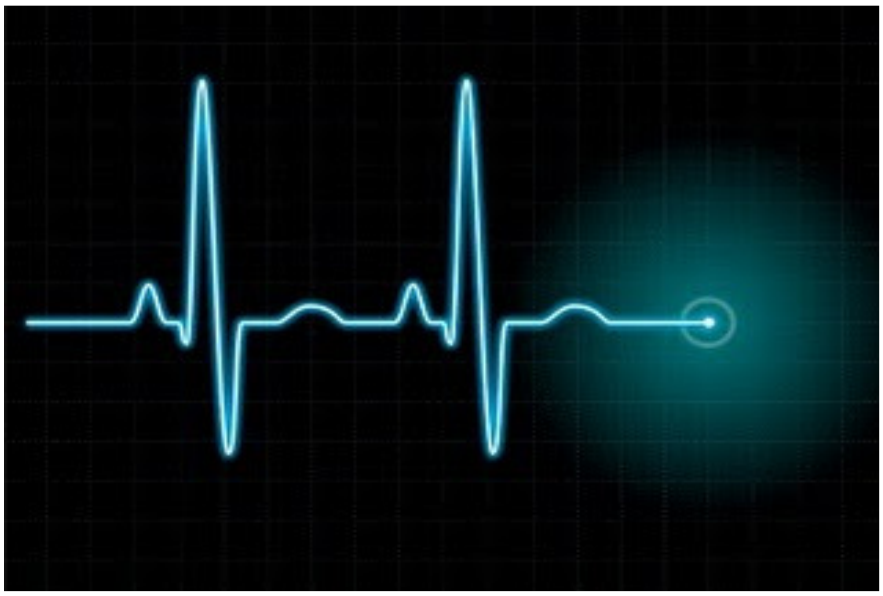
INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores

ISEL



Sistema de Reconhecimento Biométrico Baseado no Electrocardiograma

NUNO MIGUEL MARQUES ABREU
(Bacharel)

Trabalho de projecto realizado para obtenção do grau de Mestre em Engenharia Informática e de Computadores

Orientadores:
Professor David Coutinho, ISEL
Professor André Lourenço, ISEL

Júri:
Presidente: Professor Coordenador Fernando Sousa, ISEL
Vogais:
Professor Auxiliar Hugo Gamboa, FCT
Professor Adjunto David Coutinho, ISEL
Professor Assistente André Lourenço, ISEL

Março de 2012

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores

ISEL

Sistema de Reconhecimento Biométrico Baseado no Electrocardiograma

Trabalho de projecto realizado para obtenção do grau de Mestre em Engenharia Informática e de Computadores

Aluno:

Nuno Miguel Marques Abreu

Orientadores:

David Pereira Coutinho

André Lourenço

Março de 2012

Índice

Resumo

Este projecto pretende criar uma plataforma do tipo *framework*, para desenvolvimento de *software* que permita a implementação de sistemas biométricos de identificação e autenticação pessoal, usando sinais electrofisiológicos. O sinal electrocardiograma (*ECG*) é uma característica biométrica em ascensão, existindo fortes indícios de que contém informação suficiente para discriminar um indivíduo de um conjunto vasto de população.
Usa-se a *framework* desenvolvida para criar aplicações que permitam avaliar o desempenho de várias abordagens do estado da arte do reconhecimento biométrico, baseadas no *ECG*. A arquitectura típica destes sistemas biométricos inclui blocos de aquisição, pré-processamento, extracção de características e classificação de sinais *ECG*, utilizando tipicamente duas abordagens distintas. Uma das abordagens (*fiducial*) assenta em pormenores dos diferentes segmentos da forma de onda do sinal *ECG*, enquanto que a outra abordagem (*non-fiducial*) tem a vantagem de não depender criticamente desses pormenores.
Neste projecto ainda será explorada uma nova variante numa abordagem (*non-fiducial*) baseada em compressão de dados. Finalmente, pretende-se ainda estudar o desempenho destas abordagens em sinais *ECG* adquiridos nas mãos, o que constitui um desafio, dado não existirem actualmente estudos sistemáticos usando este tipo de sinais.

Palavras-chave: *ECG*, biometria, sistemas biométricos, *framework*, abordagem *non-fiducial*, *LZW*.

Abstract

This project aims to create a framework for software development, to allow the development of biometric systems for personal identification and authentication, using electrophysiological signals. The electrocardiogram (*ECG*) signal is an emerging biometric trait and there is strong evidence that it contains enough information to distinguish one individual from a wide range of people.
The framework developed is used to create applications designed to evaluate the state of the art performance of various approaches of biometric recognition based on the *ECG*. The typical architecture of these systems includes blocks, such as, the acquisition, preprocessing, feature extraction and classification of *ECG* signals. For classification typically there are two distinct approaches. One approach (*fiducial*) is based on details of the different segments of the waveform of the *ECG* signal, while the other approach (*non-fiducial*) has the advantage of not relying critically on these details.
A new approach (*non-fiducial*) variant, based on data compression, is also explored in this project. Finally, it is intended to further study the performance of these approaches in *ECG* signals acquired in the hands, which is a challenge, since there are no current systematic studies using this type of signals.

Keywords: ECG, biometrics, biometric systems, framework, non-fiducial approach, LZW.

Agradecimentos

Em primeiro lugar quero agradecer ao Prof. David Pereira Coutinho e Prof. André Lourenço pela dedicação e acompanhamento extremo, estando sempre disponíveis para ajudar e aconselhar nas alturas mais conturbadas deste projecto.

À Rita por me ter apoiado e ter dado a paz de espírito necessária para a realização do trabalho de projecto.

Aos meus pais e irmã, que sempre me apoiaram na realização do mestrado.

Ao Nuno Martins pela partilha de ideias e contribuições que ajudaram a enfrentar as longas noites de trabalho no projecto.

Ao meu chefe Rogério Silva pela disponibilidade que me foi dada nos dias de maior aperto no projecto. E a todos os meus colegas de trabalho que contribuíram para a discussão de alguns temas relevantes neste trabalho, em especial para o Ilídio Miranda e Carlos Silva.

Por fim agradeço a todos aqueles que deram contributos para este projecto, em especial à Maria Segurado.

Lista de Figuras

Lista de Tabelas

Lista de Equações

Abreviaturas

ADN	Ácido Desoxirribonucleico.
ANSI	*American National Standards Institute*: instituto nacional americano de padrões.
BSP	*Biometric Service Provider: fornecedor de serviço biométrico.*
DB	*Data Base*: base de dados.
DBNN	*Decision Based Neural Network*: rede neuronal de decisão.
DCT	*Discrete Cosine Transform*: transformada discreta do co-seno.
DWT	*Discrete Wavelet Transform*: transformada correspondente à transformada contínua de *Wavelet* para funções discretas.
ECG	Electrocardiography: electrocardiograma.
EER	Equal Error Rate: taxa de erro quando as taxas de erro FAR e FRR são iguais.
EER UT	Equal Error Rate User Tune: taxa de erro igual dentro do domínio de cada utilizador.
FAR	False Acceptance Rate: taxa de erro de aceitação. Quando a autenticação foi positiva para um impostor.
FRR	False Rejection Rate: taxa de erro de rejeição. Quando a autenticação foi negativa para um utilizador válido.
HTTP	*Hypertext Transfer Protocol*: protocolo de transferência de hipertexto.
IdError	*Identification Error*: taxa de erro para a identificação pessoal.
ISO	*International Organization for Standardization*: organização internacional de normalização.
IIR	*Infinit Inpulse Response*: filtro digital com resposta ao impulso de duração infinita.
KNN	K-Nearest Neighbor: método para classificação baseado na proximidade.
LDA	*Linea Discriminant Analysis.*
LZ77	Algoritmo de compressão de dados proposto por *Ziv* e *Lempel* em 1977.
LZ78	Algoritmo de compressão de dados proposto por *Ziv* e *Lempel* em 1978.
LZSS	Implementação do algoritmo *LZ77* proposta por *Storer* e

	Szymanski.
LZW	Implementação do algoritmo *LZ78* proposta por *Welch*.
MOBD	*Multiplication of the Backward Distance.*
MVC	*Model-View-Controller*: padrão de desenho de *software* para separação das camadas dados, controlo e apresentação.
NMODEL	*Number of Models*: número de amostras do sinal *ECG* a usar no modelo.
NREP	*Number of Repetitions*: número de repetições de testes de avaliação de desempenho.
NTEST	*Number of Testes*: número de amostras do sinal *ECG* a usar no teste.
PC	*Personal Computer*: computador pessoal.
PDA	*Personal Digital Assistent*: assistente digital pessoal.
RAW	Dados não processados.
SNR	*Signal-to-Noise Ratio*: relação sinal ruído.
SOAP	*Simple Object Access Protocol:* protocolo de acesso a objectos.
UI	*User interface*: *interface* do utilizador.
WBF	*Windows Biometric Framework API:* framework biométrica da *Microsoft.*
WDIST	*Wavelet distance*: distância entre sinais baseada na transformada de *Wavelets*.
WSDL	*Web Services Description Language*: linguagem baseada em *XML* para descrever de serviços disponíveis sobre a Internet.
XML	*Extensible Markup Language*: linguagem que permite definir novas linguagens baseadas em marcas.
ZMM	*Ziv-Merhav Method:* método para estimação da entropia relativa proposto por *Ziv* e *Merhav.*

1 Introdução

Ao longo dos anos os seres humanos têm usado características corporais, como as da face ou da voz, para reconhecimento uns dos outros. No século XIX o chefe de polícia *Alphonse Bertillon*, do departamento criminal em Paris, pôs em prática uma ideia sua e usou um conjunto de medições corporais para identificação de criminosos. Esta ideia ganhou popularidade mas também rapidamente a perdeu com a descoberta da impressão digital, no final século XIX, que ainda hoje é um processo de identificação bastante usual. No entanto, outros processos surgiram baseados principalmente em características faciais [1].

Com o aumento da população e da mobilidade, começamos a depender cada vez mais de documentos e segredos para confirmar a identidade. As *passwords* são actualmente um método bastante utilizado na confirmação dessa identidade. No entanto a maioria das pessoas baseia as suas *passwords* em palavras ou dígitos que possam facilmente ser lembrados tais como, nomes e datas de aniversários de familiares, filmes favoritos, estrelas da musica, etc. Essas *passwords* tornam-se assim vulneráveis a ataques de força bruta. No entanto é possível e aconselhável usar algumas técnicas para as tornar mais fortes, minimizando esses ataques. Por exemplo: a alteração frequente da *password*, não usar a mesma *password* em diferentes aplicações, usar *passwords* longas, etc. Estas técnicas tornam-na mais forte, mas por outro lado mais difíceis de lembrar, o que leva muitas vezes a que os utilizadores as tenham que escrever num papel, podendo assim ficar ao alcance de qualquer um. Neste ponto de vista, a biometria é algo extremamente conveniente, pois as características corporais sendo intrínsecas ao ser humano não podem ser "perdidas ou esquecidas".

1.1 Noções básicas de biometria

A biometria é a ciência que estuda as características físicas ou comportamentais dos seres vivos. A palavra tem origem em duas palavras gregas, "*bios*" e "*metros*", que significam vida e medidas, respectivamente.
As características biometrias podem ser divididas em características físicas ou comportamentais [1]. Características físicas, são aquelas que se mantêm sempre inalteradas independentemente do comportamento humano (tamanho do nariz, distância entre olhos, etc.). Enquanto que as características comportamentais (forma de caminhar, forma de escrever num teclado) podem variar de acordo com factores comportamentais, tais como o humor, a fadiga, etc. As características comportamentais são normalmente mais difíceis de utilizar para a identificação pessoal pela sua componente variável, no entanto foram propostos sistemas que tiram partido destas características para a obtenção do estado emocional de um indivíduo [16].

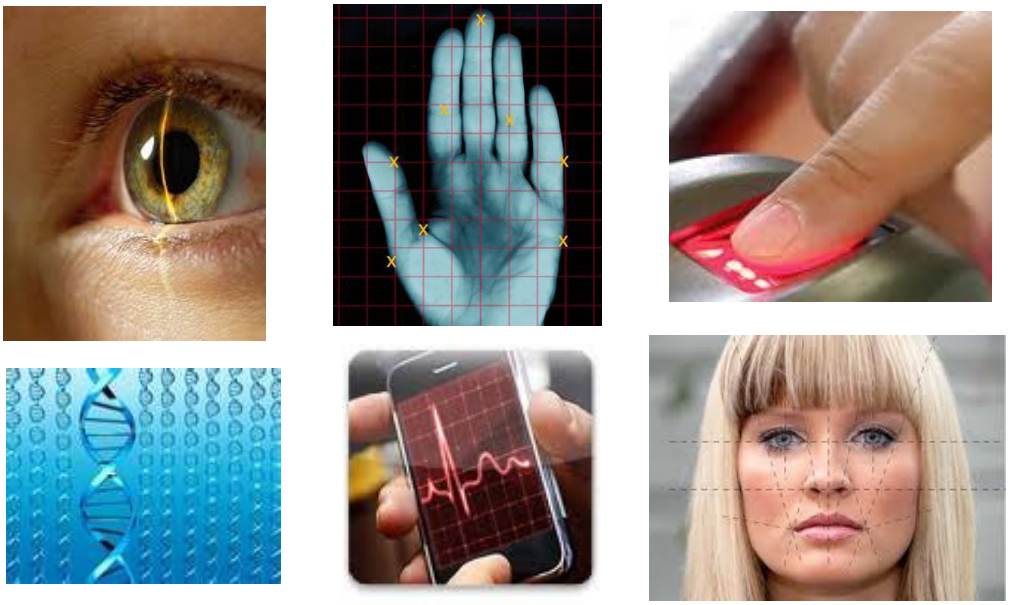


Figura 1 – Exemplos de características biométricas físicas.

