



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Relatório Projeto 2021/2022

Análise e transformação de Dados

Trabalho Realizado por:

Bruno Sequeira nº 2020235721

João Pinto nº 2020220907

Rodrigo Figueiredo nº 2020236687

INDICE

• Introdução.....	3
• Exercício 1	4
• Exercício 2	5
• Exercício 3.1	7
• Exercício 3.2	9
• Exercício 3.3	9
• Exercício 3.4	10
• Exercício 3.5	11
• Exercício 4.1	13
• Exercício 4.2	14
• Exercício 4.3	14
• Conclusão e Biografia	16

Introdução

A monitorização dos movimentos humanos pode ser útil para muitos propósitos e aplicações, uma possível aplicação é o desenvolvimento de soluções para “Active Assisted Living(ALL)”, direcionada para os cuidados de saúde e apoio à população sénior.

Com isso, o objetivo deste projeto é analisar os dados, nos domínios do tempo e da frequência, usando os sinais que foram recolhidos de acelerómetros de smartphones com o objetivo de identificar a atividade realizada. Para isso temos um DataSet que nos foi fornecido, este que contem um conjunto de experiências realizadas por utilizadores. Os acelerómetros dão-nos os dados dos eixos x,y,z numa frequência de 50Hz. A partir de 3 grupos temos 12 tipos de atividades que são as seguintes:

- *Dinâmicos*
 1. *Andando (WALKING)*
 2. *Subindo Escadas (WALKING_UPSTAIRS)*
 3. *Descendo Escadas (WALKING_DOWNSTAIRS)*
- *Estáticos:*
 4. *Sentado (SITTING)*
 5. *De pé (STANDING)*
 6. *Deitado (LAYING)*
- *Transição*
 7. *De pé → Sentado (STAND_TO_SIT)*
 8. *Sentado → De pé (SIT_TO_STAND)*
 9. *Sentado → Deitado (SIT_TO_LIE)*
 10. *Deitado → Sentado (LIE_TO_SIT)*
 11. *De pé → Deitado (STAND_TO_LIE)*
 12. *Deitado → De pé (LIE_TO_STAND)*

A realização do projeto será em Matlab, com auxílio de ferramentas matemáticas que aprendemos ao longo do semestre, tais como a Short-Time Fourier Transform(STFT), e a transformada de Fourier discreta(DFT). A função "main" é a PROJ.m.

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.

O DataSet deste projeto foi obtido por 30 pessoas que se submeteram a gravações com os sensores enquanto praticavam algumas atividades ao longo do dia.

Os ficheiros que iremos utilizar ao longo do nosso projeto são:

- acc_exp54_user27.txt
- acc_exp55_user27.txt
- acc_exp56_user28.txt
- acc_exp57_user28.txt
- acc_exp58_user29.txt
- acc_exp59_user29.txt
- acc_exp60_user30.txt
- acc_exp61_user30.txt
 - Pra importar essa informação, criamos uma função “leitura(nomedoficheiro)” que retorna 3 arrays de uma experiência, cada um representa um eixo(x,y,z).

Para além destes ficheiros usamos também o ficheiro “labels.txt” que tem a seguinte estrutura:

- Coluna1: Número da Experiência.
- Coluna2: Número do Utilizador.
- Coluna3: ID da atividade.
- Coluna4: Ponto de Início da atividade.
- Coluna5: Ponto de Fim da atividade.
 - Para importar essa informação, criamos uma função “leituraLabel(nomedoficheiro,númerodaexperiencia)” esta que retorna os ID’s de atividade, os tempos de início e os tempos finais acerca dessa experiência.

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.

A partir das funções acima mencionadas, que servem para importar os sinais necessários para a resolução do projeto, neste exercício é necessário representá-los graficamente identificando adequadamente a atividade.

Para isso percorremos o array de ficheiros, importamos o sinal do ficheiro a partir da função "leituraFile(nomefile)" e vamos importar do ficheiro labels.txt os dados que pertençam a essa experiência.

Chamamos uma função "Exercicio2(,,,,,)" este que tem como argumentos o número da figura, o array do tempo, eixo x, eixo y, eixo z, as cores que representam as atividades, os ID's das atividades, os tempos de início da atividade, os tempos de fim da atividade, nomes das atividades, e o nome a que queremos dar ao gráfico.

```
ficheiros = [54,27,55,27,56,28,57,28,58,29,59,29,60,30,61,30];  
coreszita = ['r','g','b','c','m','y','*','o','+','-','.', 'c'];  
nomes = ["w", "wu", "wd", "SIT", "STAND", "LAY", "SSIT", "SSTAND", "SLIE", "LSIT", "STLIE", "LSTAND"];
```

Esta função vai de intervalo a intervalo a colocar no gráfico as cores relativamente à atividade que está a ser representada.

