Marta Pereira Neves

TÍTULO DA TESE (MÁXIMO 130 CARACTERES)

THESIS TITLE (MAX 130 CHARACTERS)

DOCUMENTO PROVISÓRIO

Marta Pereira Neves

TÍTULO DA TESE (MÁXIMO 130 CARACTERES)

THESIS TITLE (MAX 130 CHARACTERS)

o júri / the jury

presidente / president

Prof. Doutor João Antunes da Silva
professor associado da Universidade de Aveiro

vogais / examiners committee

Prof. Doutor João Antunes da Silva
professor associado da Universidade de Aveiro

Prof. Doutor João Antunes da Silva
professor associado da Universidade de Aveiro

Prof. Doutor João Antunes da Silva
professor associado da Universidade de Aveiro

Prof. Doutor João Antunes da Silva
professor associado da Universidade de Aveiro

Prof. Doutor João Antunes da Silva
professor associado da Universidade de Aveiro

**agradecimentos /
acknowledgements**

Agradeço toda a ajuda a todos os meus colegas e companheiros.

palavras-chave

texto livro, arquitetura, história, construção, materiais de construção, saber tradicional.

resumo

Um resumo é um pequeno apanhado de um trabalho mais longo (como uma tese, dissertação ou trabalho de pesquisa). O resumo relata de forma concisa os objetivos e resultados da sua pesquisa, para que os leitores saibam exatamente o que se aborda no seu documento.

Embora a estrutura possa variar um pouco dependendo da sua área de estudo, o seu resumo deve descrever o propósito do seu trabalho, os métodos que você usou e as conclusões a que chegou.

Uma maneira comum de estruturar um resumo é usar a estrutura IMRaD. Isso significa:

- Introdução

- Métodos

- Resultados

- Discussão

Veja mais pormenores aqui:
`https://www.scribbr.com/dissertation/abstract/`

keywords   textbook, architecture, history, construction, construction materials, traditional knowledge.

abstract   An abstract is a short summary of a longer work (such as a thesis, dissertation or research paper).
The abstract concisely reports the aims and outcomes of your research, so that readers know exactly what your paper is about.
Although the structure may vary slightly depending on your discipline, your abstract should describe the purpose of your work, the methods you've used, and the conclusions you've drawn.
One common way to structure your abstract is to use the IMRaD structure. This stands for:

- Introduction

- Methods

- Results

- Discussion

Check for more details here:
`https://www.scribbr.com/dissertation/abstract/`

acknowledgement of use of AI tools

Recognition of the use of generative Artificial Intelligence technologies and tools, software and other support tools.

I acknowledge the use of [insert AI system(s) and link] to [specific use of generative artificial intelligence or other tasks]. I acknowledge the use of [software, codes or platforms] to [specific use software, codes or platforms or to other tasks].

Example 1: I acknowledge the use of ChatGPT 3.5 (Open AI, https://chat.openai.com) to summarise the initial notes and to proofread the final draft and the use of Office365 (Microsoft, https://www.office.com) for text writing and productivity.

Example 2: No content generated by AI technologies has been used in this Thesis.

Contents

List of Figures

List of Tables

Glossário

ECG   Electrocardiogram

Introduction

A short description of the chapter.
A memorable quote can also be used.

1.1 Fontes

- Tiny
- Scriptsize
- Footnotes
- Small
- Normal
- large
- Large
- LARGE
- huge
- Huge

1.2 Unidades

Utilizando o pacote `siunitx` é possível utilizar unidades do Sistema Internacional. Exemplo: a aceleração da gravidade é de $9.8\,\mathrm{m\,s^{-2}}$ e um ficheiro ocupa $1\,\mathrm{MiB}$.

1.3 Code Blocks

Uma listagem pode ser apresentada com o ambiente `listing`, que é um float (objeto flutuante, tal como uma figura ou uma tabela).

A listagem em Código **??** mostra um exemplo em C.

