

Análise e transformação de dados 2021/2022



Realizado por:

Rui Pedro Capelas Santos - 2020225542-PL1

Ricardo Rafael Ferreira Guegan – 2020211358-PL1

Tomás Bernardo Martins Dias 2020215701-PL1

Introdução

Para este projeto foi-nos pedido para, através da análise de um DataSet fornecido, construir uma solução computacional que nos permita identificar uma determinada atividade. O DataSet contém um conjunto de experiências realizadas por vários utilizadores em que foram obtidos os movimentos deles através de um acelerómetro colocado à cintura (smartphone).

Com estes acelerómetros, recolhemos dados dos eixos X Y e Z numa frequência de 50Hz.

Temos 12 tipos de atividades que estão divididas em 3 grupos distintos:

- *Dinâmicos*
 - *Andando (WALKING)*
 - *Subindo Escadas (WALKING_UPSTAIRS)*
 - *Descendo Escadas (WALKING_DOWNSTAIRS)*
- *Estáticos:*
 - *Sentado (SITTING)*
 - *De pé (STANDING)*
 - *Deitado (LAYING)*
- *Transição*
 - *De pé → Sentado (STAND_TO_SIT)*
 - *Sentado → De pé (SIT_TO_STAND)*
 - *Sentado → Deitado (SIT_TO_LIE)*
 - *Deitado → Sentado (LIE_TO_SIT)*
 - *De pé → Deitado (STAND_TO_LIE)*
 - *Deitado → De pé (LIE_TO_STAND)*

O código desenvolvido para a realização do projeto vai ser feito em Matlab com auxílio de ferramentas matemáticas como a transformada de Fourier discreta (DTF) e a transformada de Fourier em janelas (STFT) usando o algoritmo Fast Fourier Transform(FFT).

Guião do projeto:

1. Obter os sinais dos acelerómetros nos 3 eixos (X, Y e Z) relativos à sua turma PL e desenvolver o código necessário para os importar.

Como mencionado no enunciado, os ficheiros a considerar estão relacionados com o número da nossa PL (PL1) que no caso são os seguintes:

- acc_exp01_user01.txt
- acc_exp02_user01.txt
- acc_exp03_user02.txt
- acc_exp04_user02.txt
- acc_exp05_user03.txt
- acc_exp06_user03.txt
- acc_exp07_user04.txt
- acc_exp08_user04.txt

Para além destes ficheiros usamos também o ficheiro “labels.txt” que contém as informações necessárias para identificar a atividade a decorrer num dado instante numa dada experiência. Cada linha deste ficheiro está estruturada pela seguinte ordem:

- Número da experiência;
- Número do utilizador;
- Número da atividade;
- Início da atividade;
- Fim da atividade

Existe ainda o ficheiro “activity_labels.txt” que nos atribui um número a cada atividade.

2. Representar graficamente os sinais importados para cada eixo (X, Y e Z), identificando adequadamente a atividade a que cada segmento do sinal corresponde. Considerar o exemplo representado na figura 2.

Após uma análise dos dados fornecidos observamos que os ficheiros com os resultados dos acelerómetros estão divididos em três colunas, correspondentes aos eixos X, Y, Z e o número de linhas será igual ao número de amostragens (N).

O tempo de tempo entre amostragens é $T_s = 1/f_s = 0.02s$.

Para dar plot às figuras recorreremos a um ciclo while a percorrer os ficheiros a analisar, importamos os dados desse ficheiro recorrendo à função `importdata()` e vamos importar do ficheiro `labels.txt` os dados relativos à experiência que estamos a considerar.

De seguida vamos a ficheiro labels ver o próximo intervalo de tempo a analisar e o número da atividade respetiva que depois vai servir de índice para encontrar o nome da atividade no vetor atividade:

```
atividade = ["W", "WU", "WD", "S", "ST", "L", "STSit", "SitTS", "SitTL", "LTSit", "STL", "LTS"];
```

Com isto, resta-nos apenas dar o plot das informações.