A Figura 1 apresenta um conjunto de exemplos de características físicas, tais como: íris, geometria da mão, impressão digital, *ADN*, sinal de electrocardiograma e medidas faciais. Por outro lado na Figura 2 é apresentado um conjunto de características comportamentais: forma de andar, suor, forma da caligrafia e modo de escrita, forma de escrever num teclado (força exercida) e mais uma vez o sinal de electrocardiograma. Note-se que este último reflecte tanto características físicas como comportamentais.

Figura 2 – Exemplos de características biométricas comportamentais.

Desde há muito tempo que o Homem tem utilizado diferentes métodos para restringir o acesso a recursos. A forma de o fazer sofreu sucessivas evoluções, com o objectivo de aumentar o nível de segurança no acesso ao recurso. A Figura 3 ilustra a evolução do método utilizado com o respectivo nível de segurança.

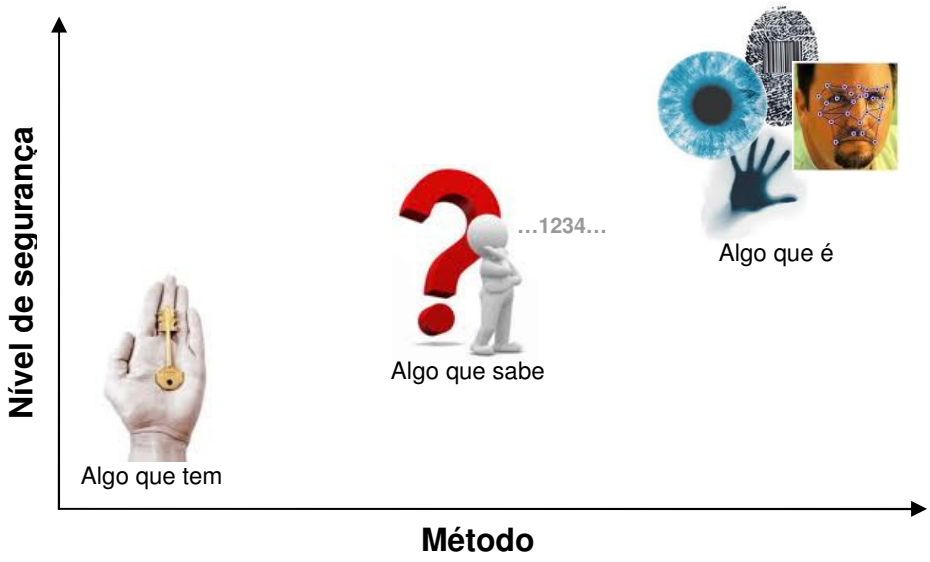


Figura 3 – Relação entre o nível de segurança e o método de acesso.

Inicialmente o método mais usado era a posse de algo que permitia o acesso ao recurso, como por exemplo uma chave. Pausou-se para um outro método, em que a pessoa teria que memorizar um código para aceder ao recurso. Actualmente o método passou a basear-se naquilo que as pessoas são, ou seja começamos a usar a biometria.

Apesar da biometria ter muitas vantagens, como o aumento do nível de segurança, apresenta ainda algumas desvantagens. A lista seguinte enumera as vantagens e desvantagem de usar a biometria no acesso a recursos.

- **Vantagens**:
  - Aumento do nível de segurança.
  - Desencorajamento de fraudes.
  - Algo difícil de ser transferido, esquecido, perdido ou copiado.
  - Comodidade.
- **Desvantagens**:
  - O resultado não é SIM ou NÃO, mas um grau de certeza.
  - Se for comprometido não pode ser feito "*reset*", tal como a uma *password*.
  - O sistema biométrico em si pode ser "atacado".
  - Levanta problemas de privacidade de dados [10].

1.2 Arquitectura de um sistema biométrico

Um sistema biométrico é normalmente baseado num sistema de reconhecimento de padrões com três etapas (ver Figura 4): pré-processamento, extracção de características e classificador [2]. A etapa de pré-processamento realiza as operações de filtragem, normalização, eliminação de ruído e eventualmente outras operações que pretendem dotar o sinal adquirido de uma representação conveniente para a próxima etapa. Na extracção de características são retiradas as características que facilitem a distinção entre os diversos indivíduos ou classes, usando a designação comum em reconhecimento de padrões. Esta etapa, num sistema de reconhecimento de padrões clássico, é considerada a mais importante visto que, depende directamente dos padrões envolvidos no problema e é onde devem ser seleccionadas as características que melhor distinguem as diversas classes. A etapa do classificador analisa as características e decide qual a classe a que os padrões adquiridos pertence. O desenho dos classificadores é normalmente independente do problema [2]. Existem diversas abordagens mas as mais simples, apenas calculam a distância mais próxima entre um conjunto de pontos, os padrões adquiridos para identificação e os representativos da classe (*templates*).

Figura 4 – Estrutura clássica de um sistema de reconhecimento de padrões aplicado ao problema da identificação.

Os sistemas biométricos possuem normalmente três modos de operação: Registo, Autenticação e Identificação. A implementação de qualquer uma destas duas últimas operações obriga à sua realização em duas fases distintas: primeiro a fase de registo de *templates* e posteriormente a fase de testes.
A fase de registo, consiste no carregamento de dados biométricos (chamados *templates*) que servirão de referência para a tomada de decisão a decorrer durante a fase de testes. Esta por sua vez consiste na acção de reconhecimento de padrões com vista à execução de operações de identificação e/ou autenticação pessoal.
No entanto, durante a fase de teste também podem existir operações de registo, para o carregamento de *templates* de novos utilizadores ou até mesmo de utilizadores já existentes (actualização ou substituição de *templates*).
Na Figura 5 ilustra-se o fluxo do processamento do sinal *ECG* através da sequência de blocos utilizados em cada uma das operações de um sistema biométrico.
Na operação de registo, o sistema realiza o pré-processamento do sinal biométrico, a extracção de características e finalmente decide se estas

são ou não válidas para concluir o registo com sucesso. Caso a validação seja positiva, os dados biométricos são guardados na base de dados como *templates*.
Na autenticação, o sistema valida a identidade dos utilizadores por comparação entre os dados biométricos extraídos e os dados biométricos guardados numa base de dados (*templates*). Neste caso o utilizador terá que apresentar a sua identificação, para que o sistema possa obter os dados biométricos (*templates)* e compara-los com os obtidos nesse instante.
Finalmente na operação de identificação o sistema determina uma identidade apenas com base nos dados biométricos extraídos, tendo o sistema que realizar diversas comparações, percorrendo os dados biométricos (*templates*) de todos os utilizadores, guardados na base de dados, e decidir qual a identidade do utilizador. Nesta operação o sistema pode devolver a identidade do utilizador ou uma indicação de erro no caso de impossibilidade de identificação.

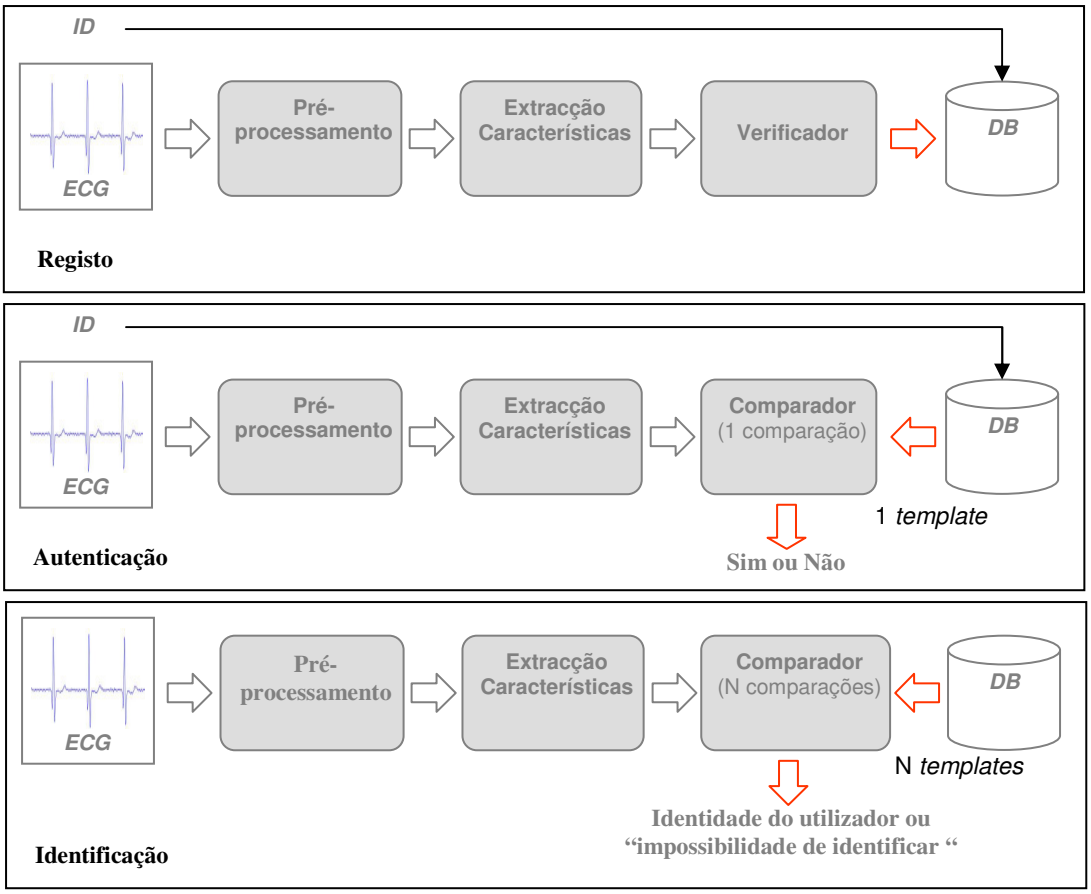


Figura 5 – Diagrama de blocos para as operações de registo, autenticação e de identificação de um sistema biométrico.

Como se pode ver na Figura 5, os blocos utilizados num sistema clássico de reconhecimento de padrões, já descritos na Figura 4, são utilizados de igual forma em cada uma das três operações biométricas, existindo apenas pequenas diferenças de funcionamento no bloco do classificador

para cada operação (assinaladas a cor vermelha na Figura 5), conforme se passa a descrever:

- **Registo** – o classificador é substituído por uma bloco de verificação da qualidade das características extraídas do sinal biométrico, que determina se o registo foi ou não bem sucedido e tem a qualidade necessária para armazenamento na base de dados.
- **Autenticação** – o classificador compara um único *template* com o sinal biométrico apresentado ao sistema decidindo se a autenticação é ou não válida.
- **Identificação** – o classificador compara o sinal biométrico apresentado ao sistema com todos os *templates* existente na base de dados, decidindo qual a identidade do utilizador.

1.3 Sinal do electrocardiograma

O electrocardiograma (*ECG*) é o registo do sinal eléctrico produzido pelo coração durante a sua actividade (ver Figura 6). Cada pessoa apresenta um sinal *ECG* distinto, que pode servir como característica biométrica. Existem fortes evidências de que este sinal é suficientemente discriminativo para identificar um indivíduo num vasto grupo populacional. A própria medição deste sinal possui inerentemente a verificação de que a pessoa está viva. Outras informações podem ser obtidas deste sinal, tais como, diferentes estados de emoção ou *stress* [3].

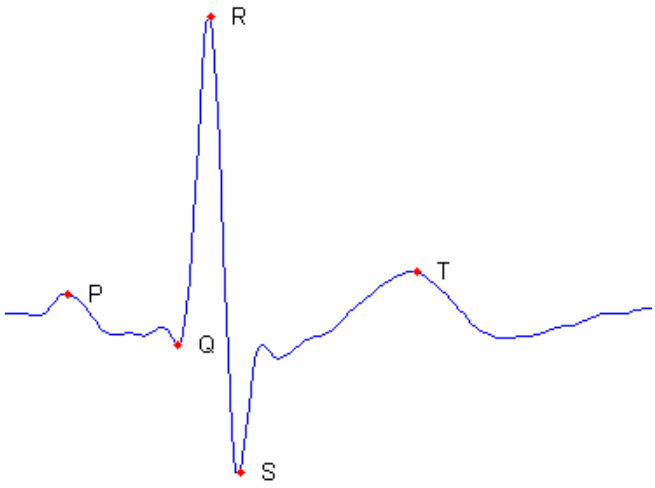


Figura 6 – Forma de onda típica de um segmento do sinal ECG[1].

A Figura 7 apresenta cinco segmentos do sinal *ECG* sobrepostos, para quatro indivíduos. Podemos observar que as formas de onda do sinal são significativamente diferentes entre indivíduos, o que facilmente os distingue uns dos outros. No entanto para o mesmo indivíduo, os cinco

[1] Imagem gerada pelo protótipo desenvolvido em *MATLAB*, recorrendo à abordagem *fiducial*, com dados obtidos através de um eléctrodo colocado junto ao peito (derivaçãoV2*)*.

segmentos apresentam praticamente a mesma forma de onda. Isto indica que a variação do sinal *ECG* para o mesmo indivíduo é normalmente mínima.
Estes sinais foram obtidos em indivíduos em repouso, utilizando um eléctrodo colocado junto ao peito na derivação *V2*, conforme se mostra na Figura 8.