1.4 Citações

Algumas formas distintas de citar:

- **Apenas referência: rfc44**
- **Apenas data: rfc44**
- **Apenas ano: rfc44**
- **Apenas autor: rfc44**
- **Apenas editor:rfc44**
- **Autor e referência:rfc44**

Theoretical Introduction

Methods and Procedure

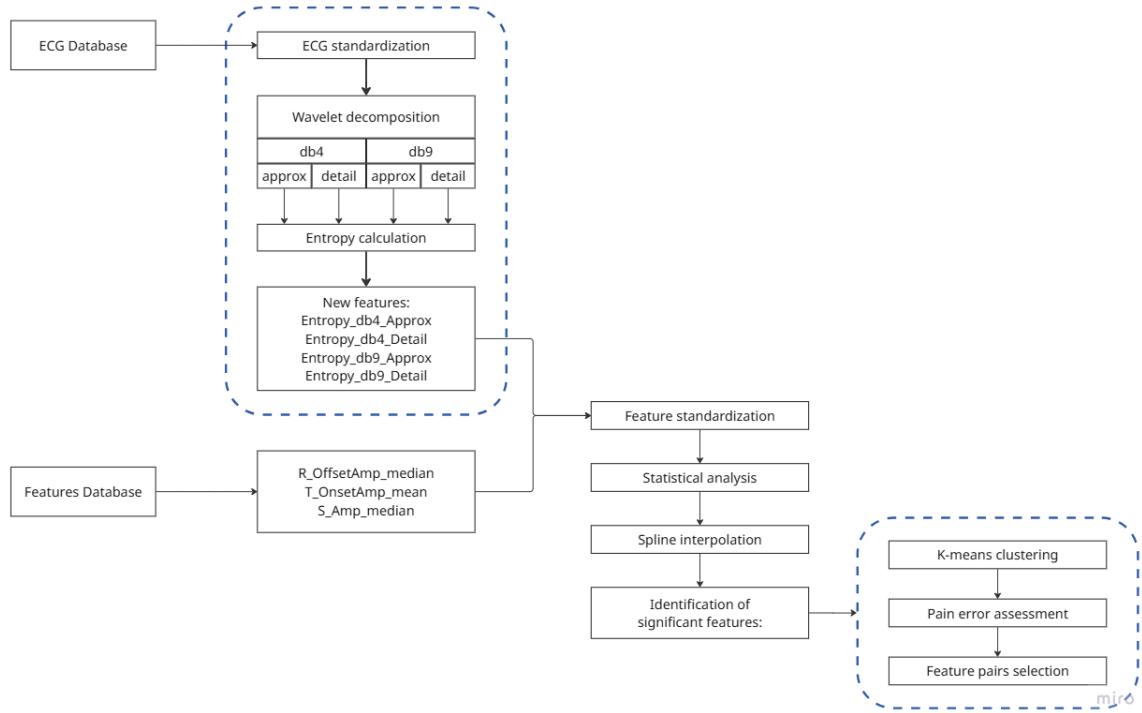In figure 3.1, the processing pipeline for this project is shown.

Figure 3.1: Processing pipeline.

3.1 DATABASE ACQUISITION

In this project, two databases were used: one has data from an Electrocardiogram (ECG) signal while the other contains features extracted from that ECG, as described in the article by Alves et al [1].

The data collection protocol followed a structured sequence designed to elicit and record both emotional and pain-related responses. It began with a 5-minute baseline period, during

which participants remained seated in a relaxed position without any external stimuli. During this time, only physiological signals were recorded to establish a baseline. Subsequently, participants watched a 10-minute video, composed of excerpts from comedy, horror, or documentary films, aimed at inducing positive, negative, or neutral emotional states, respectively. This was immediately followed by a second 5-minute stimulus-free period, allowing for recovery and stabilization of physiological responses. Next, participants underwent a Cold Pressor Test (CPT), in which they immersed their non-dominant hand in a tank of cold water maintained at 7 ± 1°C. This procedure aimed to induce pain, and participants reported their experience using the Numerical Pain Scale (NPS) at four key time points: (1) before immersing the hand, (2) at the moment pain was first perceived (Pain Threshold), (3) when the pain became unbearable (Pain Tolerance), and (4) three minutes after hand removal. The CPT phase concluded either when the participant reached their pain tolerance or after a maximum of 2 minutes, whichever occurred first. Finally, the protocol concluded with a third 5-minute period without stimuli, serving as a rest phase to observe post-task physiological responses. This process is depicted in figure 3.2.
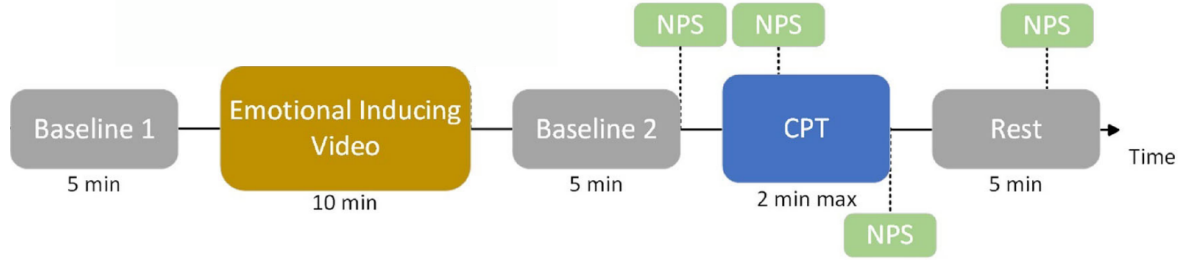
(Referenciar todos os acrónimos?)