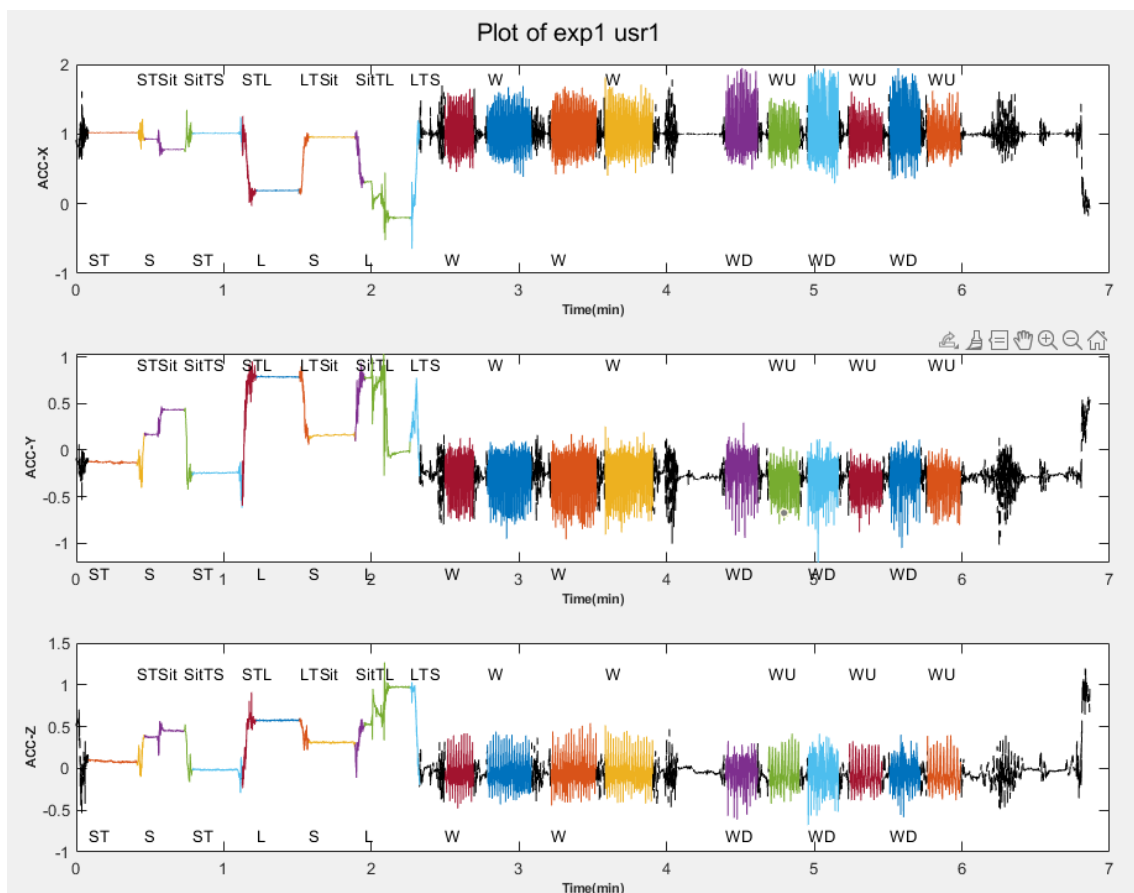


Figura 1plot da informação do ficheiro acc_exp01_user01

3. Pretende-se calcular a DFT do segmento do sinal associado a cada atividade, para os vários utilizadores, e identificar as características principais de cada atividade.

3.1. Calcular a DFT de cada segmento do sinal para as várias atividades e para os vários utilizadores. Organizar os resultados por atividade.

Para este exercício construímos a função `dft` que faz o cálculo dos coeficientes da DTF com o uso do FFT e dá plot dos mesmos para cada atividade da experiência, nos três eixos.

