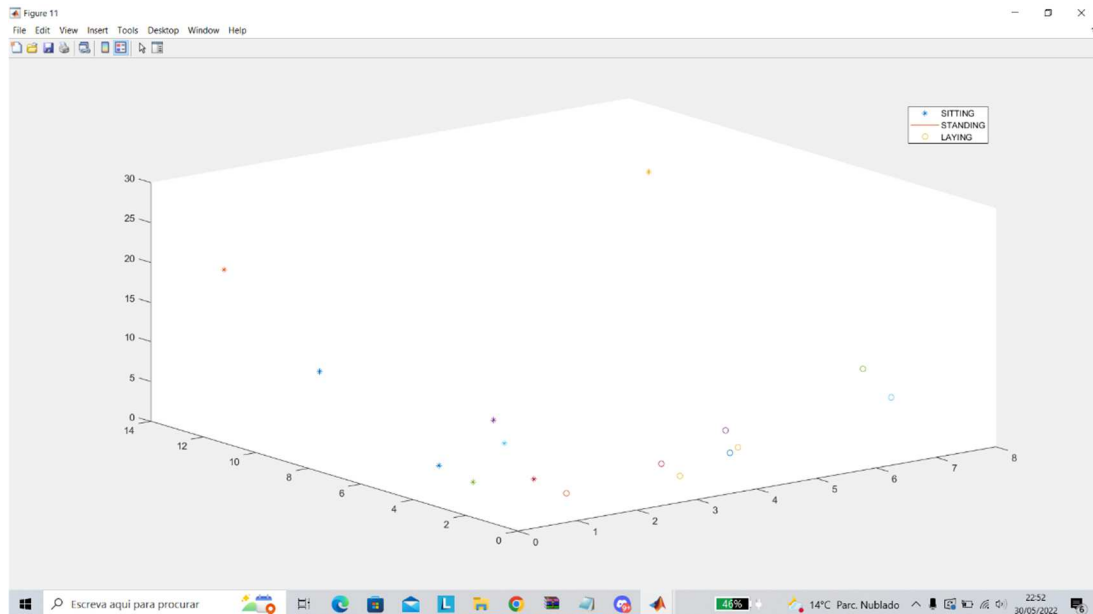
E por fim, para as atividades dinâmicas, todos os casos se encontram dentro da zona esperada e, como não tem valores sobrepostos, conclui-se que tem uma boa sensibilidade e especificidade.

- **Pergunta 3.5**

Assim, no exercício 3.5, a fechar a pergunta 3 pede-se para realizar o mesmo que foi anteriormente feito para a pergunta 3.4, mas agora para distinguir todas as atividades existentes e não apenas as suas categorias.

- **Atividades estáticas**

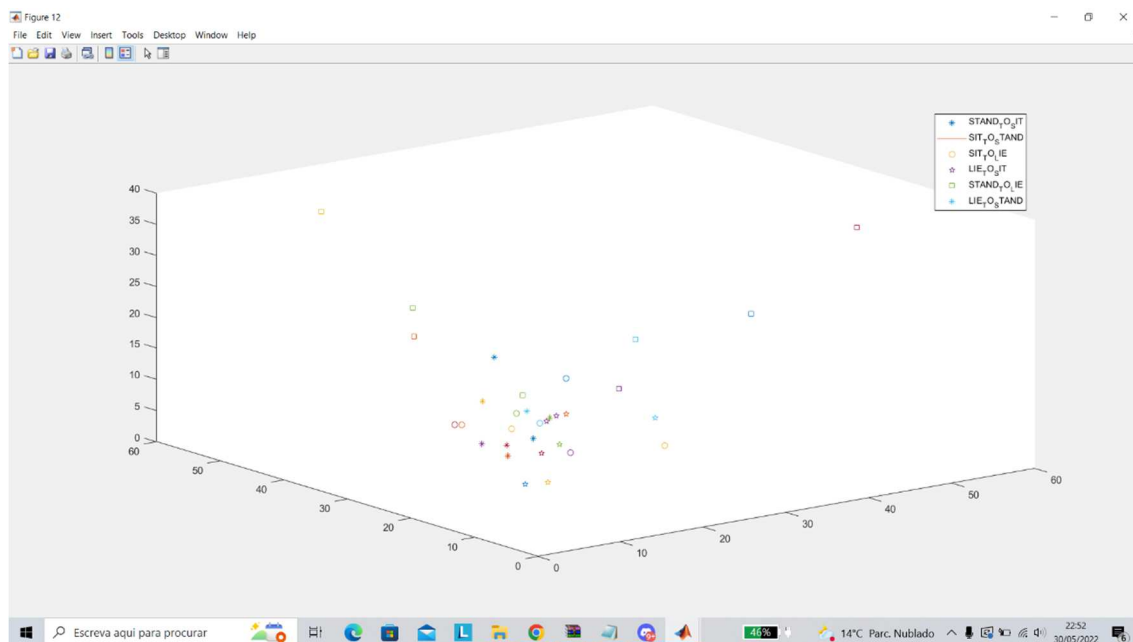
Nas atividades estáticas, as frequências mais relevantes e as magnitudes apresentam-se muito semelhantes, ou seja baixas, o que as torna muito difíceis de distinguir uma da outra, isto nos 3 ou 2 eixos.