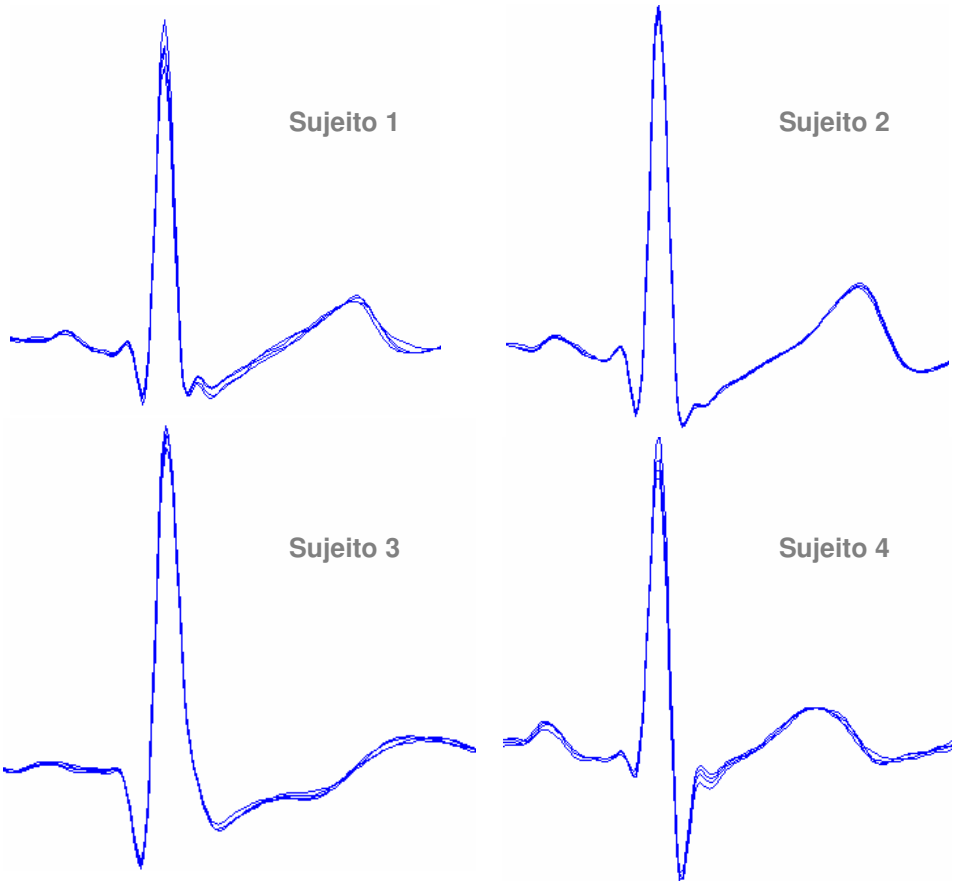


Figura 7 – Cinco segmentos do sinal *ECG*[1] sobrepostos, de quatro sujeitos diferentes.

O sinal *ECG* pode ser dividido em três partes, correspondentes à despolarização e polarização das fibras musculares que compõem o coração: a onda *P* corresponde à despolarização das aurículas do coração; a onde *QRS* corresponde à despolarização do ventrículo; e por fim, a onda *T* correspondem à polarização do ventrículo [4].
Normalmente o sinal *ECG* é obtido através da colocação de dez eléctrodos no corpo humano. Seis, são colocados sobre o peito e os restantes quatro nas extremidades dos braços e pernas, tal como podemos ver na Figura 8.

[1] Imagens gerada pelo protótipo desenvolvido em *MATLAB*, com sinais *ECG* obtidos através de um eléctrodo colocado junto ao peito (derivação *V2)*.

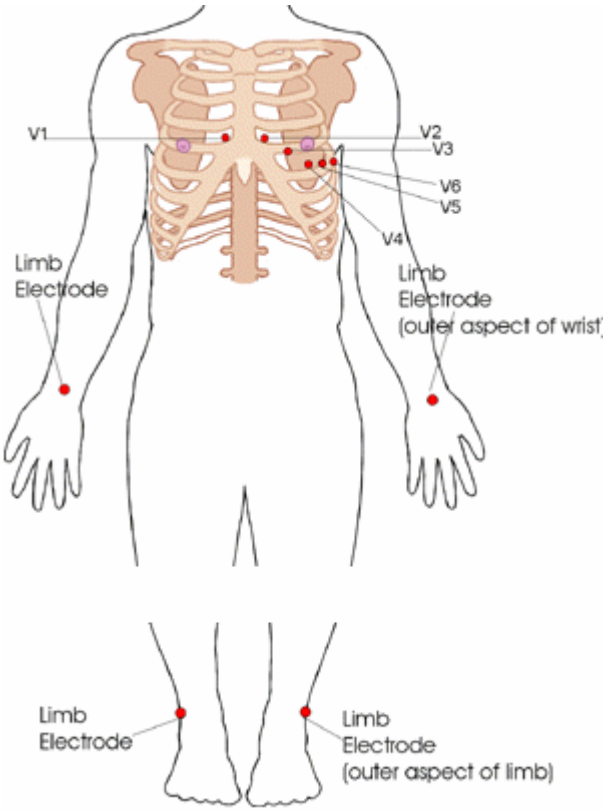


Figura 8 – Pontos convencionais onde colocar os eléctrodos para aquisição do sinal *ECG*[1].

Esta forma de obtenção do sinal *ECG* não é prática, sobretudo quando usada em sistemas biométricos que requerem uma menor intrusão. Existem alguns estudos que demonstram uma boa eficácia dos sistemas biométricos, que utilizaram sinais *ECG* obtidos com um número mais reduzido de eléctrodos [5], [6], [7] e [8]. Por exemplo [8] foi um desses estudos, que utilizou um único eléctrodo, junto ao peito na *V2*, para obter as amostras do sinal *ECG*.

1.4 Estado da arte

Os trabalhos anteriores realizados na área do reconhecimento biométrico baseados no sinal *ECG* podem ser classificados de acordo com a abordagem seguida em *fiducial* ou *non-fiducial*. Na abordagem *fiducial* usam-se pontos de referência a partir dos quais se medem o tempo ou a diferença de amplitude entre esses pontos, para a criação de um modelo (*template*). Por outro lado, a abordagem *non-fiducial* trata o sinal *ECG* ou os vários segmentos do batimento cardíaco como um todo, obtendo características de base estatística relativas à morfologia geral do sinal. Esta distinção tem uma analogia directa nas duas abordagens possíveis para os sistemas biométricos baseados no reconhecimento da face, onde se pode operar localmente e extrair características biométricas como a distância entre os olhos ou o tamanho da boca, ou por outro

[1] Imagem obtida em http://www.ambulancetechnicianstudy.co.uk/ecgbasics.html (acedido em Janeiro de 2011) e posteriormente alterada.

lado, na abordagem *non-fiducial* seria analisada a imagem facial de uma forma global.
Ambas as abordagens têm vantagens e desvantagens. Enquanto que o risco da abordagem *fiducial* é a perda de informação escondida por de trás da morfologia da biometria, na abordagem *non-fiducial* lida-se com uma grande quantidade de informação redundante que precisa de ser eliminada. O desafio está em remover essa informação de forma a minimizar a variação entre amostras do mesmo sujeito e maximizar a variação entre sujeitos diferentes.

1.4.1 Abordagens de base *fiducial*

Entre os primeiros trabalhos na área, foi a proposta de *Biel et. al.* [4] em 2001, de um algoritmo de extracção de características *fiducial* que demonstrou a viabilidade da utilização de sinais *ECG* para a identificação pessoal. Usaram o sistema de 12 eléctrodos para adquirir sinais de 20 indivíduos de várias idades. Realizaram experiências adicionais para testar o efeito da variação do local da colocação dos eléctrodos bem como do operador que os coloca. Das 30 características de diagnóstico clínico estimadas para cada um dos eléctrodos, apenas 12 foram utilizadas para comparação no treino e classificação. Compararam os resultados da combinação de diferentes características para demonstrar que a melhor taxa de classificação era de 100%, com apenas 10 características.

Nesse mesmo ano de 2001 *Kyoso & Uchiyama* [23] também propuseram uma abordagem *onde* foram seleccionadas quatro características, a duração da onda *P*, o intervalo *PQ*, o complexo *QRS* e a duração *QT*. Estas características foram identificadas nos picos através da aplicação de um limitador de derivação de segunda ordem. O sujeito com a menor distância de *Mahalanobis* [41] entre cada duas de quatro características é o seleccionado. O melhor desempenho obtido foi de 94,2%, usando apenas o *QRS* e os intervalos *QT*.

Em 2002, *Shen et al.* [24] propuseram um método de reconhecimento através do sinal *ECG* utilizando sete características baseadas no complexo *QRS*. A ideia subjacente foi que esta onda é pouco afectada por variações do batimento cardíaco, e portanto é apropriada para o reconhecimento biométrico baseado no sinal *ECG*. A metodologia proposta consistia em dois passos: no primeiro passo fazia-se a comparação de modelos para calcular o coeficiente de correlação entre os complexos *QRS*, para encontrar possíveis candidatos e diminuir o espaço de procura; no segundo passo, uma decisão baseada em redes neuronais (*DBNN*) foi então utilizada para fortalecer a validação da identificação resultante do primeiro passo. Os resultados mostraram que no primeiro passo identificou-se correctamente 85% dos casos, já com a rede neuronal obtiveram resultados de reconhecimento de 100%.

Um estudo mais completo de reconhecimento biométrico foi publicado em 2005, por *Israel et al.* [5]. Neste propôs-se uma solução com três

etapas para o reconhecimento biométrico baseado no sinal *ECG*: pré-processamento, extracção de características e a classificação. Além disso, descreveram um conjunto de várias experiências com vista ao estudo da influência do local da colocação dos eléctrodos e do stress físico. O sistema proposto utilizava apenas características temporais, obtidas depois de se aplicar ao sinal *ECG* adquirido um filtro para manter a informação discriminativa do sinal entre as frequências 1,1 – 40Hz e eliminar o restante espectro considerado ruído. Depois de observar que a melhor taxa de identificação estava perto dos 100%, concluíram que o sinal *ECG* seria uma característica biométrica que suporta variações do ritmo cardíaco.

Uma abordagem semelhante foi apresentada no mesmo ano por *Palaniappan & Krishan* [31]. Para além das características comuns usadas dentro do complexo *QRS*, um factor de forma, que é uma medida de complexidade de sinal, foi proposto e testado como entrada para o classificador de redes neuronais. Registaram uma taxa de sucesso para a identificação de 97,6% sobre registos de 10 indivíduos.

Zhang & Wei [26] em 2006, sugeriram a utilização de um método de classificação baseado no teorema de *Bayes*. Foram usados sinais *ECG* de 502 sujeitos, utilizando 14 características. Os resultados obtidos indicaram um desempenho superior na classificação baseada no teorema de *Bayes* comparada como a distância de *Mahalanobis* (de 3,5% para 13%).

Singh & Gupta [27] em 2008, propuseram um método para obtenção das ondas *P* e *T* baseado na derivação temporal e delineação adaptativa, para a obtenção de 19 características. O sistema foi avaliado sobre um conjunto de 25 indivíduos, obtendo uma taxa de desempenho de 99%.

Em 2009, *Boumbarov et al.* [28] apresentou diferentes abordagens de base *fiducial* para o sinal *ECG*, utilizando diferentes modelos. A classificação foi baseada em redes neuronais e obteve-se uma taxa de identificação que variou entre 62% e 94% para diferentes indivíduos.

Ting & Salleh [32] em 2010 apresentaram uma nova abordagem para a identificação pessoal baseada no sinal *ECG* através da utilização de um filtro. Foi anunciada uma taxa de identificação de 87,5% para um total de 13 indivíduos em repouso. Foi também anunciado que este método é robusto ao ruído acima de 20dB *SNR*.

Venkatesh & Jayaraman [33] em 2010 propuseram uma nova abordagem para a identificação e autenticação pessoal baseado do sinal *ECG*. Foram extraídas do sinal *ECG* 9 características no domínio do tempo e utilizadas na classificação. Usaram várias abordagens com diferentes classificadores sobre uma base de dados com 15 indivíduos, obtendo uma taxa de desempenho para a identificação de 100% e de 96% para a autenticação.

Tawfik et al. [34] em 2010 apresentou um estudo que propôs três métodos para a identificação pessoal baseado no sinal *ECG*. Foi utilizado uma base de dados com sinais de 22 indivíduos saudáveis. As taxas de desempenho nos três métodos variaram entre 97,73% e 99,09%.

1.4.2 Abordagens de base *non-fiducial*

Entre os primeiros estudos na área, está o estudo proposto por *Plataniotis et al.* [29] em 2006, baseado na auto-correlação como modo para obtenção de características. Com o objectivo de identificar padrões repetitivos no sinal *ECG*, o autor sugeriu a auto-correlação de uma amostra *ECG* como uma forma de evitar a obtenção de pontos de referência utilizados em abordagens de base *fiducial*. Foi demonstrado que a auto-correlação de amostras do sinal *ECG*, possui informação discriminativa dentro de um grupo populacional. Contudo, dependendo da frequência de amostragem do sinal, o número de dimensões de uma amostra resultante da auto-correlação é consideravelmente alta para uma utilização eficiente. De forma a diminuir este número de dimensões e reter apenas a informação útil para o reconhecimento de características, foi aplicado a transformada discreta de co-seno. Este método foi testado com sinais *ECG* de 14 indivíduos adquiridos há alguns anos atrás, obtendo-se uma taxa de desempenho para a identificação de 100%.

Wübbeler et al. [35] publicou em 2007 um estudo sobre a autenticação pessoal baseado no sinal *ECG*, no qual extraíram características biométricas com a ajuda de um conjunto de eléctrodos, produzindo um vector a duas dimensões de características do electrocardiograma. Utilizou-se um procedimento de *thresholding* para localizar e extrair picos. Para a classificação, foi calculada a distância entre dois vectores de amostras e calculada a primeira e segunda derivada temporal. A taxa de desempenho do sistema para a autenticação foi de 99% para um conjunto de 74 indivíduos.

Molina et al. [36] em 2007 propôs uma metodologia de síntese do sinal *ECG* para a autenticação pessoal. O segmento do batimento cardíaco é normalizado e comparado com uma estimativa, calculada a partir do próprio segmento e do modelo referente à entidade declarada. O classificador utilizou a distancia euclidiana como medida de semelhança, obtendo uma taxa de desempenho de 98%.

Em 2008, *Chan et al.* [6] publicou um conjunto de sinais *ECG* adquiridos nos dedos (polegar e indicador) com dois eléctrodos. Foi utilizada a distância *Wavelet* como medida de semelhança obtendo uma taxa de sucesso de 89,1%, que superou outros métodos de cálculo da distância usados neste estudo. Além disso, realizou-se uma nova sessão de leituras para indivíduos mal classificados, o que melhorou o desempenho do sistema para 95%.