Figure 3.2: Scheme of the protocol applied to obtain the database. (Adapted from [1])

During the study, three physiological signals were recorded: electrocardiogram (ECG), electrodermal activity (EDA), and electromyography (EMG), with EMG signals collected specifically from the trapezius and triceps muscles. Among these, the ECG data were stored in a dedicated database, which was subsequently used in the present project. From the ECG signals, several time series were extracted, including heart rate (HR), amplitude of wave peaks, amplitude of onsets and offsets, and the intervals between consecutive onsets, offsets, and peaks. These features were computed using 10-second sliding windows with a 50% overlap. For each window, a set of statistical metrics—namely the mean, median, and variance—was calculated, resulting in a total of 237 features. These were compiled into a features database, which was also employed in the current project.

3.2 Feature Extraction

Four of all the extrated features revealed to be more related to pain assessment, by the use of <diz o nome dos métodos que a Bruna usou para identificar as features mais importantes>. So, under the scope of this project, the median of the R wave offset amplitude, the median of the T wave onset amplitude, and the mean and median of the S wave peak amplitude were

used as a base for subsequent analysis. Considering that the previously described features are examples of the state of the art features in ECG analysis, and knowing that the pain quantification or identification is still an open area, this work wants to understand and find out other possible features that may describe the pain process. So, to accomplish such goal, an exploratory analysis to the collected ECG was performed.

3.2.1 ECG Data Exploration

When looking at the time series of the signal, in figure 3.3, it's noticeable that the range of amplitudes of the QRS complex is way bigger than that of the P and T waves. To tackle this, the ECG signal was standardized according to equation 3.1, where X is the original ECG, μ is the mean, σ is the standard deviation and Z is the standardized signal, which has mean zero and standard deviation one. This means the range of values is the same for all participants, reducing the intervariability that is induced by the initial state of the participant, since it is an uncontrolled variable.
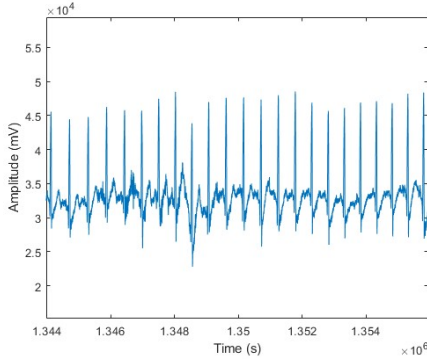
$$Z = \frac{X - \mu}{\sigma} \tag{3.1}$$

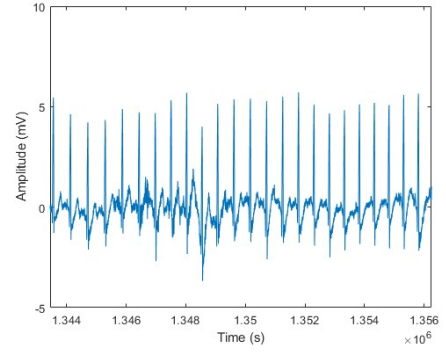**Figure 3.3:** Section of the ECG signal of a participant.

Figure 3.4: Section of the standardized ECG signal of a participant.

Once the signal was standardized, fluctuations were noticed in the time series. By removing high frequencies from the signal, by filtering for example, these would be alleviated. However, there could be information related to pain associated to those frequencies.

(Falar das famílias de wavelets para análise de ECGs?)