○ Atividades de transição

Por último, este tipo de atividades, apesar de se destacarem quando comparadas com os outros tipos de atividades, dentro dos diferentes tipos nesta categoria, estas não apresentam muito em comum.

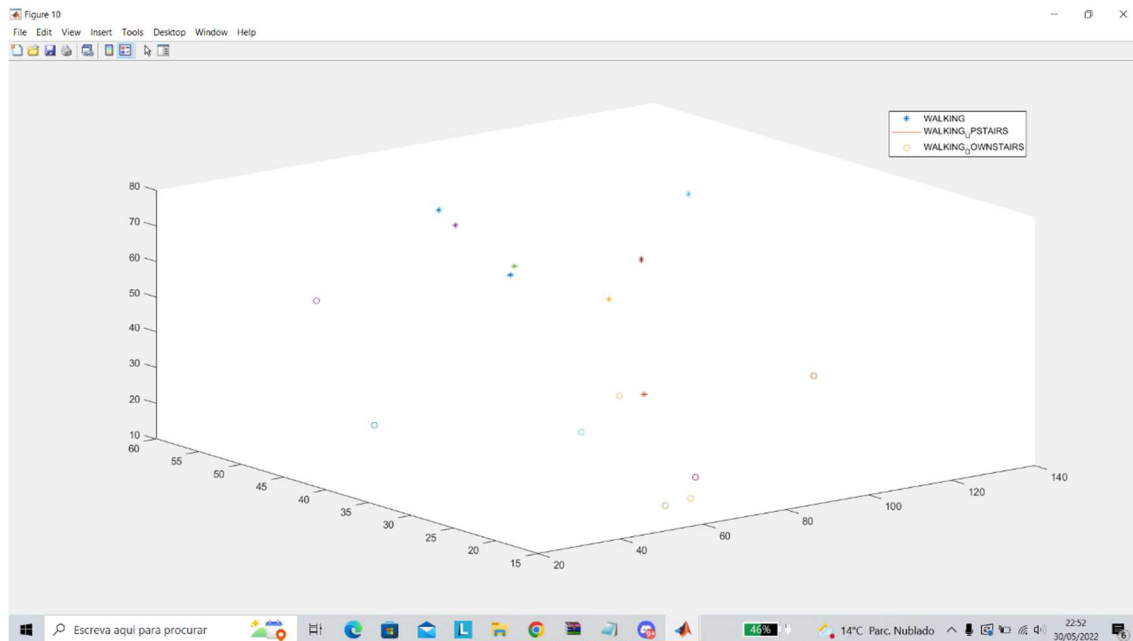
É importante referir que a atividade de transição stand to lie apresenta muitos pontos com magnitudes bastante elevadas, sendo este o padrão mais notório e comum nelas.