No mesmo ano, *Chiu et al.* [7] propões a utilização da transformada discreta de Wavelet (*DWT*) para extrair características e a distância euclidiana, como medida de semelhança para a classificação. Quando o método proposto foi aplicado a uma base de dados de 35 indivíduos, obteve-se uma taxa de 100% na autenticação pessoal. O autor referiu ainda que esta taxa terá um valor mais baixo, caso sejam acrescentados 10 novos indivíduos com problemas de arritmia cardíaca.

Fatemian & Hatzinakos [30] em 2009 sugeriram igualmente a utilização da transformada *Wavelet* (*DWT*) para reduzir o ruído e segmentar o sinal *ECG,* passando posteriormente por um processo no qual cada batimento cardíaco é amostrado, normalizado, alinhado e calculada a média, criando assim um modelo para cada indivíduo. A taxa de autenticação anunciada foi de 99,6% para o conjunto de indivíduos, cada um com 2 modelos na base de dados.

Odinaka et al. [37] em 2010 publicou os resultados da análise do sinal *ECG* para reconhecimento biométrico. Os métodos propostos neste estudo foram aplicados a sinais *ECG* obtidos em três ocasiões distintas durante um período de sete meses, de forma a estudar o impacto da variabilidade a longo prazo. O método proposto neste estudo registou uma taxa de erro (*EER*) de 0,37% para a autenticação e 99% para a identificação, sobre registos obtidos no mesmo dia dos testes. Já em diferentes dias, a taxa de erro foi de 5,58% e 76,9% respectivamente.

Ye et al. [38] em 2010 apresentou um estudo sobre a aplicabilidade do sinal *ECG* para a identificação pessoal, através de um método para extracção de características morfológicas que oferecem informação suficiente para descriminar um grupo de indivíduos. Foram usadas três bases de dados com sinais *ECG* adquiridos através de dois eléctrodos, criadas com longos intervalos de tempo entre si. Sabe-se ainda que alguns indivíduos dessa base de dados possuem arritmia cardíaca. A taxa de classificação foi de 99,6% para sinais *ECG* normais e também com arritmia cardíaca.

Coutinho et al. [8] em 2010 segmentou o sinal *ECG* e aplicou uma quantização uniforme a 8 bits para transformar as amostras em sequências de 256 símbolos. A classificação baseou-se na procura do modelo em base de dados que obtivesse o menor tamanho para descrever a amostra de teste, calculada através do algoritmo de *Ziv-Merhav*. Neste estudo foi referido uma precisão de classificação de 100% para um conjunto de 19 indivíduos com variação do estado emocional.

1.5 Problemas ou lacunas

Uma das características fundamentais dos sistemas biométricos é a capacidade de não intrusão na aquisição dos sinais biométricos. Os sistemas biométricos baseados no sinal *ECG*, tendem cada vez mais a seguir este objectivo, diminuindo o número de eléctrodos utilizados e até mesmo o local onde estes são aplicados (junto ao peito, mãos, etc.).

Para cumprir este objectivo e atingir bons resultados é necessário realizar estudos, aplicar novas técnicas e algoritmos de forma a conceber sistemas com um maior desempenho.
Nos estudos as técnicas propostas são normalmente postas em prática recorrendo a aplicações que testam e avaliam o sistema de forma sistemática. O tempo e esforço despendido no desenvolvimento dessas aplicações pode ser relevante e necessário para aplicar noutras tarefas consideradas de maior relevância para o objectivo que se pretende atingir. Por outro lado nem sempre são obtidas as melhores métricas para a avaliação correcta do sistema, bem como a falta de uniformização dos resultados, podem inviabilizar a comparação destes com os de outros estudos. A fonte de dados biométricos aplicada nos testes, varia de estudo para estudo e os testes em ambiente real são pouco frequentes.

1.6 Objectivo do trabalho

Este trabalho tem como objectivo o desenvolvimento de uma *framework* que possibilite a criação de sistemas biométricos baseados no sinal *ECG*. Esses sistemas biométricos suportados na *framework* implementam a autenticação e identificação pessoal, baseada em duas abordagens distintas: *fiducial* e *non-fiducial*.
Pretende-se que a *framework* possa ser configurada, nas várias abordagens de base *fiducial* e *non-fiducial*, e os sistemas biométricos implementados usados para testes em ambiente real, servindo desta forma para colmatar a lacuna entre os ambientes de desenvolvimento de investigação e de teste em ambiente real.
A *framework* será constituída por um conjunto de funcionalidades base, necessárias para o funcionamento de qualquer sistema biométrico. As funcionalidades específicas de suporte a cada abordagem são realizadas por componentes externos e integrados pela *framework* através de configuração. Esta permitirá não só configurar componentes isolados mas também, grupos de componentes interligados de forma mais complexa, como por exemplo a combinação de duas abordagens distintas.
Foram desenvolvidos os componentes necessários à implementação de algoritmos representativos das abordagens *fiducial* e *non-fiducial*, ficando em aberto a implementação das funcionalidades específicas, inerentes a outras soluções de ambas as abordagens.

Neste trabalho, será também proposta uma nova solução de implementação de um classificador, baseado na abordagem *non-fiducial*, para classificação de sinais *ECG* adquiridos nas mãos. Trata-se, da utilização do método *Ziv-Merhav (ZMM),* baseada numa versão modificada do algoritmo Lempel-Ziv (*LZ78*).

1.7 Organização

O presente documento tem a seguinte estrutura: O capítulo 2 descreve a arquitectura de um sistema biométrico para autenticação e identificação pessoal baseado no sinal *ECG*, recorrendo a duas abordagens distintas. É proposta uma nova implementação para a abordagem *non-fiducial.* No capítulo 3 apresenta-se uma proposta de *framework* para suporte à criação de um sistema biométrico baseado numa das abordagens, através da criação e configuração de componentes externos. No capítulo 4 são apresentados os resultados dos testes de avaliação da *framework*. O capítulo 5 apresenta as conclusões deste trabalho. O capítulo 6 descreve futuros trabalhos.

2 Arquitectura do sistema biométrico

Um sistema biométrico pode operar em dois modos: autenticação (*authentication*) e identificação (*identification*) pessoal. Estas são as duas operações mais utilizadas e visíveis para o utilizador. No entanto, existem outras para suporte à gestão e controlo do sistema como sejam o registo (*enrollment*) e a avaliação (*evaluation*) de desempenho do sistema biométrico, conforme se propõe neste trabalho e é representado na Figura 9.
O registo permite a introdução de novos utilizadores e respectivos dados biométricos (sinal *ECG*) no sistema. Esta operação é realizada com a colaboração do utilizador e com a supervisão do gestor do sistema, de forma a garantir nomeadamente a autenticidade do sinal *ECG*.
A avaliação permite, ao gestor do sistema, obter métricas relativas ao desempenho do sistema.

Figura 9 – Operações disponíveis no sistema biométrico desenvolvido.

As operações do sistema podem-se resumir da seguinte forma:

- **Identificação**: atribuição da identidade dado um sinal biométrico; o sistema recebe o sinal *ECG* e devolve a identidade (*ID*) estimada para o utilizador.
- **Autenticação**: verificar se a identidade que o utilizador reclama corresponde à identidade estimada dado um sinal biométrico; o sistema recebe o sinal *ECG* e a identidade (*ID*) do utilizador. Como resultado, devolve uma resposta do tipo verdadeiro ou falso.
- **Registo**: adicionar no sistema novos utilizadores com o respectivo sinal biométrico; O sistema recebe o sinal *ECG* e a identidade (*ID*) do utilizador e como resultado, este indica se o processo teve ou não sucesso.
- **Avaliação**: quantificar o desempenho dos sistema com base num conjunto de métricas a definir.

A arquitectura do sistema biométrico baseia-se num conjunto de blocos, seguindo a estrutura clássica de um sistema de reconhecimento de padrões: aquisição, pré-processamento, extracção de características e classificador. Para além dos blocos que eventualmente fazem parte de uma estrutura clássica e descrita na Figura 4, esta arquitectura acrescenta ainda o bloco de aquisição responsável por recolha do sinal biométrico.
De acordo com a abordagem adoptada, *fiducial* ou *non-fiducial*, esses blocos podem ter diferentes implementações como se apresenta na Figura 10.

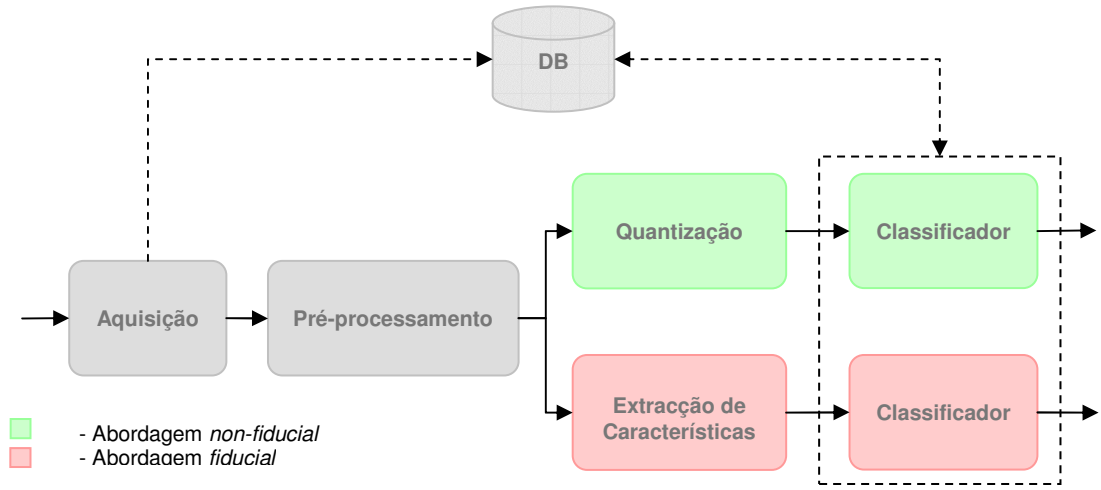


Figura 10 – Arquitectura do sistema biométrico para autenticação e identificação, através da abordagem *fiducial* e *non-fiducial*.

Esta arquitectura do sistema biométrico, de acordo com as diversas operações, recebe o sinal *ECG* na entrada do bloco de aquisição e processa-o nas diversas etapas representadas por cada bloco. No final, o bloco de classificação devolve um resultado de acordo com a operação envolvida. Durante este processo determinada informação (dados RAW e os modelos) é guardada na base de dados durante a operação de registo para na operação de identificação, autenticação ou avaliação ser consultada.

Neste trabalho serão desenvolvidas duas implementações, uma por cada abordagem. Os blocos de aquisição e pré-processamento são comuns para ambas as abordagens. Já a extracção de características, que se resume apenas à quantização na abordagem *non-fiducial* por exemplo, e por consequência o classificador, terão implementações distintas da abordagem *fiducial*.
As próximas secções descrevem, com mais detalhe, cada um destes blocos para as diversas abordagens.

2.1 Aquisição

A aquisição é o bloco responsável pela recolha do sinal *ECG*. Este pode ser obtido, para além dos pontos convencionais pouco práticos apresentados na Figura 8, junto ao peito, nas mãos ou em qualquer parte do copo humano onde seja possível obter os sinais eléctricos associados aos batimentos cardíacos, suficientes para o reconhecimento do sistema biométrico. Na Figura 11 podemos ver dois exemplos de possíveis locais do copo humano onde é possível adquirir o sinal *ECG*. A imagem da esquerda utiliza dois eléctrodos para a aquisição na *V1* (eléctrodo da esquerda) e na *V2* (eléctrodo da direita) enquanto que na imagem da direita são usados quatro eléctrodos, dois em cada mão, um na palma da mão e o outro no dedo indicador.

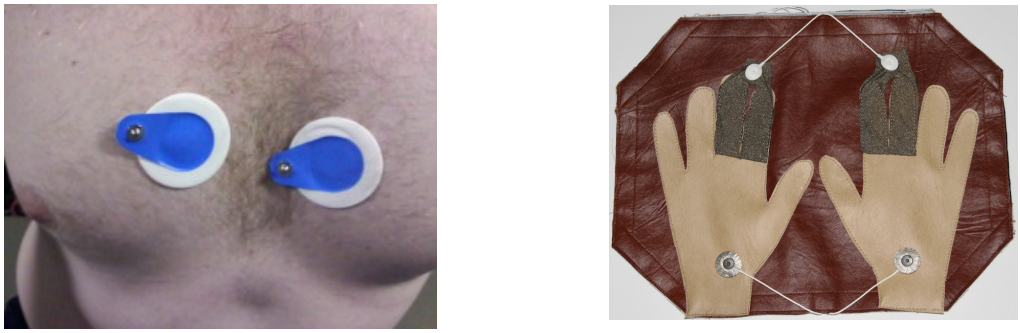


Figura 11 – Aplicação de dois tipos de *setup* para a aquisição do sinal *ECG*. Na imagem da esquerda são usados dois eléctrodos (na *V1* o eléctrodo da esquerda e na *V2* o eléctrodo da direita). Na imagem da direita é usada a palma das mãos (um eléctrodo em cada palma e um eléctrodo em cada dedo indicador).

Neste trabalho será usado um conjunto de sinais *ECG*, obtidos com dois eléctrodos um em cada palma da mão. Utilizando um *setup* idêntico ao da imagem direita da Figura 11, usando apenas os eléctrodos da palma da mão e ignorando os eléctrodos do dedo indicador.
Com este *setup* o sinal *ECG* é adquirido pelos sensores e convertido de analógico para digital (12 Bits de resolução e amostragem de 1000*Hz*) através do dispositivo *BioPLUX* [9]. Na saída do *BioPLUX* o sinal *ECG* obtido é um sinal *RAW* sem qualquer processamento. Esse processamento será realizado posteriormente pelo próximo bloco (pré-processamento), onde este sinal está disponível à entrada. Este é guardado na base de dados (ver Figura 10), permitindo que implementações futuras (de pré-processamento) possam manipular este sinal *ECG RAW*.
O sinal *ECG RAW* à saída do *BioPLUX* é enviado, através de uma interface *Bluetooth*, para o próximo bloco de pré-processamento existente no dispositivo móvel, como mostra a Figura 12.