Apresentamos agora o resultado, com o exemplo de dois ficheiros:

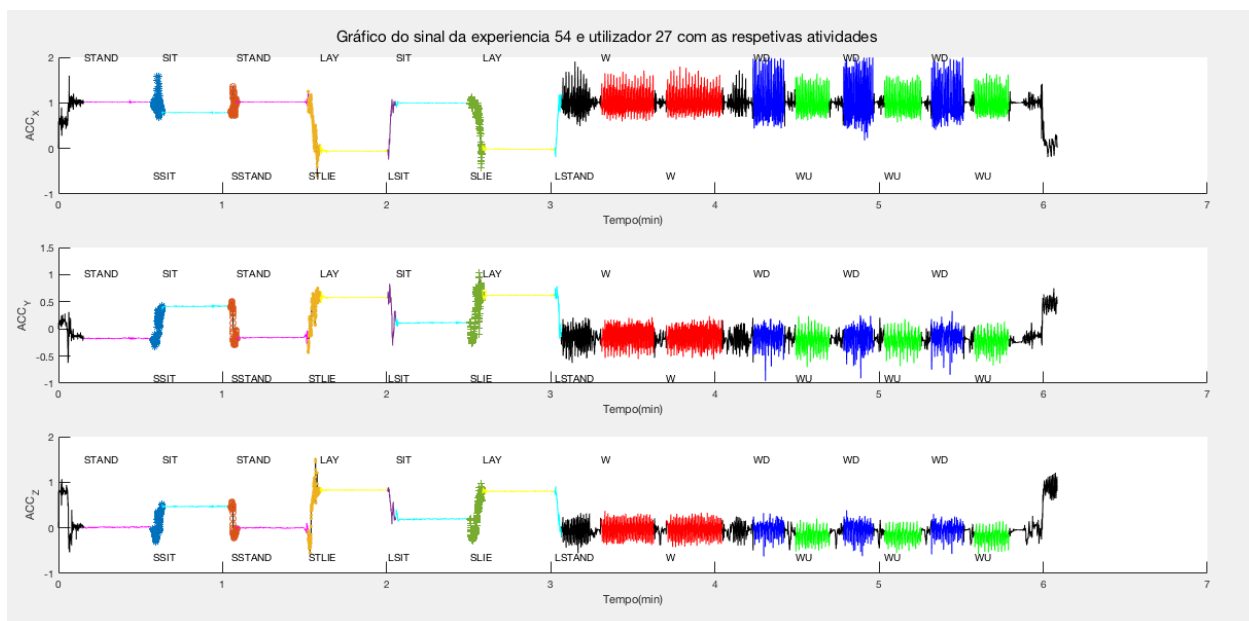


Figura nº1- Representação do ficheiro "acc_exp54_user27.txt"