○ Atividades dinâmicas

Desta forma, olhando para as atividades dinâmicas, segue-se o seguinte gráfico, a atividade walking up apresenta, uma magnitude mais baixa do que todas as outras atividades dinâmicas.

Da mesma maneira, as atividades walking e walking down são, no entanto mais difíceis de distinguir sobre os 3 eixos, sendo que deveria optar-se por algo mais a duas dimensões.



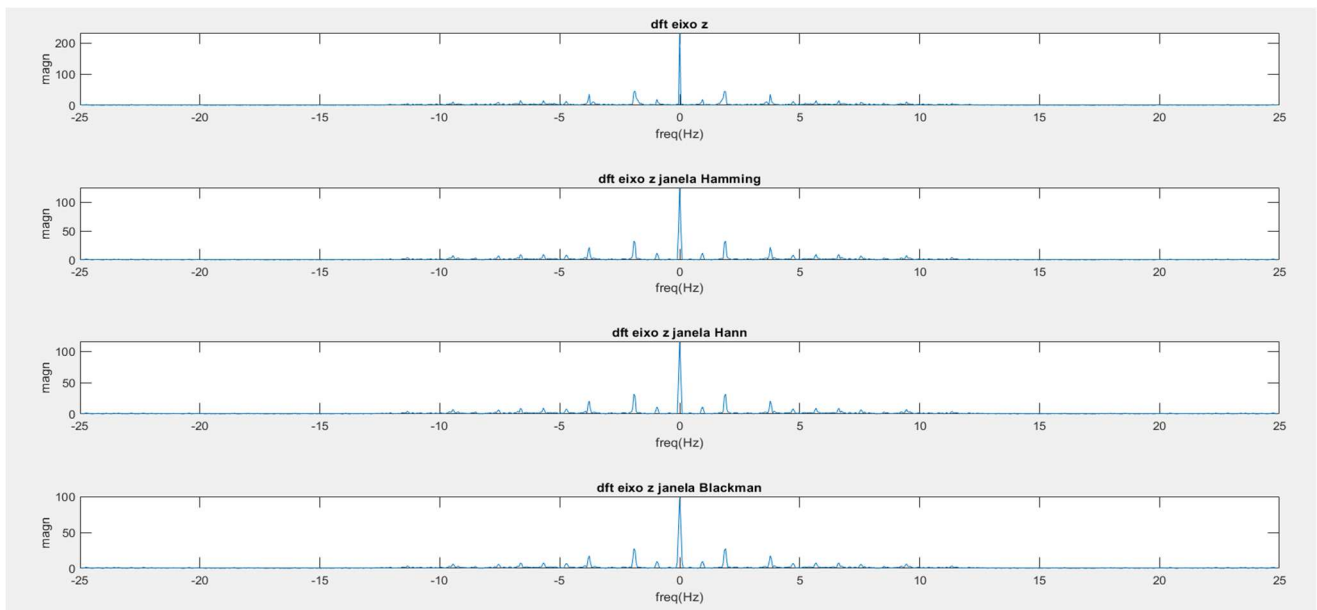
• **Pergunta 4.1 e 4.2**

A pergunta 4 refere-se à obtenção das distribuições tempo-frequência para o sinal do acelerómetro no “eixo Z” para um ficheiro de dados à nossa escolha.

Assim, como primeiramente tempos de escolher a janela a considerar na aplicação da STFT consideramos as janelas de Hamming, Blackman, Hann e retangular, procedendo-se posteriormente à sua comparação para a experiência escolhida (para este caso decidimos escolher a experiência 42) e obtiveram-se os gráficos resultantes da DFT aplicada a cada uma destas janelas.

Tendo os gráficos calculados, é possível perceber primeiramente que os valores de magnitude registados vão diminuindo de janela para a janela, isto é, são maiores para a janela retangular, menores para a janela de hamming, blackman e hann.

Concomitantemente, é fácil concluir que a janela de blackman é a mais indicada para o projeto visto que a sua aplicação ao sinal permitiu torná-lo mais suave e com menos descontinuidades, eliminando no processo picos de baixa magnitude que dificultavam a visualização e interpretação da DFT.