Figura 12 – Aquisição do sinal biométrico *ECG* nas palmas das mãos através do dispositivo *BioPLUX.*

Como o sistema biométrico requer alguma capacidade de processamento, este é normalmente realizado de forma distribuída num servidor remoto. No entanto a recolha do sinal *ECG* deve poder ser realizada em qualquer local. Assim, para que o sinal gerado pelo *BioPLUX* possa chegar a este *PC*/Servidor é necessário recorrer à Internet através de um outro *PC*, *PDA* ou *Smartphone*. Este apenas receberá o sinal *Bluetooth* do *BioPLUX* e envia para o sistema biométrico num *PC/Servidor* situado em qualquer local da Internet, com mostra a Figura 13.

Figura 13 – Envio do sinal *ECG* pelo *BioPLUX* para um dispositivo móvel que redirecciona para o sistema biométrico algures na Internet.

Este tipo de comunicações, com dados biométricos considerados sensíveis, deve ser protegido através de formas de comunicação seguras.

Nos últimos tempos têm surgido diversos dispositivos no mercado que permitem a aquisição do sinal *ECG* em tempo real, tendo estes, diversas aplicações práticas como por exemplo a visualização do sinal de electrocardiograma.

A Figura 14 apresenta um dispositivo móvel (*IPhone4*[1]) capaz de adquirir, processar e mostrar o sinal *ECG* em tempo real. Foi acoplando a este dispositivo uma capa com dois eléctrodos que detectam o impulso eléctrico, gerado pelo sistema cardiovascular. Este sinal é enviado posteriormente para o dispositivo móvel que através de uma aplicação, reproduz o sinal *ECG* no ecrã.

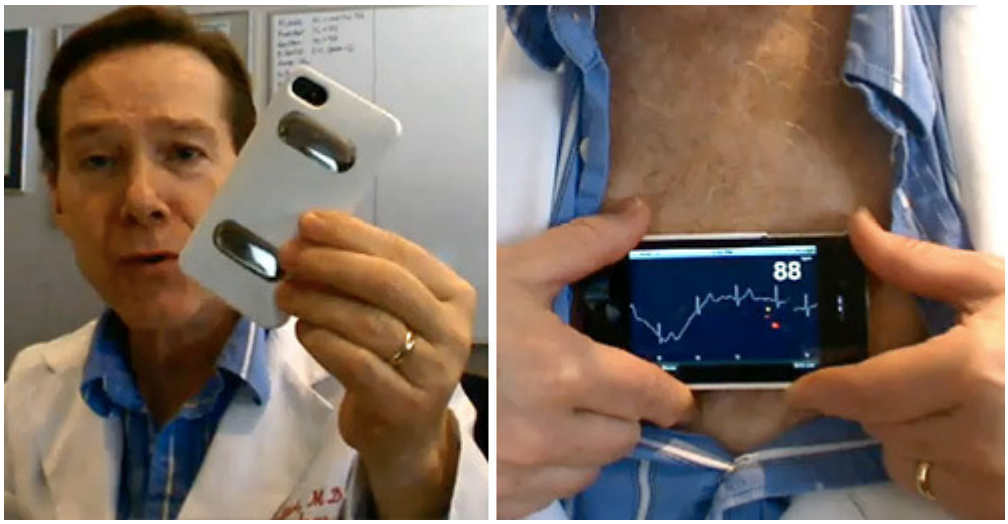


Figura 14 – Aquisição[2] do sinal *ECG* com o *IPhone4*, através de dois eléctrodos acoplados na parte de trás do dispositivo.

A tendência passa pela redução dos dispositivos para a aquisição dos sinais *ECG* e a integração destes em equipamentos móveis, tais como o *Smartphone* e assim um conjunto de aplicações poderão surgir, como por exemplo, a identificação pessoal.

2.2 Pré-processamento

O bloco de pré-processamento elimina o ruído (banda de interesse) e segmenta o sinal *ECG* em conjuntos de amostras correspondentes a ciclos cardíacos.

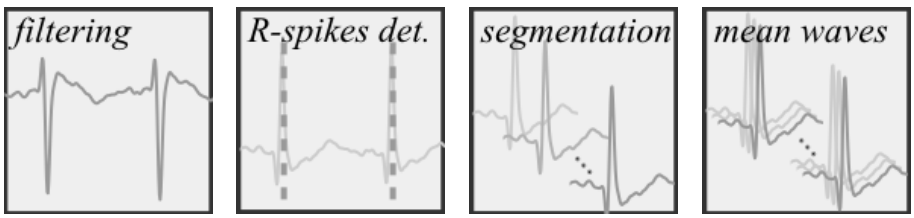


Figura 15 – Sequência típica de operações realizadas pelo bloco de pré-processamento.

[1] Modelo do *Smartphone* desenvolvido pela empresa *Apple* – http://www.apple.com/iphone/
[2] Imagens obtidas em http://alivecor.com/ (acedido em Setembro de 2011).

Tipicamente o sinal *ECG* possui um espectro de frequência entre 2 a 30 *Hz*, fora desse espectro, qualquer frequência é considerada ruído e deverá ser eliminada.
Para remover este ruído é utilizado um filtro passa-banda do tipo *Infinit Inpulse Response* (*IIR*) *Butterworth* [12], em dois passos. Primeiro, para remover as frequências acima de 30 *Hz* é aplicado um filtro passa baixo de 4ª ordem. Depois, para remover as frequências abaixo de 2 *Hz* é aplicado um filtro passa alto de 2ª ordem. Como forma de correcção de desvio e normalização do sinal é aplicado um outro filtro.

O próximo passo é a segmentação do sinal *ECG*, com a detecção dos picos R, de forma a obter um conjunto de segmentos *ECG*, para tal, foi utilizado uma derivação do algoritmo *Multiplication of the Backward Distance* (*MOBD*) usado em [13], [14] e [15].

2.2.1 Detecção dos picos R

Nesta secção faz-se a descrição de um método para detecção dos picos *R*, adaptado do algoritmo de *Engelese and Zeelenberg* [22], tendo como base o desenho de "*Algorithms Based On Digital Filters*" usado no estudo em [17].

O princípio deste algoritmo é a utilização de diferenças de derivadas para mais facilmente possibilitar a detecção do pico *R*. Assim é utilizada a Equação 1, sendo *X(n)* a derivada do sinal original.

$$Y0(n) = X(n) - X(n-4) \qquad 4 < n < 8191.$$

Equação 1 – Diferença de dois pontos da derivada do sinal original.

O sinal resultante, passará agora por um filtro digital passa baixo, descrito na Equação 2:

$$Y1(n) = Y0(n) + 4Y0(n-1) + 6Y0(n-2) + 4Y0(n-3) + Y0(n-4).$$

Equação 2 – Filtro digital passa baixo.

São escolhidos dois valores de referência (*threshold*), iguais em amplitude mas opostos em sinal. Do sinal resultante do filtro é procurado o ponto com a maior amplitude, no qual é usado como valor de referência positivo (Equação 3). Este ponto de referência inicia uma região de pesquisa com dimensão aproximada de 160 *ms*.
Cada possível valor de referência, pode ser classificado como uma mudança de fase, um possível candidato para o complexo *QRS* ou simplesmente como ruído.

$$\text{If } Y1(i) > 21.0, \text{ then search region onset} = i.$$

Equação 3 – Detecção da zona de pesquisa.

Caso não existam mais pontos de referência que passem pela zona de pesquisa (160 *ms*), então a ocorrência é considerada uma mudança de fase, caso contrário a três condições testadas (Equação 4).

$$\begin{aligned}
&\text{Condition 1: If } Y1(i+j) < -21.0 \qquad 0 < j < 40\\
&\text{Condition 2: If } Y1(i+j) < -21.0 \qquad 0 < j < 40,\\
&\qquad\qquad \text{and}\\
&\qquad\qquad Y1(1+k) > 21.0 \qquad j < k < 40\\
&\text{Condition 3: If } Y1(i+j) < -21.0 \qquad 0 < j < 40,\\
&\qquad\qquad \text{and}\\
&\qquad\qquad Y1(i+k) > 21.0 \qquad j < k < 40\\
&\qquad\qquad \text{and}\\
&\qquad\qquad Y1(i+1) < -21.0 \qquad k < 1 < 40.
\end{aligned}$$

Equação 4 – Condições para a detecção de um candidato *QRS*.

Se alguma das anteriores condições se verificar, então a ocorrência é classificado como um candidato *QRS*, No caso de existirem outros pontos de referência então é considerado ruído.

2.3 Extracção de características

Este bloco difere de acordo com a abordagem adoptada. Na abordagem *fiducial*, são extraídas características de cada conjunto de amostras (segmentos) do sinal *ECG* [4]. Enquanto que na abordagem *non-fiducial*, o conjunto de amostras do sinal *ECG* é analisado como um todo, por exemplo através da transformada para o domínio da frequência [6] [7] [29] ou através da normalização e redução de resolução (quantização a 8 bits com 256 símbolos) [8].
Em qualquer das abordagens, os dados produzidos neste bloco são guardados na base de dados como modelos, durante o processo de registo. E posteriormente são usados durante as operações de identificação e autenticação pessoal.
Estas duas abordagens são descritas com maior detalhe nas próximas duas sessões.

2.3.1 Abordagem *fiducial*

O bloco de extracção de características, na abordagem *fiducial*, extrai um conjunto de características em cada segmento do sinal *ECG*. Essas características devem ser, o mais possível, distintas entre todos os

modelos existentes na base de dados. Desta forma o desempenho do sistema será maximizado, minimizando as falhas de classificação.

Nem sempre a utilização de um grande número de características permite obter melhores resultados. Em determinados casos esse número torna-se um problema e a solução passa pela selecção de novas características. É o caso do estudo [4] de *Biel et al.* que obteve melhores resultados com a selecção de 10 (Figura 16) características num total de 30.

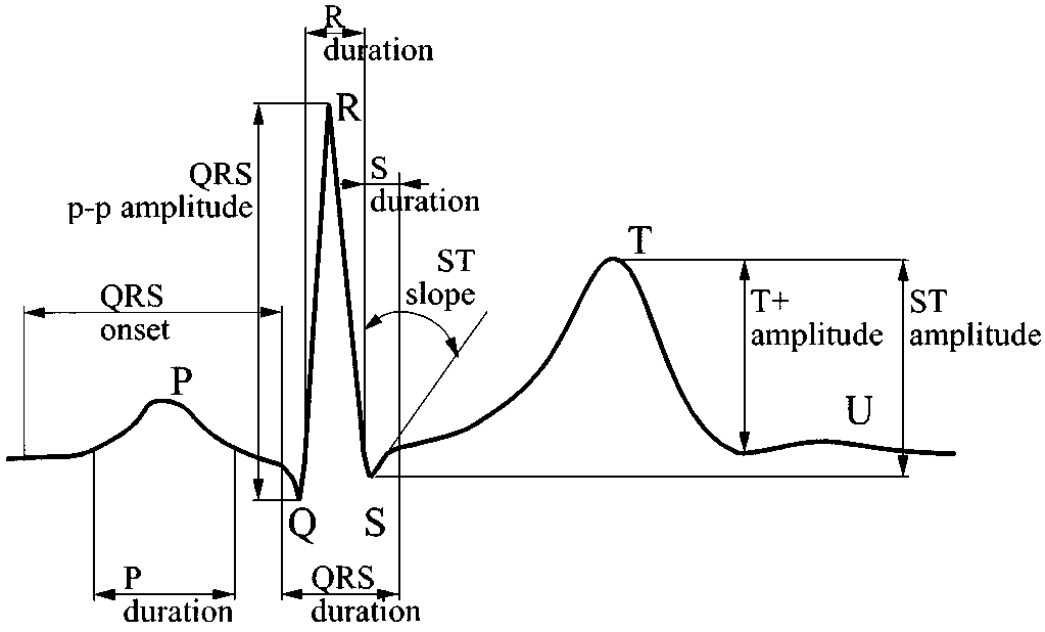


Figura 16 – Extracção[1] de características utilizadas no estudo em [4].

Uma grande quantidade de características pode levar a um conjunto de problemas, nomeadamente na classificação, pois um número elevado de dimensões dificulta a decisão do classificador. A este problema dá-se o nome de "*curse of dimensionality*".

2.3.2 Abordagem *non-fiducial*

Na abordagem non-fiducial não são extraídas características, baseadas em pontos particulares (*fiducia*) das amostras do sinal *ECG*, mas utiliza-se todo o sinal como uma única característica.

Recentemente foi proposta [8] uma nova abordagem baseada na comparação de sequências de símbolos, utilizando uma variante do método *Ziv-Merhav* (*ZMM*) para a estimação da entropia relativa, que já foi usada como ferramenta para a classificação de textos. Desta forma é possível obter uma medida da semelhança entre dois textos e utilizá-la num classificador. Para isso o sinal *ECG* foi transformado numa sequência de símbolos, através de um processo conhecido como

[1] Imagem obtida do estudo de *Biel et al.* [4]

quantização [11]. Foi utilizada a quantização escalar uniforme de 8 *bits*, apresentada de forma grosseira na Figura 17, e assim do *ECG* resulta uma sequência de símbolos de um conjunto de 256 símbolos possíveis.

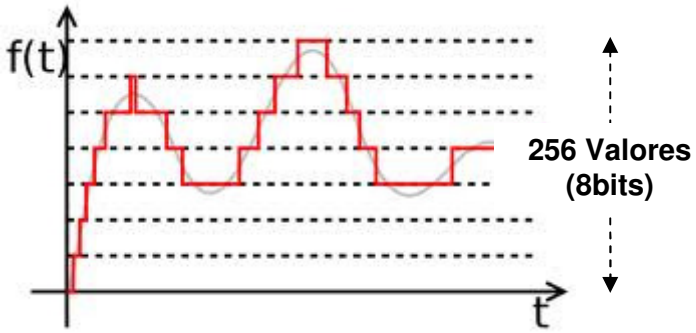


Figura 17 – Exemplo[1] de uma quantização escalar uniforme.