To choose the best type of wavelet, two criteria were established: one's format should be similar to that of the ECG wave and the other should have a higher frequency, while still maintaining its smoothness. Accordingly, Daubechies-4 ('db4') and Daubechies-9 ('db9') were chosen. Regarding the number of levels of decomposition, as can be seen in figure 3.5, when there are two levels the R peak is noticeable in the first detail, which is not something this project means to highlight. Therefore, only one level of decomposition was applied.
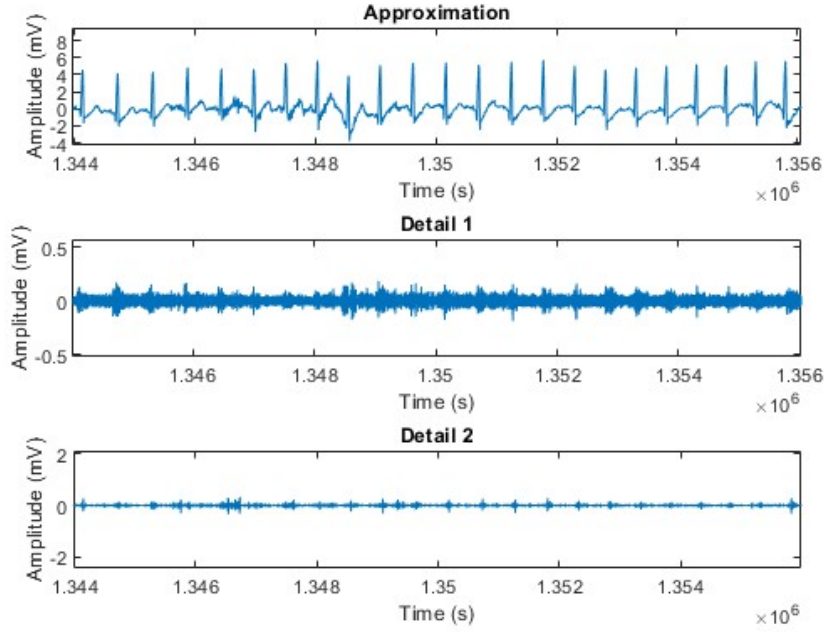
Figure 3.5: Daubechies-4 Wavelet Decomposition - level 2.

3.2.2 Entropy Calculation

Entropy quantifies the order or disorder of information that a signal presents. Since this quantity might be affected in the ECG when someone feels pain, it was extracted from the resulting wavelets, functioning as a new feature. In order to keep consistency with the other features, a same sized window was used, that is, 10 seconds with 50% overlap.

(Justificar tipo de entropia que escolhi? - approximation)

3.3 DATA ANALYSIS

To do the optimal processing of the features, fifteen participants were selected at random. After standardizing the features, so that they're comparable with each other, graphics of the time series of the features were plotted for each participant. An example can be seen in figure 3.6, in which a clear change can be seen when the participant dips their hand in water, feeling pain. Since the timeseries of the mean and median of the S wave amplitude were so similar, it was decided that only the median would be analysed.

(Falta ajustar imagem para eixo do y não ficar sobreposto.)

Subsequently, a statistical analysis of the participants' features was conducted using histograms, as illustrated in Figure 3.7. The histograms revealed that the distributions of the features deviate from normality, justifying the use of the median and interquartile range (IQR) as measures of central tendency and dispersion, respectively. These values were computed for each feature and considered as derived features that will integrate the analysis.

The goal of this project is to select features that distinguish pain from no pain. To do this, it's important to guarantee the synchronization between segments, so that they're comparable.
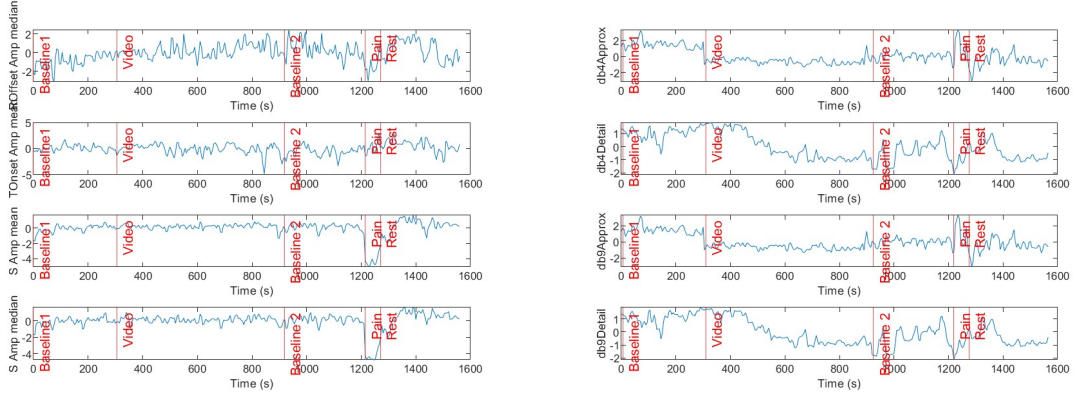
7