O que fazemos é um ciclo com o número de atividades e procuramos no ficheiro `labels` as ocorrências dessa atividade na experiência. Se houver mais que uma agrupamos. Esses coeficientes vão ser guardados num `cellarray` em que para aceder aos dados basta simplesmente `user` o `cell2mat(coefi{A,1, B})` onde `A` corresponde ao número da atividade e `B` ao acelerómetro (1=X, 2=Y, 3=Z). Após o cálculo dos coeficientes resta nos apenas dar plot a eles. Cada figura vai ter então o plot dos 3 eixos em que cada um deles tem 12 gráficos, pela ordem de cada uma das atividades.

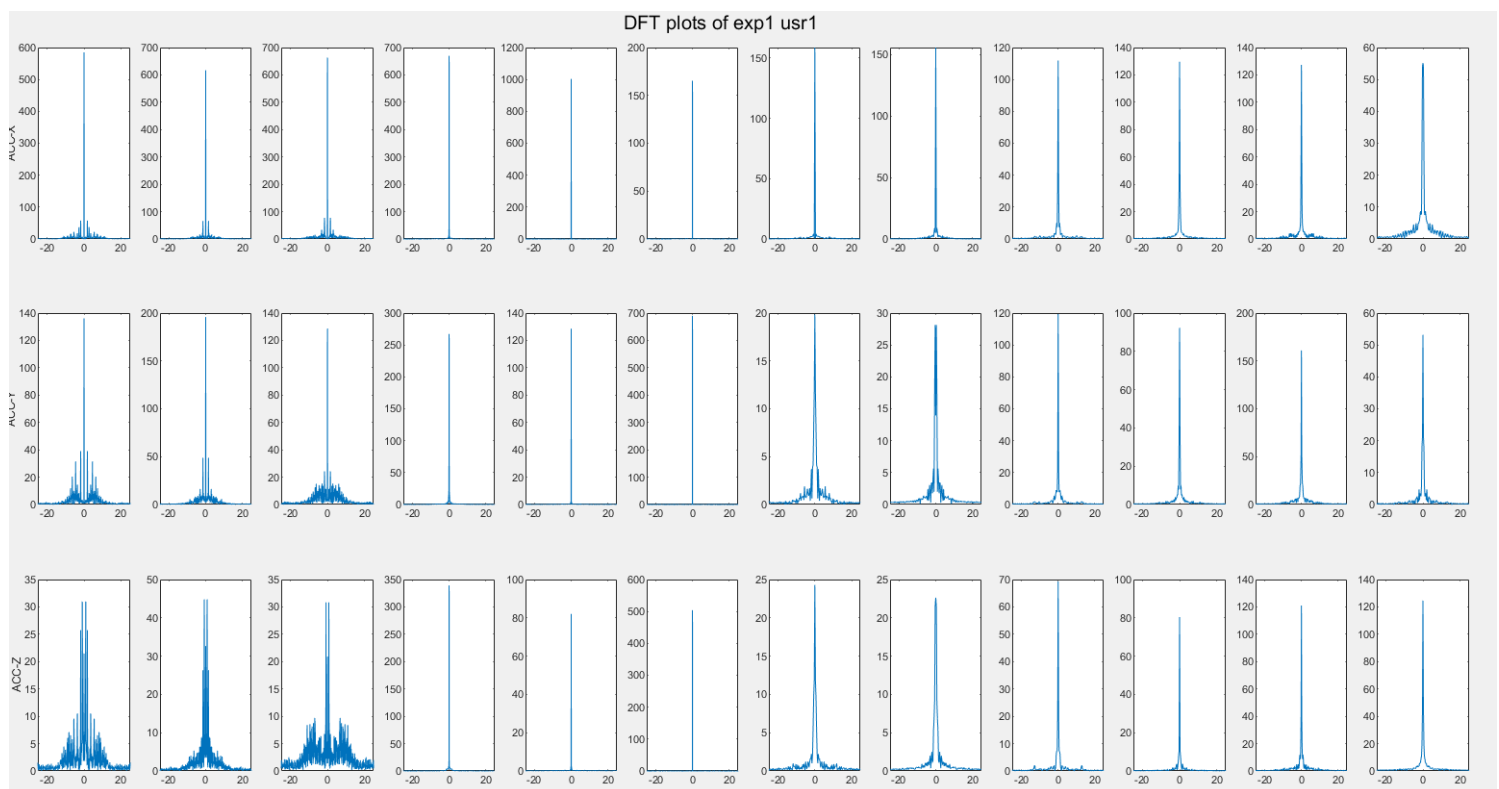


Figura 2 plot dos coeficientes resultantes da DTF para o ficheiro `acc_exp01_user01`