2.4 Classificador

O classificador tem como principal objectivo classificar as amostras de teste, indicando a que modelo pertence. A forma como é realizada, baseia-se na semelhança entre dois conjuntos de amostras (amostras de teste e amostras do modelo). A distância entre conjuntos, pode ser usada como medida de semelhança. Esta é tanto maior quanto menor for o valor da distância que as separa. A distância é um valor numérico e o seu cálculo difere com a abordagem adoptada. Nas duas próximas secções, são apresentadas duas formas distintas para o cálculo da distância: abordagem *fiducial* e *non-fiducial*.

2.4.1 Abordagem *fiducial*

A classificação na abordagem *fiducial* baseia-se na distância entre as características das amostras de teste e as características das amostras do modelo.
O cálculo dessa distância é baseado no método do vizinho mais próximo *K-NN* (K=1) e utiliza como métrica, a distância euclidiana [2] e apresentado na Equação 5.

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}.$$

Equação 5 – Cálculo da distância euclidiana entre o ponto P e Q com n dimensões.

[1] Imagem obtida em http://pt.wikipedia.org/wiki/Quantiza%C3%A7%C3%A3o (acedido em Setembro de 2011) e posteriormente alterada.

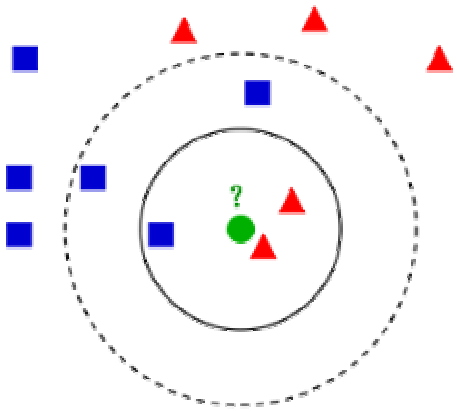


Figura 18 – Exemplo[1] do cálculo do vizinho mais próximo (*K-NN*).

Na Figura 18 vemos o ponto de teste (circulo verde) e a procura do modelo mais próximo (K=1) a este (quadrados azuis ou triângulos vermelhos).

2.4.2 Abordagem *non-fiducial*

A classificação na abordagem *non-fiducial* é baseada na semelhança entre dois conjuntos de símbolos (256 valores): as amostras de teste e as amostras do modelo. Ao contrário da abordagem *fiducial*, em que apenas são comparadas características das amostras do sinal *ECG*, na abordagem *non-fiducial* é comparada toda a amostra do sinal *ECG* (após quantização) obtendo um valor de semelhança (distância).

Em [8] é proposto um estudo que aborda a utilização do método *Ziv-Merhav* (*ZMM*) [60] para a estimativa da entropia relativa através de sequencias de símbolos, como uma ferramenta para a classificação de texto. Nesse estudo e representado na Figura 19, é descrita uma implementação do *ZMM,* baseada numa versão modificada do algoritmo *Lempel-Ziv* (*LZ77*) [61].

[1] Imagem obtida em http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm (acedido em Setembro de 2011)

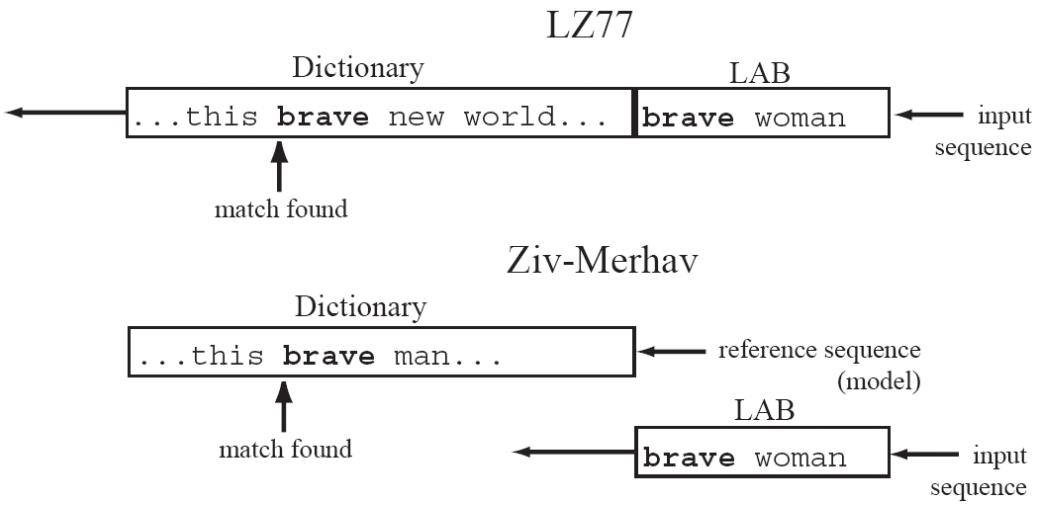


Figura 19 – Imagem[1] do Algoritmo *LZ77* original que usa uma *sliding window* sobre sequência de entrada para actualizar o dicionário (em cima). E em baixo o algoritmo *ZMM* usado no estudo em [8].

2.4.3 Nova abordagem *non-fiducial*

A nova abordagem apresentada neste trabalho, para a classificação non-fiducial, baseia-se no estudo realizado em [8]. Nesta nova abordagem, propomos implementar uma ferramenta para a classificação de texto baseada numa versão modificada do algoritmo *Lempel-Ziv* (*LZ78*) [62].

Este algoritmo constrói um dicionário com os símbolos alfabéticos resultantes da quantização para as amostras do modelo, fazendo passar pelo compressor. Terminado este passo, o dicionário é "congelado" e não mais é alterado. Finalmente as amostras de teste passam pelo compressor, sendo contabilizadas as referências utilizadas para o dicionário. Esse valor de contagem de referências, representa a distância entre as amostras de teste e as amostras do modelo. Esta distância é tanto menor quanto mais semelhantes são os textos (conjunto de símbolos alfanuméricos) do teste e do modelo.

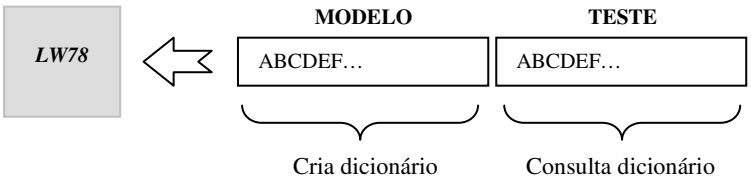


Figura 20 – Representação do cálculo da distância entre as amostras de teste e modelo, baseada no algoritmo de compressão *LZ78*.

[1] Imagem obtida do estudo de Coutinho, D.P. [8].

O *LZ78* é um dos algoritmos de compressão de dados desenvolvidos por *Abraham Lempel* e *Jacob Ziv* em 1978 [62]. Baseia-se na construção de um dicionário de caracteres. Cada sequência de caracteres que não exista no dicionário é lá colocada, associando-lhe um código, que será usado na codificação dessa sequência. Sempre que qualquer sequência exista no dicionário a sua codificação é substituída pelo código aí correspondente.
Existem diversas implementações deste algoritmo, sendo uma delas o *LZW*, que se tornou famosa pela facilidade de implementação. Neste trabalho é usada uma implementação do *LZW* alterada, permitindo que o dicionário seja apenas construído durante a compressão das amostras do modelo e por outro lado seja contabilizado o número de referências ao dicionário durante as amostras de teste.

2.5 Avaliação de desempenho

A avaliação de desempenho de um sistema biométrico permite determinar o grau de desempenho desse sistema, para as operações de autenticação e identificação pessoal. O grau de desempenho é um dos principais factores que define a qualidade e aceitação de um sistema biométrico [10].

As métricas, mais comuns, usadas para esta avaliação são:

- *False Acceptance Rate* (*FAR*): taxas de aceitação de entidades que não são quem realmente dizem ser. Falsas aceitações.
- *False Rejection Rate* (*FRR*): taxa de rejeição de entidades que são quem realmente dizem ser. Falsas rejeições.
- *Equal Error Rate* (*EER*): taxa de erro igual. Corresponde ao valor no qual o *FAR* e o *FRR* são iguais.
- *IdErro:* taxa de erro na identificação.

A matriz de confusão, na Figura 21, serve como base para a obtenção de um conjunto de métricas, sendo esta constituída por duas linhas e duas colunas que registam: as ocorrências verdadeiras que o sistema aceitou – correcto (verdadeiras aceites – *VA*), as ocorrências verdadeiras mas que o sistema não aceitou – erro (Falsas rejeitadas – *FR*), ocorrências falsas que o sistema não aceitou – correcto (Verdadeiras rejeitadas – *VR*) e as ocorrências falsas que o sistema aceitou – erro (Falsas aceites – *FA*).

		Análise	
		Verdadeiros	**Falsos**
Teste	**Aceites**	**VA** (certo – correctamente aceite)	**FA** (erro – falsamente aceite)
	Rejeitados	**FR** (erro – falsamente rejeitado)	**VR** (certo – correctamente rejeitado)

Figura 21 – Matriz de confusão.

A taxa de erro *FAR* é obtida, somando as ocorrências de falso aceites (*FA*) sobre o número total de casos analisados. Por outro lado a taxa de erro *FRR* é obtida, somando as ocorrências verdadeiras rejeitadas (*FR*) sobre o número de total de casos analisados.

Estes valores são calculados para um valor de *threshold* que normalmente variar entre zero e um. Deste modo obtemos a curva da taxa de erro *FAR* e *FRR*, tal como apresenta a Figura 22.

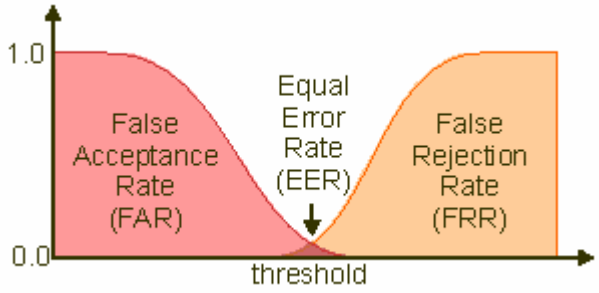


Figura 22 – Imagem[1] da curva típica das taxas de erro *FAR* e *FRR*.

A taxa de erro *EER* resulta do valor da intercepção das duas taxas de erro *FAR* e *FRR*. Esse ponto de intercepção dá-se o nome de *EER threshold*, e é o limiar de decisão para se classificar e contabilizar as ocorrências (*VA*, *FA*, *FR* e *VR*).

As métricas *FAR* e *FRR* podem também ser conhecidas como *False Match Rate* (*FMR*) e *False Non-Match Rate* (*FNMR*) respectivamente.

[1] Imagem obtida em http://support.bioid.com/sdk/docs/About_EER.htm (acedido em Setembro de 2011).

A curva *ROC* (*Relative Operation Caracteristic*) é uma referência bastante utilizada na avaliação de desempenho de sistemas biométricos, na medida em que permite analisar o sistema de identificação face à sua sensibilidade. É utilizada como uma ferramenta para medir e especificar problemas de desempenho num sistema baseado na decisão.

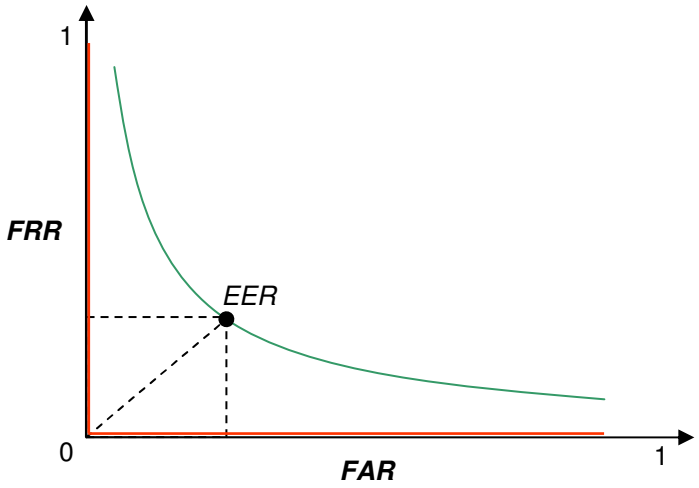


Figura 23 – Curvas *ROC.*

A curva verde, da Figura 23, representa uma típica curva *ROC* em que o *EER* é o ponto de intercepção do *FAR* com o *FRR*.
Um classificador é tanto mais preciso quanto mais a curva *ROC* for próxima do zero das ordenadas e das abcissas (curva a vermelho).

3 *Framework* para suporte de um sistema biométrico

O objectivo da *framework* é criar uma plataforma de desenvolvimento de *software* que permita a implementação de sistemas biométricos de identificação e autenticação pessoal. Com esta plataforma pretendemos disponibilizar um conjunto de funcionalidades base, comuns aos sistemas biométricos e também mecanismos que permitam facilmente integrar novos componentes externos de *software*, que possibilitem o teste em ambiente real de diferentes abordagens de reconhecimento.