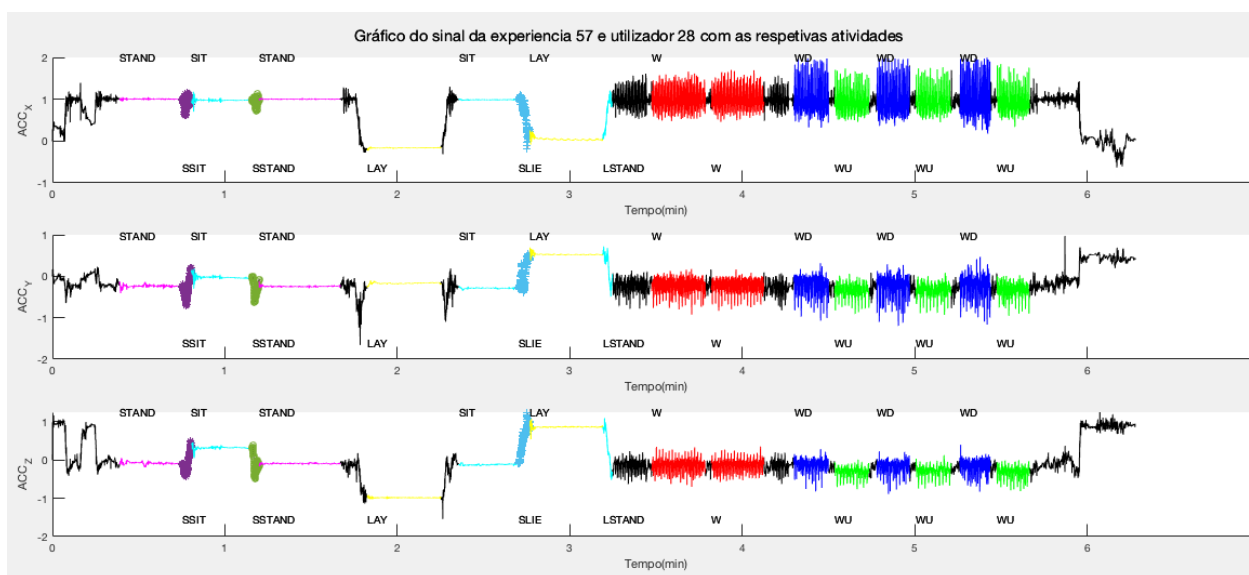


Figura nº2 -Representação do ficheiro "acc_exp57_user26.txt"

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 realização deste exercício, foi criado uma função “DFT(eixo,fs)”, que faz o cálculo dos coeficientes da DFT retornando assim a frequência e a magnitude do sinal.

Foi criado uma função “Exercicio3()” que percorre os ID’s retornados da leituraLabel e retorna os tempos de todas as ocorrências de um ID, e criou-se uma função “DFT_PLOTS()” que junta a uma figura os subplots correspondentes a um ID.

Para isso fazemos um ciclo 3 ciclos, o primeiro percorre os ficheiros tal como no exercicio1, e outro que envia para a função “Exercicio3()” o ID da atividade, para que lhe seja retornado o array dos tempos.

Juntando o assim toda a informação desse ID nesse ficheiro num array auxiliar. Finalizando é enviado para a função DFT_PLOTS() todos os eixos(x,y,z).

Apresentamos agora o resultado, com o exemplo de dois ficheiros:

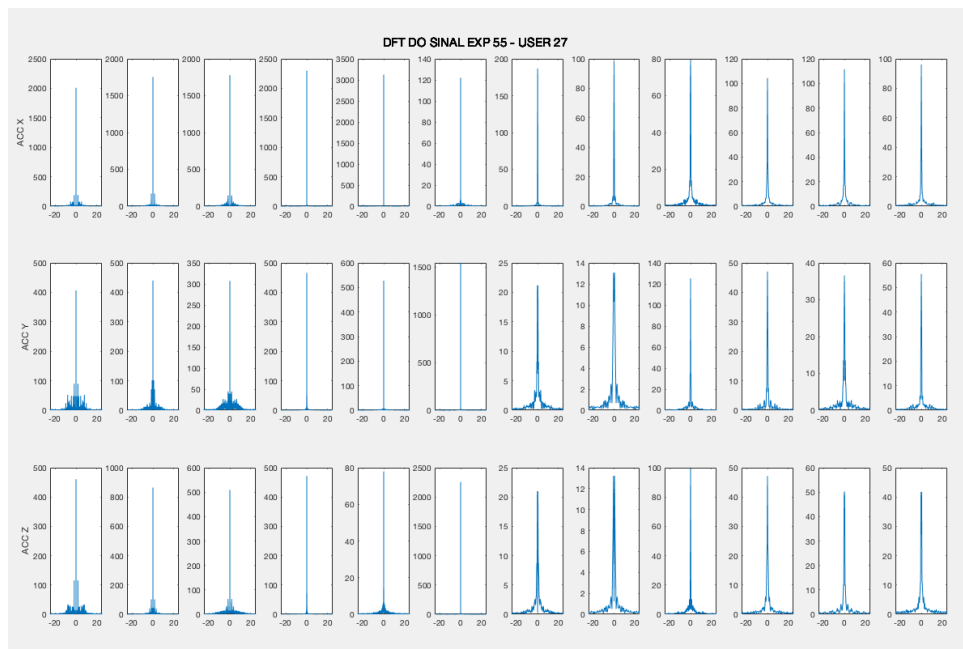


Figura nº3 Representação dos coeficientes resultantes da DFT de cada segmento do sinal para as várias atividades