• Pergunta 4.3

Por fim, tendo em conta a janela escolhida na questão anterior, aplicámos a STFT ao sinal da experiência 42 usando os parâmetros corretos para visualizar de forma conveniente as diferentes atividades.

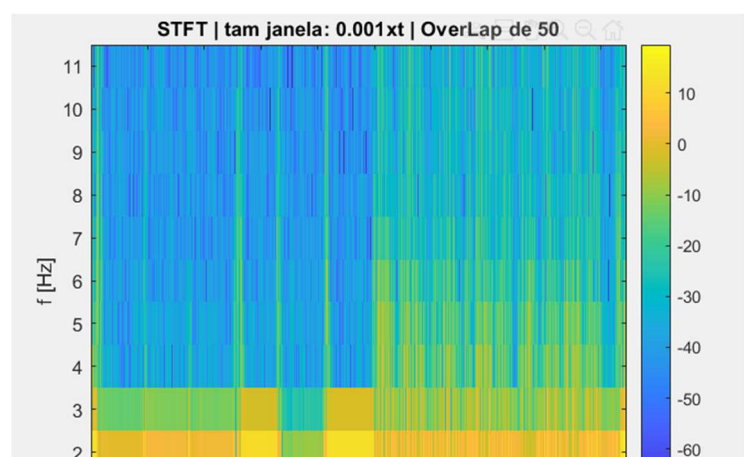
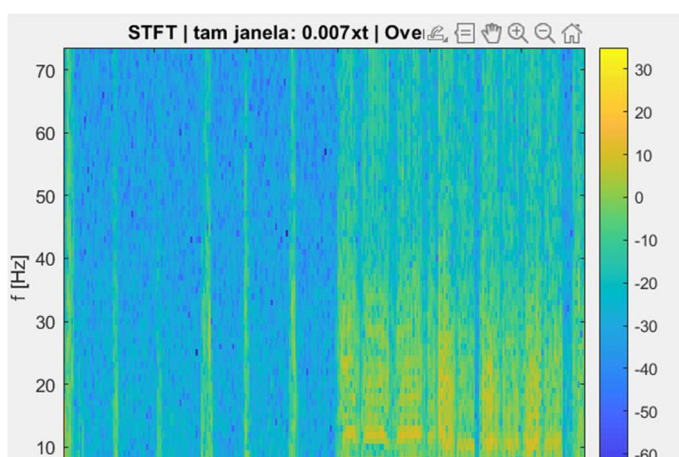
Contudo, é importante ainda referir que o tamanho da janela vai afetar a resolução na frequência e no tempo podendo constatar-se nalguns casos que quando o tamanho da janela diminui, a resolução da frequência piora.

Desta forma, é preciso encontrar o tamanho de janela indicado para que se consiga analisar o sinal tendo em conta a resolução do tempo e da frequência.

A janela mais adequada tem de ser aquela que for pequena o suficiente para apanhar as frequências mais altas das atividades dinâmicas, mas simultaneamente grande o suficiente para apanhar as frequências mais baixas das atividades estáticas.

Assim, usando tamanhos de janela de 0.007 e 0.001, obtiveram-se resultados satisfatórios visto que é possível observar em que intervalos de tempo em que a frequência se encontra e qual o seu valor.

Observando os gráficos ao detalhe também é possível distinguir quais são as atividades estáticas (valores de frequência baixo), atividades de transição (aquelas que apresentam um pico entre duas estáticas) e as atividades dinâmicas (aquelas que mantêm várias frequências mais elevadas por um maior período de tempo).



- **Conclusão**

Concluimos, dizendo que encontramos algumas dificuldades na resolução deste projeto, nomeadamente no exercício 3, visto que os valores obtidos não se mostraram os melhores, principalmente no desvio padrão no exercício 3.3. Desta forma, também a forma como os resultados foram geridos e usados ao longo do projeto não foi a mais conveniente.

No entanto, é importante referir que este trabalho nos ajudou a entender melhor a linguagem Matlab, assim como aplicações da matéria lecionada em ambientes reais.