3.2. Identificar as características mais relevantes, nomeadamente espectrais, para cada atividade, analisando e apresentando os resultados obtidos por atividade.

Consideramos que os aspetos principais a ter em conta são o número de picos de magnitude porque em teoria, as atividades dinâmicas irão apresentar mais picos então permite distinguir atividades dinâmicas de estáticas. Para além disso, a própria magnitude poderá ser uma boa base para distinguir as diferentes atividades.

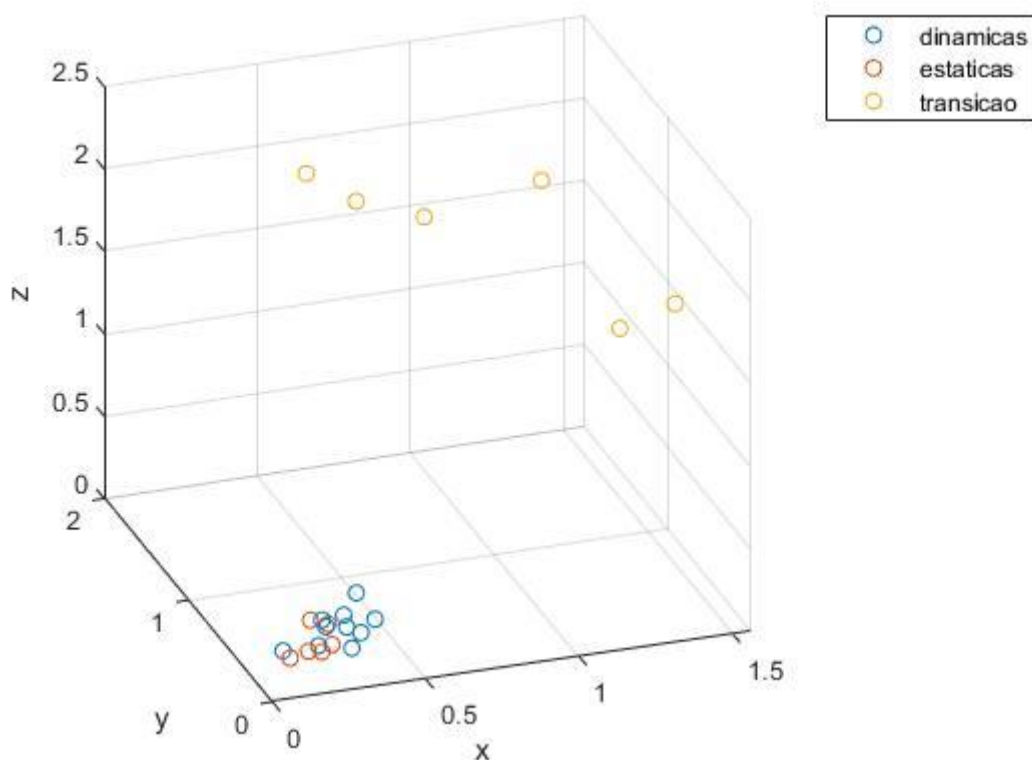
3.3. Para cada atividade dinâmica, identificar estatisticamente o número de passos por minuto para cada utilizador. Criar uma tabela com os resultados obtidos e calcular o valor médio e o desvio padrão por atividade