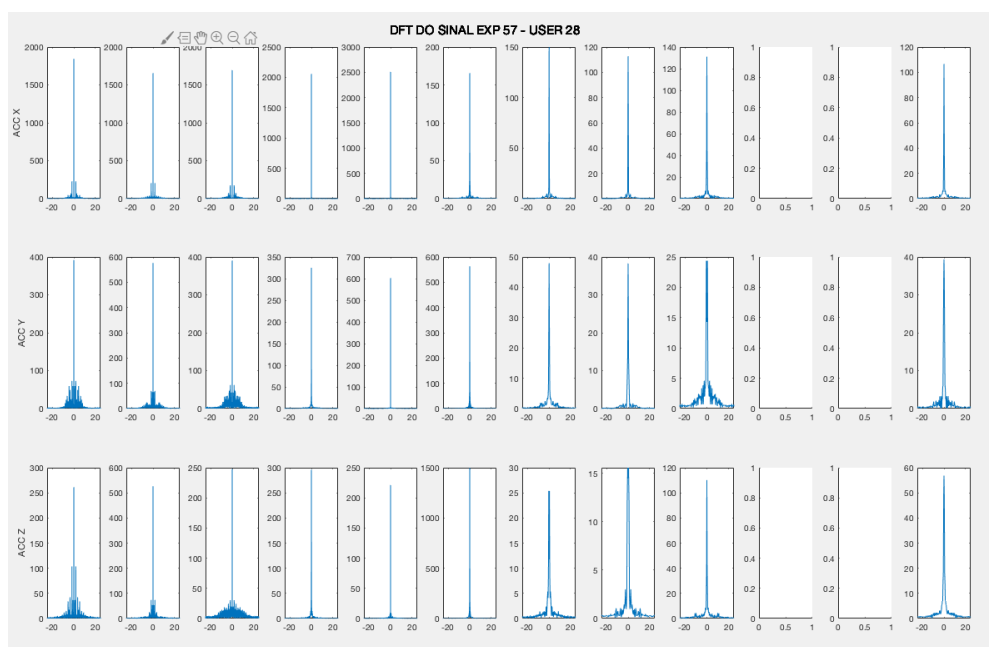


Figura nº4 Representação dos coeficientes resultantes da DFT de cada segmento do sinal para as várias atividades

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

As características mais relevantes, a ter em conta são os números de picos de magnitude, visto que nos ajudam a distinguir as diferentes atividades, estáticas, dinâmicas e transição, pois teoricamente, as dinâmicas vão apresentar um número maior de picos. Igualmente, a magnitude poderá ser um bom auxílio 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.

Foi criada uma função chamada “calculaPassos()”. Para cada utilizador foi calculado a DFT das suas atividades dinâmicas (Walking, Walking_Upstairs, Walking_Downstairs), e essa informação foi enviada para a função acima referida. Essa função de maneira a calcular o número de passos por minuto, são calculados os picos usando o “findpeaks()”, com o ‘MinPeakHeight’ de 40% do valor máximo. Após isso, utilizamos a frequência usamos o primeiro pico mais significativo, o que nos vai retornar o número de passos por segundo. Finalizando apenas é necessário multiplicar por 60 e calcular o desvio-padrão e a media.

Atividades	Eixo X	Eixo Y	Eixo Z
Walking	Média ± Desvio: 101.7451±26.3628	Média ± Desvio: 52.1785±26.3628	Média ± Desvio: 50.0497±25.0143
Walking Up	Média ± Desvio: 90.9544±13.1263	Média ± Desvio: 45.0882±6.4661	Média ± Desvio: 38.6061±20.5619
Walking Down	Média ± Desvio: 97.9883±14.007	Média ± Desvio: 35.1309±23.9388	Média ± Desvio: 8.7824±19.7941

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 pudéssemos realizar este exercício, criamos uma função "exercicio3_4(arraydeficheiros,pergunta)", sendo que para realização desta alinea é necessario que a "pergunta" seja igual a 1. Esta função que percorre todos os ficheiros, e junta as atividades dinamicas conforme o seu eixo, as atividades estáticas e as de transição. E criou-se uma função chamada "Plot3D()" e enviamos todos os arrays das atividades de todos os eixos(x,y,z) e plotamos em 3D.