O *BioAPI* consórcio foi fundado com o objectivo de desenvolver uma *API* biométrica que trouxesse independência da plataforma e dos dispositivos para programadores de aplicações biométricas e para fornecedores de serviços biométricos (*BSP*). O consórcio é constituído por um grupo de mais de 120 empresas e organizações que têm um interesse comum na promoção do crescimento do mercado da biometria. Este consórcio desenvolveu uma especificação e referências de implementação para uma *API* biométrica *standard* (*BioAPI 1.1 - ANSI/INCITS 358-2002*) [50] que fosse compatível com um amplo conjunto de programas aplicativos biométricos e tecnologias biométricas [51]. O standard *BioAPI 2.0* é uma versão completamente nova, criada pelo comité internacional de *standards* para biometria dentro da *ISO* (*ISO/IEC JTC1 SC37*). Esta versão internacional (*ISO/IEC 19784-1:2005*) [52] introduz várias melhorias sobre a versão *ANSI* da *BioAPI 1.1*. Existem várias implementações da especificação BioAPI, desenvolvidas em *C/C++* para plataformas *Win32* e *Linux/Unix* [53][54][55]. De forma a permitir o acesso à *BioAPI framework*, desenvolvidas em *C/C++*, através da plataforma Java foi desenvolvido um *Java Wrapper* recorrendo à *Java Native Interface* (*JNI*). Existem actualmente algumas implementações de *Java Wrappers* baseados em *JNI* [56] [57]. Foi também desenvolvido um *C# Wrapper* de acordo com o standard *BioAPI 1.1 - ANSI/INCITS 358-2002* [58].
A *Microsoft* também desenvolveu uma *API* biométrica chamada *Windows Biometric Framawork API* (*WBF*) com o objectivo de uniformizar todas as aplicações biométricas usadas nos seus sistemas operativos [59]. Esta *API* está integrada nos sistemas operativos *Windows Server 2008 R2* e *Windows 7*.
Todas estas referências foram tidas em conta no desenvolvimento da *framework* proposta neste trabalho, mas no entanto nenhuma foi adoptada.

No contexto deste projecto define-se *framework* como a infra-estrutura de *software* de base que permite a construção de sistemas complexos pela integração de componentes internos e externos à própria *framework*.
O objectivo desta infra-estrutura é facilitar a construção de sistemas biométricos, em particular sistemas baseados no sinal *ECG*. Pretende-se

que esta *framework* possa ser configurada nas principais abordagens das perspectivas *fiducial* e *non-fiducial* e que permita o seu teste em ambiente real.

A Figura 24 apresenta a relação entre a *framework* e os componentes externos que implementam as diferentes etapas necessárias para o reconhecimento biométrico através dos blocos de pré-processamento, extracção de características e classificador. O conjunto *framework* e blocos externos denomina-se por sistema biométricos.

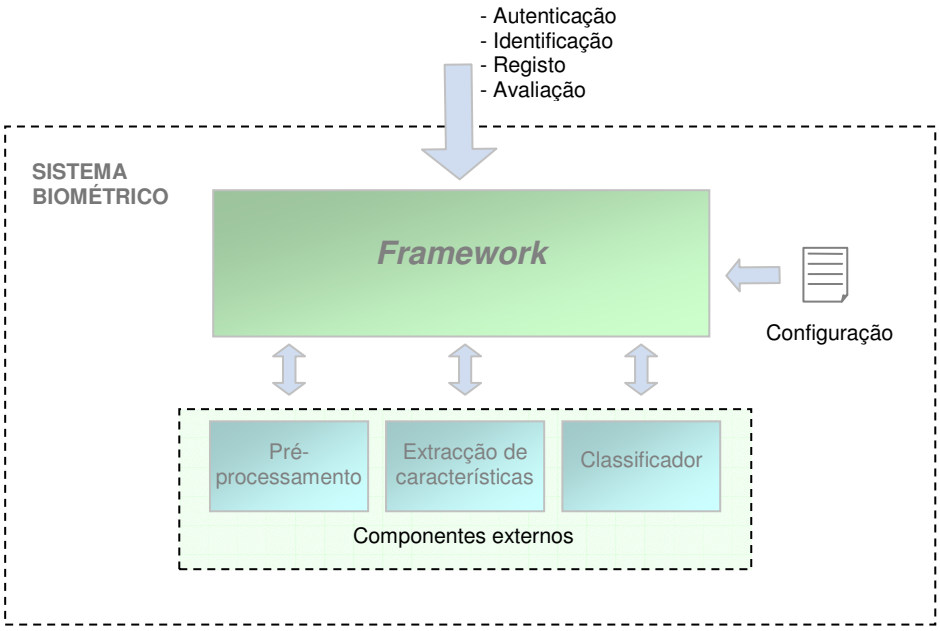


Figura 24 – Desenho do sistema biométrico baseado na *framework* e respectivos componentes externos.

Os blocos de pré-processamento, extracção de características e classificador, podem seguir as diferentes abordagens descritas na literatura.
A implementação de cada bloco (pré-processamento, extracção de características e classificador) é efectuada através da utilização de um ou mais componentes externos interligados entre si. A interligação de componentes permite definir um fluxo de informação entre cada um, mas também entre a *framework* e componentes, e vice-versa (ver exemplo da Figura 25).

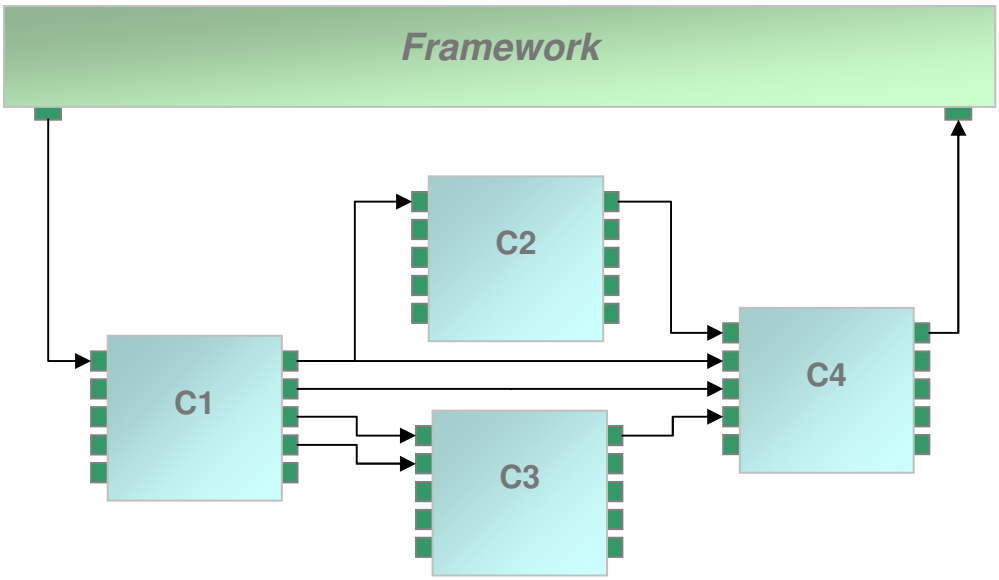


Figura 25 – Exemplo de uma possível interligação de componentes.

A função dos componentes é receber informação, processa-la, enviando o resultado desse processamento para a sua saída, passando-a para o próximo componente, sendo esse processo repetindo sucessivamente até que no final é entregue à *framework*.
A passagem de informação entre componentes é definida de forma estruturada usando a noção de configurações. Estas configurações especificam que componentes e ligações serão usados, sendo estes guardados de forma persistente num ficheiro de configurações. A disposição dos componentes (ligação entre eles) pode ser feita em série, paralela e também com uma combinação das duas. As regras e limitações definidas para os componentes e suas ligações, são apresentadas mais à frente neste documento.

A Figura 26 propõe a arquitectura do sistema biométrico desenvolvido neste projecto.

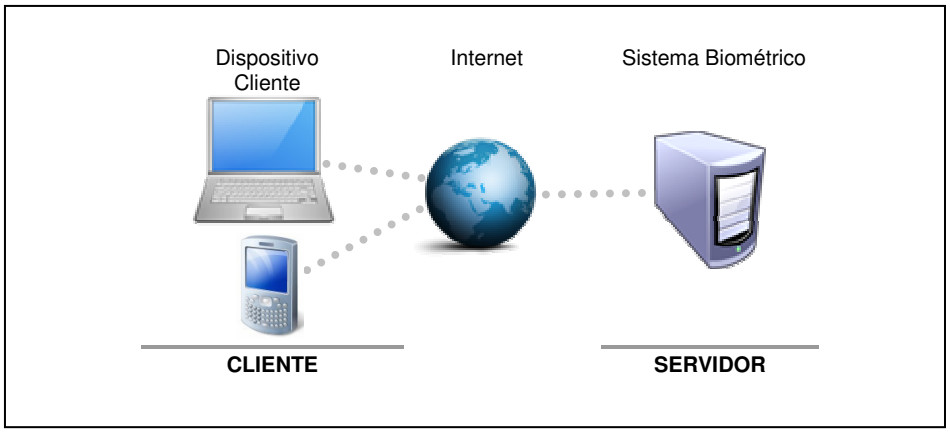


Figura 26 – Aplicação prática do sistema biométrico, com os diversos equipamentos envolvidos.

Este sistema biométrico divide-se em duas partes: cliente e servidor. O cliente é composto por uma aplicação gráfica para o utilizador efectuar pedidos de registo, identificação, autenticação e avaliação ao sistema biométrico.
O servidor é composto por um *PC*, onde está presente o sistema biométrico. Este disponibiliza um conjunto de operações (identificação, autenticação, registo e avaliação) com possibilidade de acesso remoto através de *webservices*.

A Figura 27 detalha a arquitectura proposta através de um diagrama com os respectivos blocos de *software* utilizados.

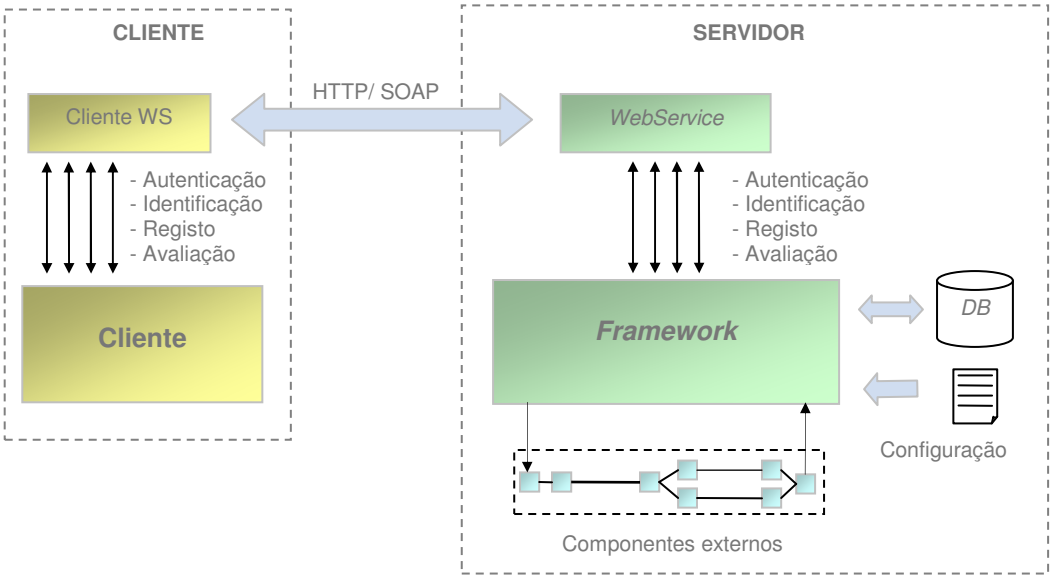


Figura 27 – Diagrama da arquitectura geral do sistema biométrico, cliente e servidor com os respectivos blocos de *software*.

O cliente é constituído por uma aplicação gráfica que interage com o utilizador, permitindo a este realizar as diversas operações disponíveis no sistema biométrico. Nesta interacção com o utilizador, a aplicação cliente recolhe informação introduzida pelo utilizador e após a execução da operação seleccionada, apresenta o respectivo resultado:

Registo: são inseridos os dados do utilizador juntamente com os dados biométricos (sinal *ECG*). Como resultado podem surgir duas mensagens possíveis, "Operação bem sucedida" ou "Falha no registo".

Autenticação: é inserida a identidade do utilizador juntamente com os seus dados biométricos. Como resposta a esta operação, podem surgir uma de duas mensagens: "Autenticação válida" ou "Autenticação inválida".

Identificação: são inseridos os dados biométricos do utilizador e como resposta é estimada a identidade a que correspondem esses

dados. Caso o sistema não consiga realizar a identificação é apresentada a seguinte frase: “Falha na identificação”.

Avaliação: esta operação não necessita de qualquer introdução de dados por parte do utilizador. No entanto após o seu pedido é apresentado ao utilizador o nome do ficheiro no qual irá ser colocado dados que servirão para a avaliação do sistema actualmente configurado. Esse ficheiro será posteriormente gerado pelo sistema biométrico e mantido num directoria específica do servidor.

As operações são executadas remotamente no servidor, sendo virtualizadas localmente no cliente pelo bloco “Cliente WS”. É utilizado um *webservice* que disponibiliza uma comunicação *SOAP* via *HTTP*. Podendo esta suportar também confidencialidade através do protocolo *HTTPS*.

O servidor é constituído pela *framework*, componentes externos, o respectivo ficheiro de configuração dos componentes e ligações, e a base de dados relacional onde é guardada a informação biométrica de cada utilizador. O *webservice* dará resposta aos pedidos realizados ao sistema biométrico de uma forma remota.

3.1 Requisitos funcionais e não funcionais

A Tabela 1 descreve os requisitos funcionais para a *framework* biométrica, os quais pretendem satisfazer as necessidades dos utilizadores deste tipo de sistema.

Identificação	Descrição
RF-1	Permitir a identificação pessoal.
RF-2	Permitir a autenticação pessoal.
RF-3	Permitir a avaliação de desempenho do sistema.
RF-4	Permitir o registo de utilizadores.
RF-5	Permitir acesso remoto ao sistema.
RF-6	Garantir a persistência da informação biométrica de cada utilizador.
RF-7	Permitir a instalação e configuração de novos componentes para suporte a novas abordagens.
RF-8	Permitir o registo de informação relevante para *log*.

Tabela 1 – Requisitos funcionais da *framework* biométrica.

Os requisitos não funcionais permitem definir as regras e limitações aplicadas ao sistema. Estas estão listadas no anexo A deste documento.