Figure 3.6: Standardized features' time series.
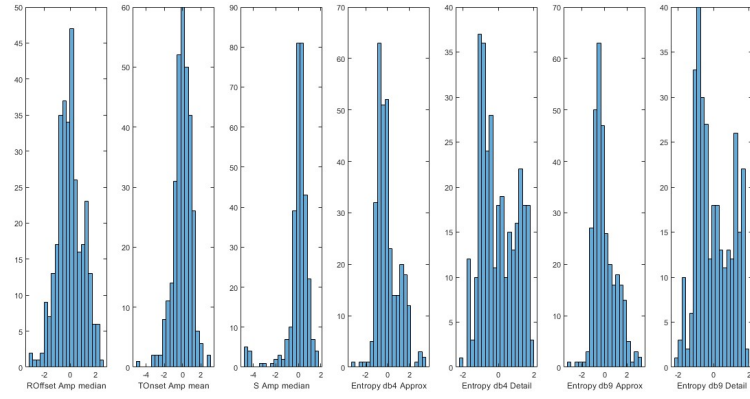

Figure 3.7: Histograms of the features.

To accomplish this goal, spline interpolation was done in each step, conducing to equalize the number of points to a 60 seconds intervals. The result of this method is portrayed in figure 3.8.
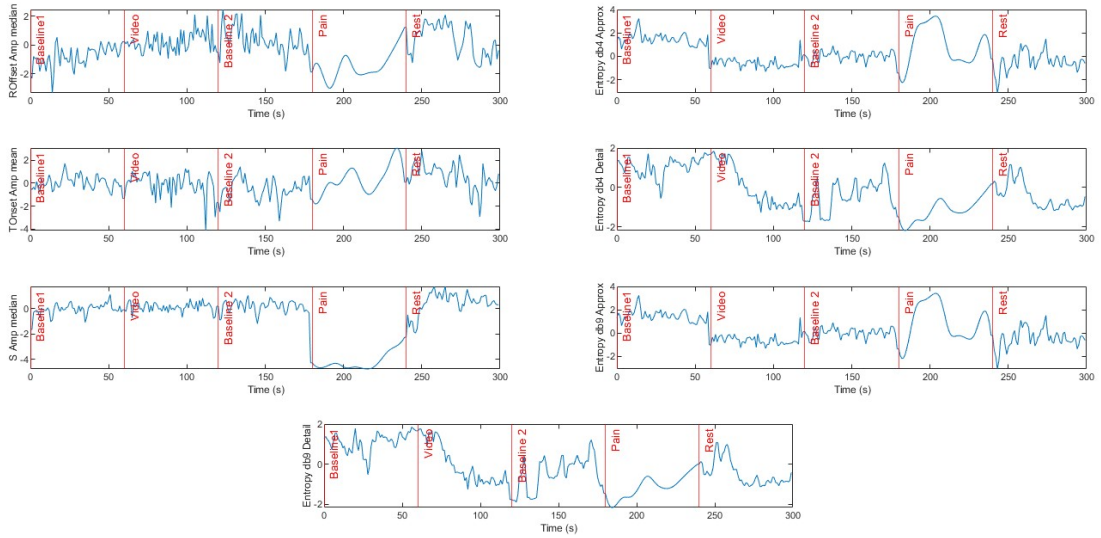

Figure 3.8: Features normalized in time after interpolation.

References

[1] B. Alves, S. Brás, and R. Sebastião, "Decoding pain: Prediction under different emotional contexts through physiological signals", *International Journal of Data Science and Analytics*, Oct. 2024, ISSN: 2364-415X. DOI: 10.1007/s41060-024-00649-z.

Additional content