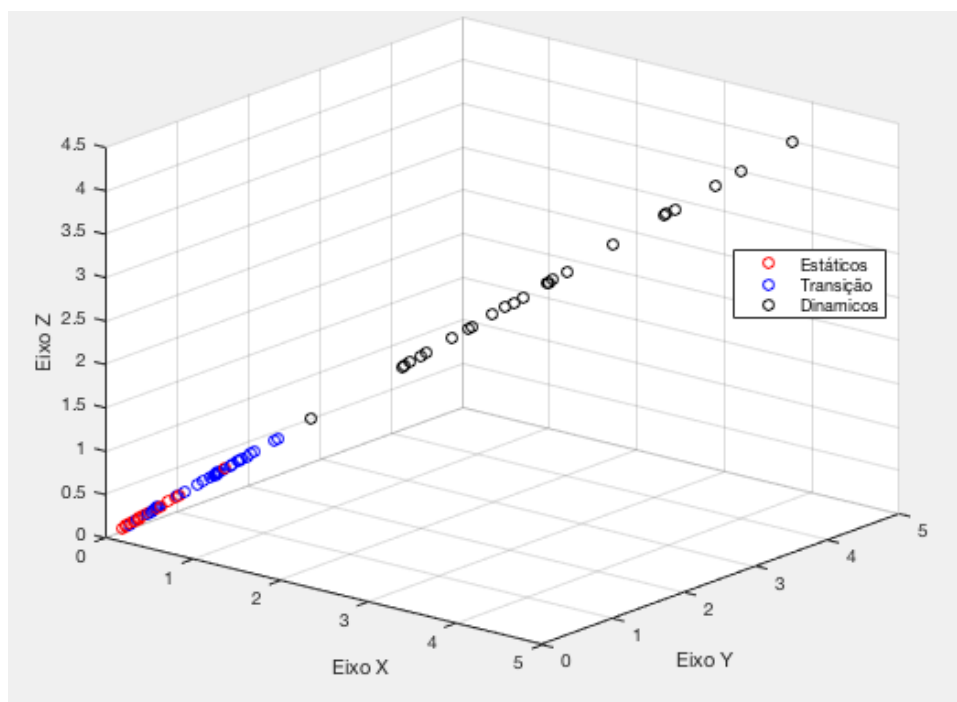


Gráfico nº1 - Gráfico 3D com a média dos picos dos 3 tipos de atividade.

O gráfico nº vai de acordo com o esperado, visto que podemos observar a discrepância das frequências entre as atividades dinâmicas e as restantes.

Avaliar os resultados obtidos em termos de sensibilidade e especificidade.

Para pudermos diferenciar, usamos a média da magnitude dos picos, mas só pequena parte dos picos são verdadeiramente válidos, ou seja, a sensibilidade das amostras da magnitude.

Em termos de especificidade, teríamos que verificar se não existe nenhuma frequência superior à média da magnitude obtida através dos picos, logo, teríamos de verificar se não haveria nenhum valor não relevante.

O desempenho esperado será ligeiramente baixo para grande parte das atividades, principalmente porque a média da magnitude também o é.

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.

Tal como no exercício anterior iremos usar a função "Exercicio3_4()" mas com a "pergunta"=2. Criou-se uma função "Plot3D_2()" que recebe por parâmetro todos os arrays (x,y,z) auxiliares dos 3 tipos de atividades.

Sendo que irá percorrer todos os arrays identificando assim cada atividade com um tipo de cor, ou forma.

O gráfico obtido foi o seguinte:

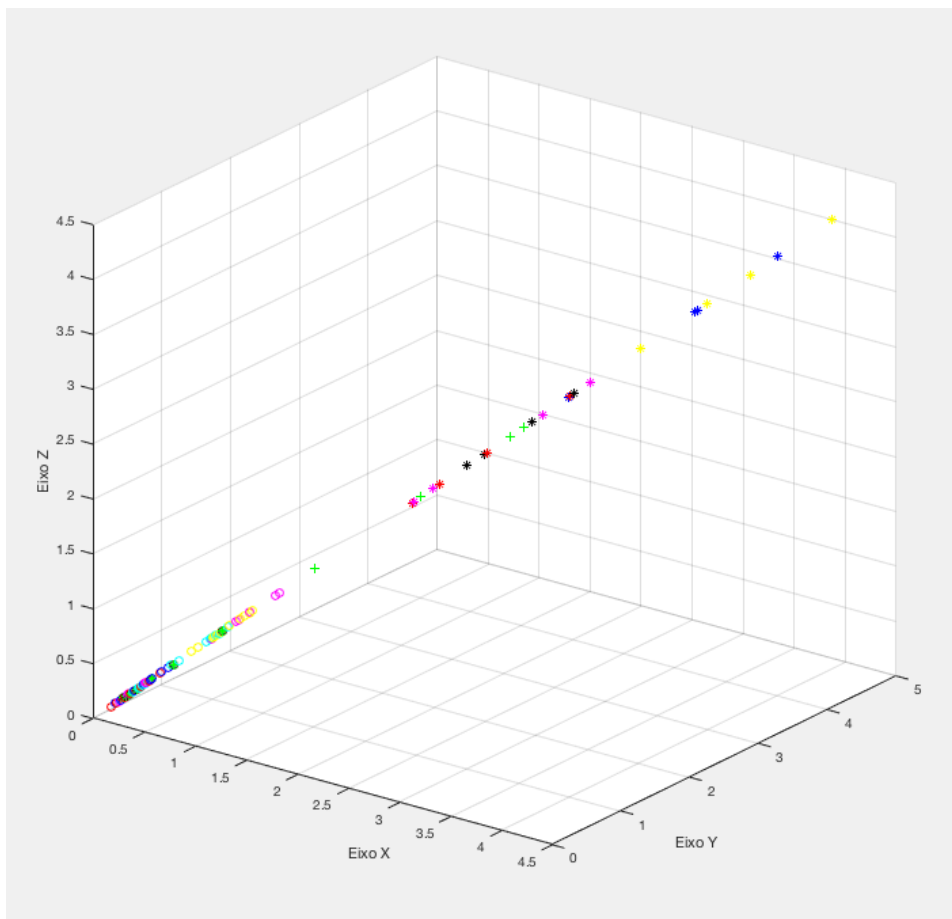


Gráfico nº 2 - Gráfico 3D com a média dos picos das várias atividades.

Legenda:(vermelho)=W;* (verde)=WU ; o (azul)=WD;(ciano)=SIT;(magenta)=STAND;(amarelo)

 =LIE;* (vermelho)=STAND_SIT; +(verde)= SIT_STAND;* (azul)=SIT_LIE;* (preto) = LIE_SIT; *(magenta) = STAND_LIE;

 *(amarelo)=LIE_STAND;

Avaliar os resultados obtidos em termos de sensibilidade e especificidade.

Em termos de sensibilidade e especificidade, concluímos o mesmo que o exercício 3.4.

4.1. Escolher, justificadamente, a janela a considerar na aplicação da STFT. Para isso, selecionar 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.

Calculamos as DFT's do segmento com as diferentes janelas que foram as seguintes:

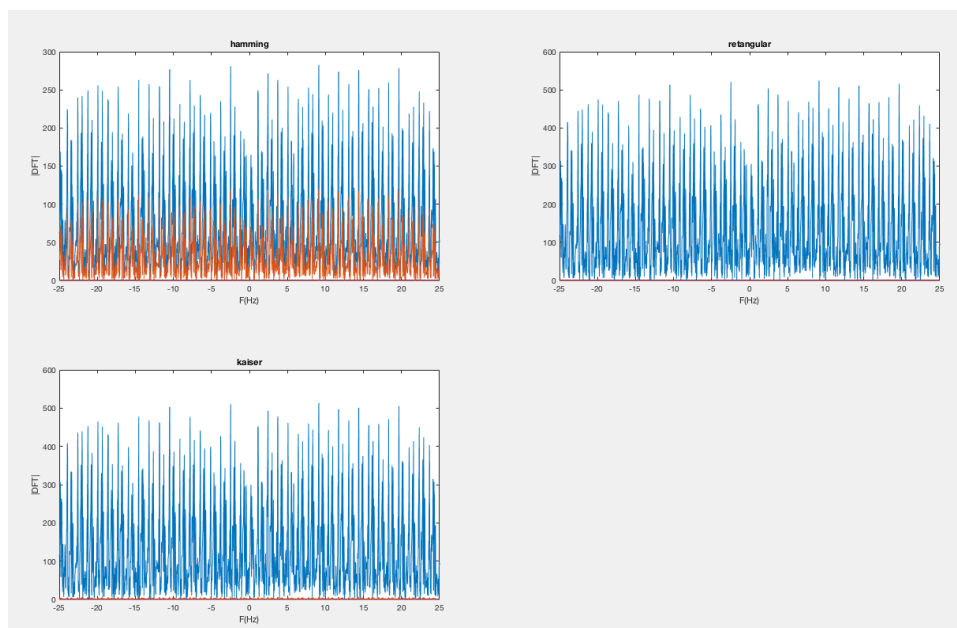


Figura nº - DFT's das diferentes janelas

Com a observação dos resultados conseguimos concluir que a aplicação de uma janela de hamming traz algumas melhorias na qualidade dos resultados, não existindo mudanças tão bruscas nas transições entre janelas. Logo iremos usar uma aplicação com uma janela de hamming.