Para resolver esta questão, foi criada a função spm. Aqui é calculada a dft das três atividades dinâmicas (Walk, Walk Upstairs e Walk Downstairs) para cada um dos utilizadores. Depois, de modo a obter as o número de passos por minuto calculamos os picos usando findpeaks, usando MinPeakHeight de 40% do valor máximo. Depois calculamos a frequência usamos o primeiro pico mais significativo, o que nos vai dar o número de passos por segundo. Para obtermos o número de passos por minuto basta apenas multiplicar por 60. Depois resta-nos apenas calcular a média e o desvio padrão.

| User/Atividade | Walk | | | Walk Upstairs | | | Walk Downstairs | | |
|----------------|------------|------------|------------|---------------|------------|------------|-----------------|------------|------------|
| | x | y | z | x | y | z | x | y | z |
| 1 | 102,916 | 102,916 | 102,916 | 92,71 | 92,71 | 92,71 | 86,8902 | 59,4512 | 59,4512 |
| 2 | 106,7416 | 106,7416 | 106,7416 | 90,9763 | 55,4734 | 55,4734 | 82,9694 | 93,8865 | 93,8865 |
| 3 | 97,7411 | 97,7411 | 97,7411 | 81,9793 | 48,7445 | 48,7445 | 95,6376 | 65,4362 | 65,4362 |
| 4 | 115,2838 | 115,2838 | 115,2838 | 87,8981 | 26,7516 | 26,7516 | 99,2218 | 35,0195 | 35,0195 |
| Média | 104,8288 | 104,8288 | 104,8288 | 89,4372 | 52,10895 | 52,10895 | 91,2639 | 62,4437 | 62,4437 |
| Desvio Padrão | 7,39426459 | 7,39426459 | 7,39426459 | 4,71488368 | 27,4225388 | 27,4225388 | 7,535628289 | 24,1840805 | 24,1840805 |

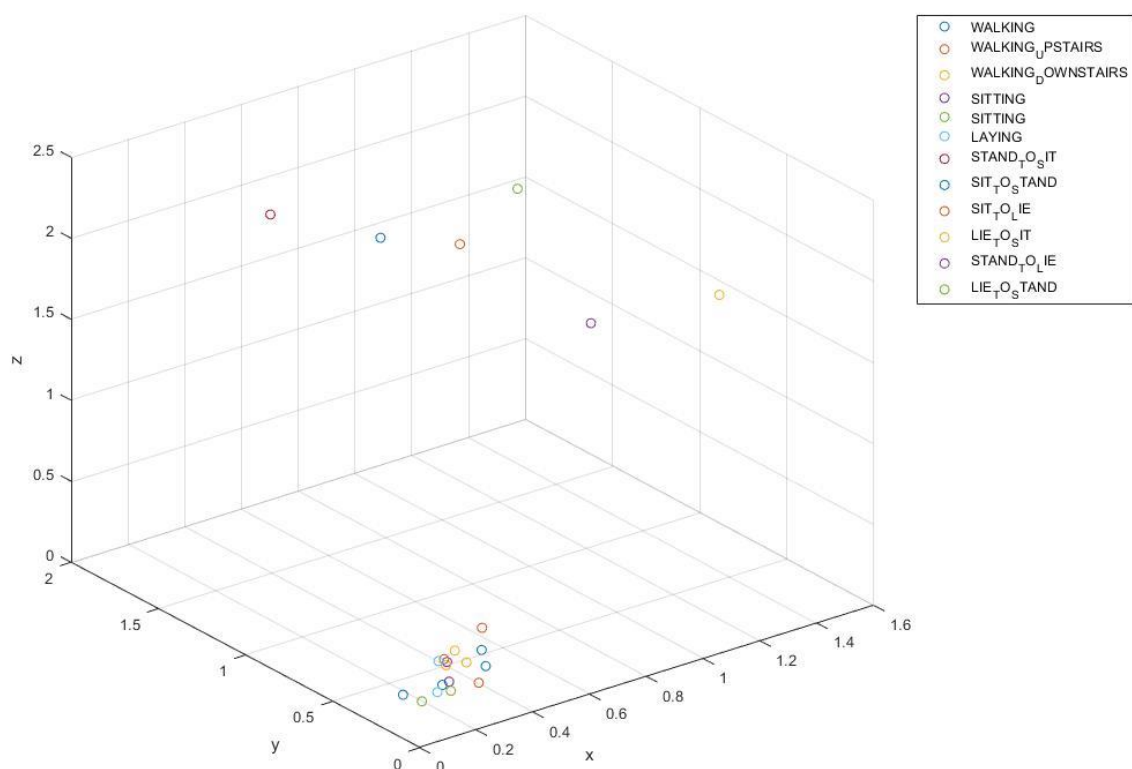
3.4. Identificar as características, por exemplo espectrais, que permitam distinguir as atividades estáticas, de transição e dinâmicas. Apresentar os resultados graficamente. Avaliar os resultados obtidos em termos de sensibilidade e especificidade.

Para resolver este exercício, calculamos o primeiro pico para as várias atividades da experiência 1, para tentar mostrar as diferenças entre as atividades estáticas das de transição e das dinâmicas. Após a observação do gráfico verificamos que conseguimos distinguir facilmente as atividades de transição das estáticas e das dinâmicas, mas entre as estáticas e as dinâmicas não conseguimos distinguir muito bem. Para além disso, os resultados não estão bem de acordo com o esperado. Realmente era esperado que as atividades estáticas estivessem mais próximas de 0 em todos os eixos, mas as dinâmicas não, e por isso não conseguimos falar de sensibilidade e especificidade.



3.5. Identificar as características, por exemplo espectrais, que permitam distinguir cada tipo de atividade. Apresentar os resultados graficamente. Avaliar os resultados obtidos em termos de sensibilidade e especificidade.

Com um processo semelhante ao usado na questão anterior, obtivemos o gráfico abaixo representado, onde tal como na questão anterior não concordo com a posição das atividades dinâmicas, e entre as dinâmicas e as estáticas também não conseguimos diferenciá-las.



Queremos voltar a salientar que para nós o expectável seria que as atividades dinâmicas tivessem os valores mais altos em todos os eixos, e que as estáticas tivessem os valores mais baixos em todos os eixos, e possivelmente o facto de não estarmos a verificar isto será um erro na nossa função.

4. Obter computacionalmente as distribuições tempo-frequência para o sinal do acelerómetro no “eixo Z” para um ficheiro de dados à sua escolha. Usar a Short-Time Fourier Transform (STFT).

O resultado apresentado nas alíneas seguintes tem como ficheiro de origem “acc_exp01_user01.txt”. A atividade dinâmica selecionada foi a “Walking”.

4.1. Escolher, justificadamente, a janela a considerar na aplicação da STFT. Para isso, seleccionar uma atividade dinâmica e usar diferentes tipos de janela para segmentar o sinal associado a essa atividade. Calcular a DFT do segmento com as diferentes janelas e comparar os resultados obtidos, procurando evidenciar o efeito das diferentes janelas.