Para resolução deste exercício é criamos uma função “stftNew()”, em que utilizamos uma janela de hamming.

Para executar esta função apenas necessita de colocar os dados a que quer a short-time fourier transform(STFT). Exemplo:”stftNew(dados)”;

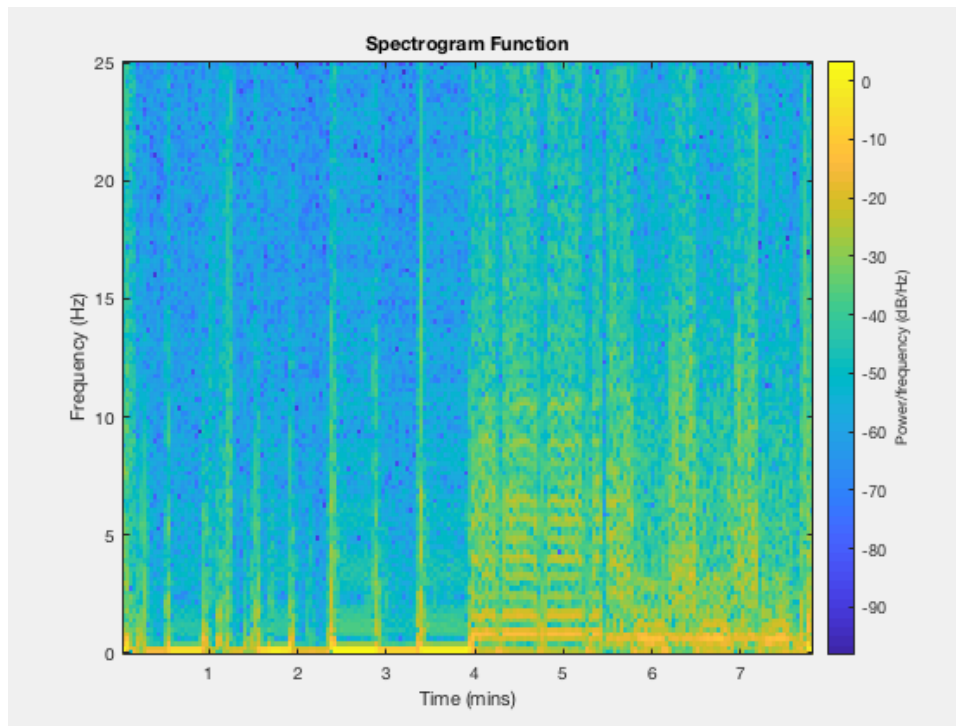
4.3. Aplicar a STFT ao sinal com a janela deslizante selecionada, usando os parâmetros corretos para visualizar de forma conveniente as diferentes atividades, não descorando a visualização apropriada da “assinatura” espectral de cada uma delas. Avaliar os resultados obtidos em termos de sensibilidade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

Neste exercício vamos usar a função criada acima, damos import do eixo Z do ficheiro a que queremos aplicar a STFT. Por exemplo se o array do eixo Z se chamar “Dados” a função fica “STFTNew(Dados)”.

Após alguns testes, chegamos à conclusão que uma janela deslizante com um tamanho 0.01x menor que o tempo total da atividade, irá permitir uma melhor resolução para identificar cada tipo de atividade no espectrograma.

Com o intuito de minimizar a atenuação da janela e a continuidade temporal nos locais afastados do seu centro usámos uma sobreposição de 50%.

Os resultados foram os seguintes:



É possível observar que dos 0 aos 4 minutos ocorrem atividades maioritariamente estáticas, e de certa forma algumas de transição, mas a partir dos 4 minutos é possível observar as atividades dinâmicas.

Conclusão:

Com este projeto, permitiu-nos melhorar as nossas capacidades e conhecimentos em relação à análise e transformação de dados e a melhorar a capacidade de programar numa nova linguagem(Matlab).

Biografia

Material de apoio à disciplina do UCstudent.