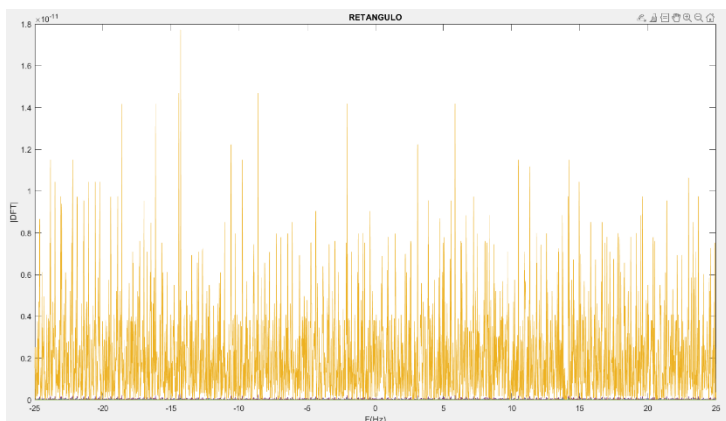


Figura 3- Janela Retangular

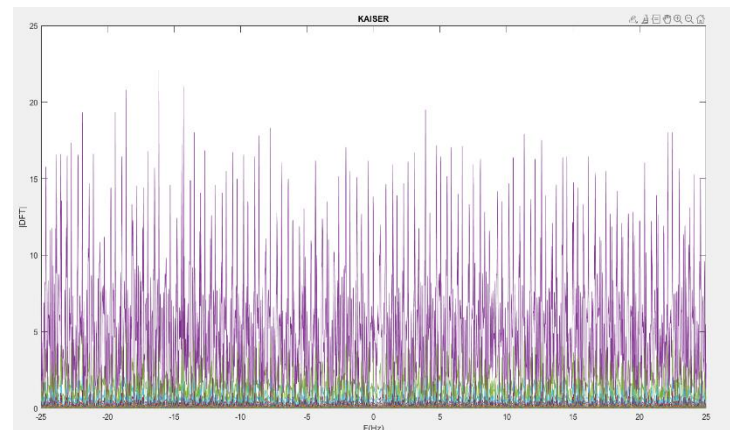


Figura 4- Janela de Kaiser

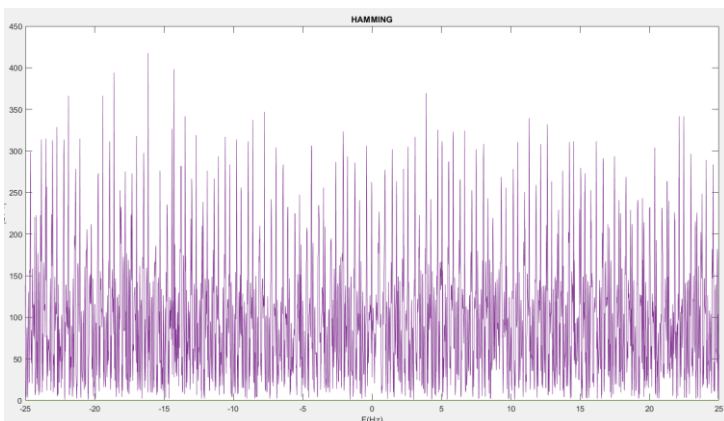


Figura 5- Janela Hamming

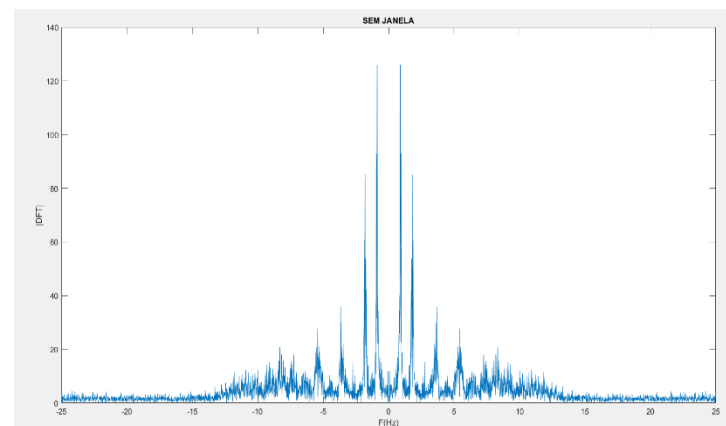


Figura 6- Sem janela

O nosso principal objetivo ao escolher a janela é que esta molde os dados de modo que os padrões sejam mais facilmente detetáveis. Das janelas analisadas concluímos que aquelas que conseguem melhor reduzir o ruído, o que é importante em atividades dinâmicas, são a janela de Hamming e a retangular. Porém a janela de Hamming consegue evidenciar melhor os padrões do movimento escolhido, neste caso “Walking”.

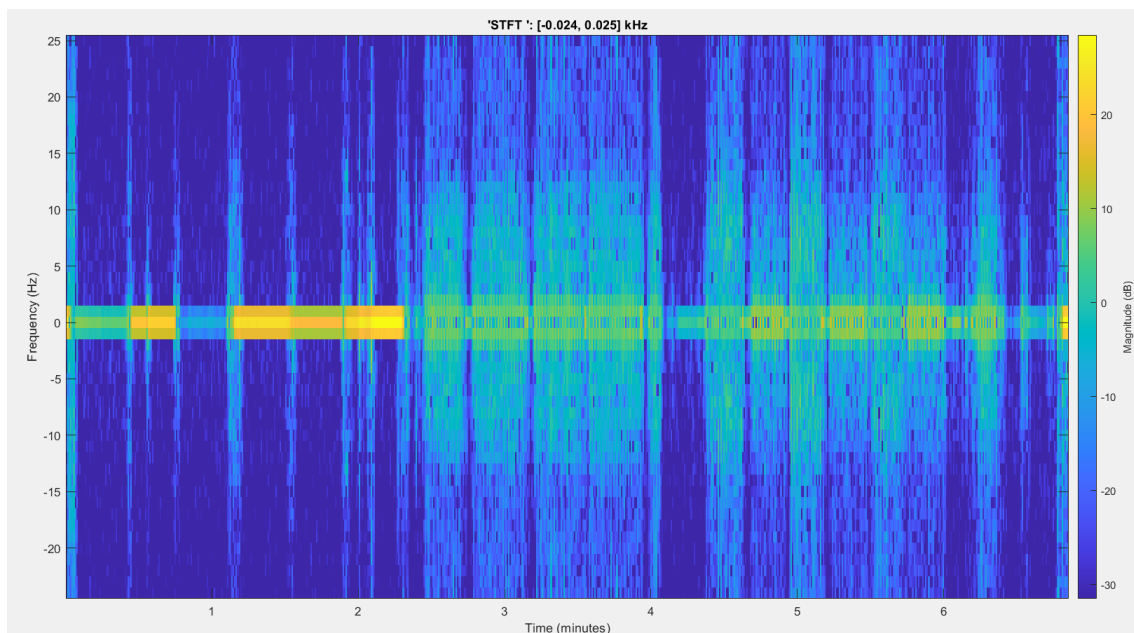
4.2. Desenvolver a sua própria função para o cálculo da STFT.

```
opts = {'Window', hamming(50, 'periodic'), 'OverlapLength', 20, 'FrequencyRange', "centered"};  
[~,f] = stft(dados_atividade,fs,opts{:});
```

Após a escolha do tipo de janela, para realizar o cálculo da STFT ainda é necessário ter em mente outros parâmetros, estes são: tamanho da janela, ou seja, número de amostras que vamos considerar em cada aplicação da DFT e o overlap que iremos ter, ou seja, o número de amostras que se sobrepõem com os elementos da DFT anterior.

Para escolher o tamanho da janela decidimos que este seria ligeiramente maior que o tempo necessário para realizar um passo. No que diz ao overlap como o passo seguinte está relacionado com o movimento anterior achamos que importante que o overlap seja elevado para se conseguir evidenciar o mesmo.

4.3. Aplicar a STFT ao sinal com a janela deslizante selecionada, usando os parâmetros corretos para visualizar de forma conveniente as diferentes atividades, não descorando a visualização apropriada da “assinatura” espectral de cada uma delas. Avaliar os resultados obtidos em termos de sensibilidade e especificidade.



Como podemos observar no gráfico conseguimos facilmente distinguir atividades estáticas e dinâmicas. As dinâmicas apresentam maior dispersão da sua frequência ao contrário das estáticas que mantem a mesma frequência ao longo da sua duração.

Para além de ser possível distinguir as atividades conseguimos também visualizar as transições entre as mesmas. Sendo estas perceptíveis pelas mudanças abruptas no campo de frequências.

Com uma observação mais atenta do gráfico conseguimos também distinguir as assinaturas de cada uma das atividades, por exemplo nas atividades dinâmicas aquela que tem maior dispersão de frequência é o “Walking” e as que tem menos dispersão “WalkingUp” e “WalkingDown”.