3.2 Casos de utilização

Este sistema biométrico apresenta o conjunto de operações que serão realizadas por três tipos de actores: Utilizador, Administrador e Configurador.
A Figura **28**, apresenta o diagrama com os casos de utilização e os respectivos actores envolvidos.

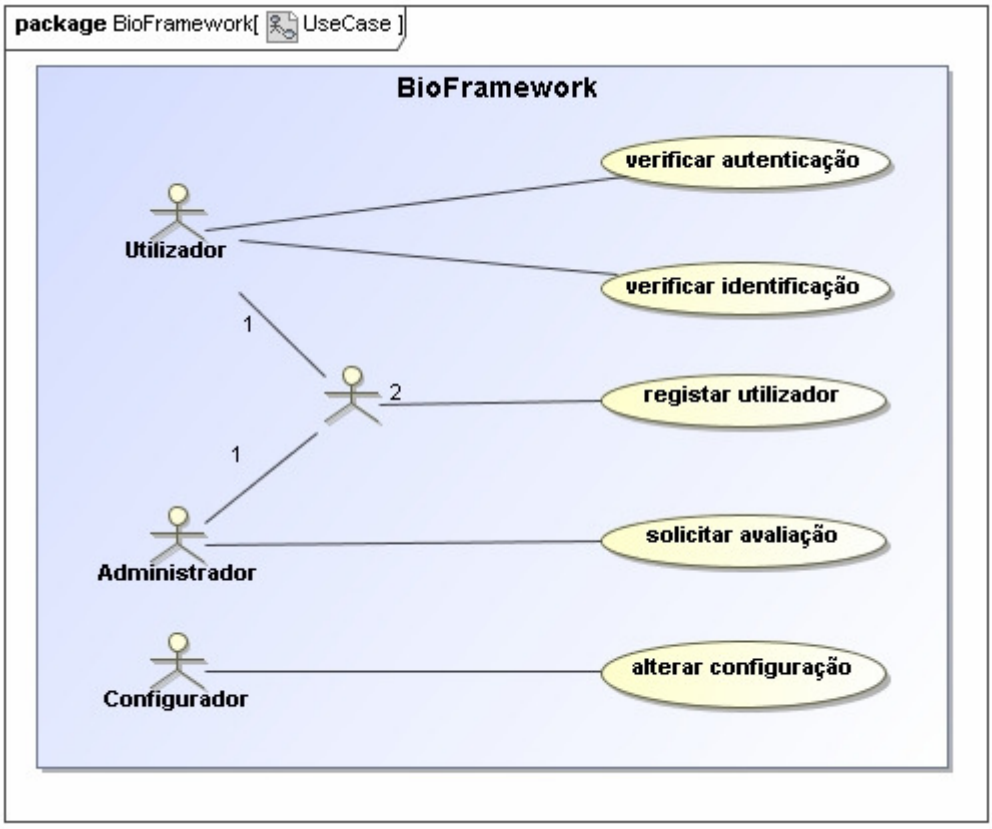


Figura 28 – Diagrama de casos de utilização.

O Utilizador efectua operações de autenticação, identificação e registo, sobre o sistema biométrico. Esta ultima é obrigatoriamente supervisionada pelo Administrador, atestando a identidade do utilizador a quando da aquisição dos sinais biométricos no registo. O Administrador pode solicitar a avaliação do sistema e o Configurador realizar alterações de configuração dos componentes externos e respectivas ligações.
A descrição dos vários casos de utilização está referida no Anexo B deste documento.

3.3 Arquitectura

A arquitectura do sistema biométrico é baseada no modelo cliente-servidor [42], suportado por *webservice* [43] (ver Figura 27). Desta forma, o conjunto de funcionalidades, disponibilizadas pela *framework*, ficam acessíveis remotamente por qualquer aplicação cliente que cumpra as

definições impostas pelo serviço (*WSDL – Web Services Description Language* [44]).

O servidor é constituído pela *framework*, que recebe os pedidos dos clientes, dirigindo-os aos diversos componentes, instalados segundo a configuração definida no ficheiro de configurações.
A base de dados permite guardar a informação de forma persistente, durante o registo e também permitir a consulta dos diversos modelos aí guardados.
O cliente terá que apresentar uma interface para os utilizadores que permita realizar todas as operações (autenticação, identificação, registo e avaliação) disponíveis no sistema biométrico.
A utilização de *webservices* possibilita uma maior flexibilidade por parte das aplicações clientes, permitindo assim o seu uso em inúmeras plataformas (*PC*, *PDA*, *SmartPhone*, etc.).

3.4 Cliente

O cliente é constituindo por uma aplicação que através de uma interface gráfica, permite realizar as diversas operações suportadas pelo sistema biométrico. As operações disponíveis nesta interface são a identificação e autenticação pessoal, o registo de novos utilizadores e a solicitação da avaliação do sistema. Esta ultima está apenas acessível ao Administrador.

Esta aplicação terá que comunicar com o sistema biométrico (localizado remotamente) através da utilização dos serviços remotos disponibilizados via *webservice* pela *framework*. Esta comunicação é feita através do protocolo *SOAP* [45] sobre o *HTTP*. Caso exista a necessidade de manter a privacidade dos dados utilizados na comunicação e de garantir a autenticidade do servidor é possível a utilização do protocolo *HTTPS* [46] em substituição do *HTTP*. No entanto neste trabalho será utilizado apenas o protocolo *HTTP* por razoes de simplicidade de implementação da versão de demonstração.
Este tipo de comunicação remota, facilita a construção de aplicações clientes para qualquer tipo de dispositivo independentemente do sistema operativo usado. Assim, se necessário, é possível que este sistema biométrico possa ser utilizado em inúmeros dispositivos móveis.

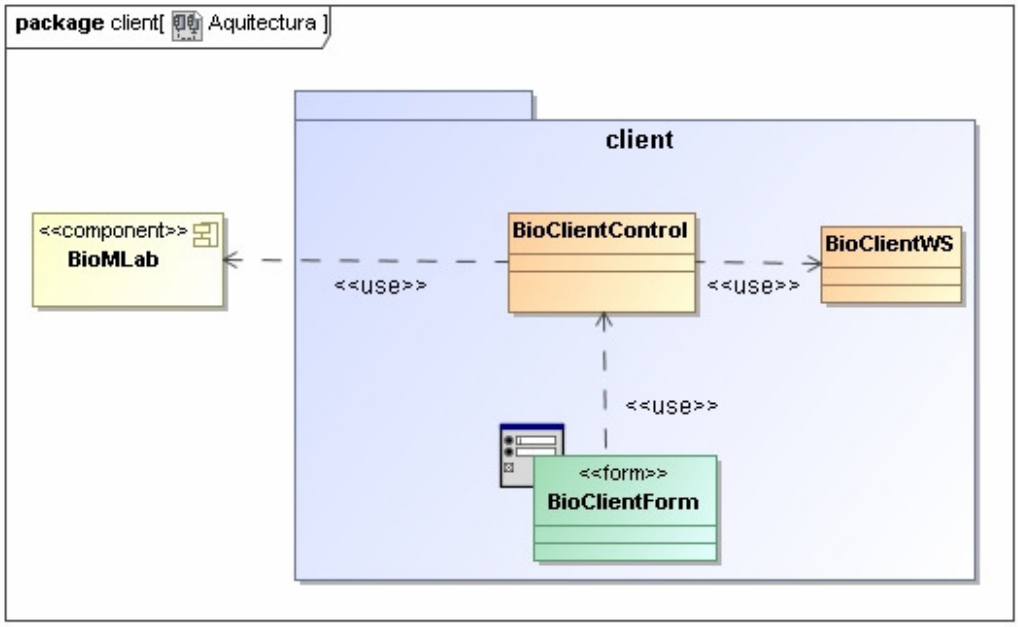


Figura 29 – Diagrama da arquitectura de classes para aplicação cliente.

A Figura 29 apresenta o diagrama da arquitectura de classes para a aplicação cliente. Esta apresenta quatro grandes partes:

- "*BioClientForm*" – interface gráfica que interage com o utilizador.
- "*BioClientControl*" – entidade que gere todos os pedidos gerados na interface do utilizador encaminhando-os para outras entidades. Desta forma interage com uma *API* de acesso ao sistema de aquisição de sinais *BioPLUX*, que adquire as amostras do sinal *ECG*, e com o "*BioClientWS*", para enviar os pedidos ao sistema biométrico.
- "*BioMLab*" – este componente permite comunicar com o dispositivo *BioPLUX* para a aquisição dos sinais *ECG*.
- "*BioClientWS*" – promove a comunicação com o sistema biométrico, através de *webservices*.

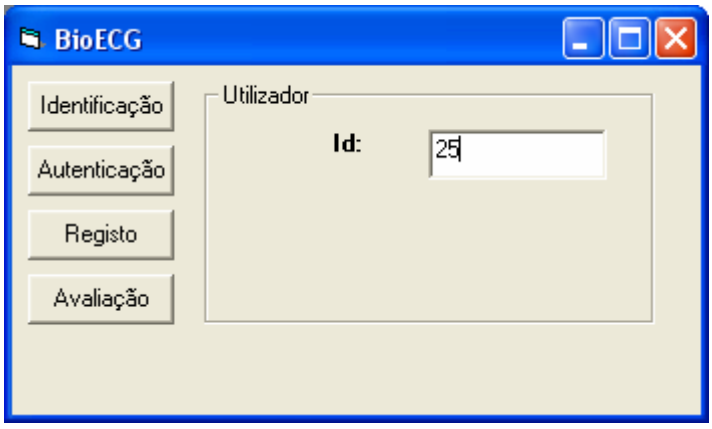


Figura 30 – Interface gráfica da aplicação cliente.

A Figura 30 apresenta a interface gráfica utilizada na aplicação cliente do sistema biométrico.

3.5 Servidor

O servidor é constituído pela *framework* do sistema biométrico e no qual fazem parte os diversos componentes externos de suporte às abordagens utilizadas, a base de dados para persistência dos dados biométricos e o *webservice* para suportar os pedidos remotos dos diversos clientes.

3.5.1 Framework

A *framework* do sistema biométrico é um módulo de *software* composto por um conjunto de funcionalidades base que permitem gerir a execução dos componentes externos, sendo responsável por:

- Gestão dos pedidos dos clientes.
- Gestão da execução e passagem de dados de cada componente externo.
- Persistência dos dados na base de dados relacional.

A solução adoptada nesta implementação, baseia-se no modelo de três camadas (3-*tier ou MVC*) [48], [49]. Composto pela camada de apresentação, camada de negócio e camada de dados, tal como podemos ver no diagrama da Figura 31, este modelo permite uma separação bem definida entre camadas, aumentando a coesão e diminuindo o acoplamento entre camadas. Desta forma cada camada saberá que papel desempenhar (separação de funcionalidades) e como interagir com as outras, seguindo uma interface bem definida (diminuição das dependências). Uma vantagem deste modelo é o baixo impacto causado, caso seja necessário a substituição de uma chamada.

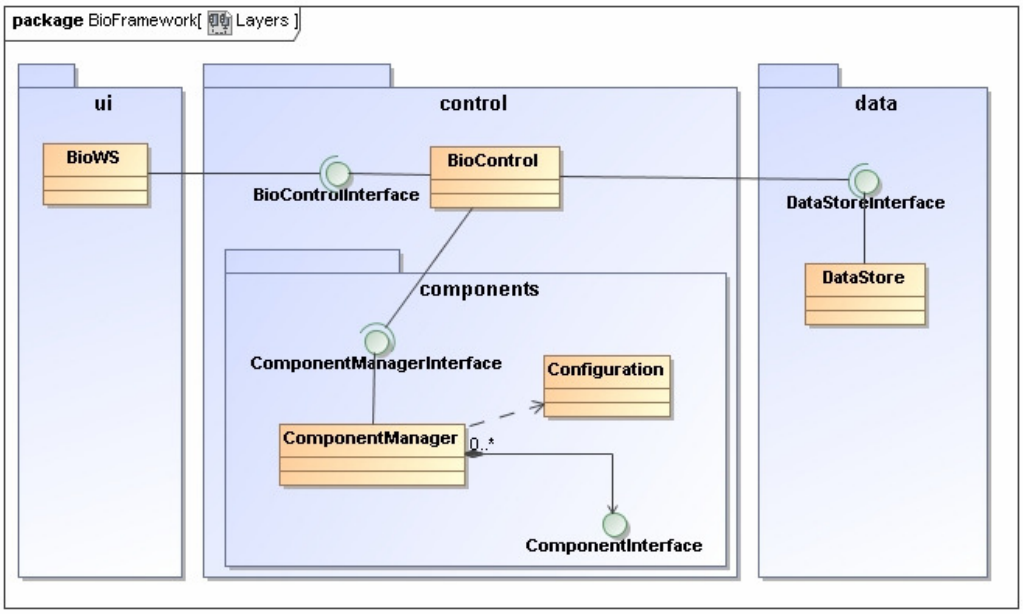


Figura 31 – Diagrama de camadas da *framework* do sistema biométrico.

A Figura 31 apresenta as diversas camadas e respectivas interfaces de acoplamento entre camada, propostas para a *framework* do sistema biométrico:

- ***ui*** – camada de apresentação, responsável pela interface com o utilizador. Sendo neste caso implementada pelo *webservice*, através da interface *BioWS,* que com a ajuda do cliente *webservice* fará a ponte entre os pedidos da aplicação cliente e o sistema biométrico.
- ***control*** – camada de negócio, que efectua o controlo do sistema biométrico. Atende os pedidos dos clientes (da camada *ui*) e encaminha-os para o gestor de componentes. Também é da responsabilidade desta camada o acesso aos dados, através da utilização da camada de dados.
- ***data*** – camada de abstracção de acesso aos dados a persistir.

A camada de controlo, presente no *package* "*control*" (ver Figura 32) é composta por uma interface que permite o acesso a esta por parte da camada de apresentação (ui). Essa interface chamada "*BioControlInterface*", disponibiliza todos os métodos necessários à realização das diversas operações. Esta é realizada pela classe "*BioControl*", que é composta por uma classe gestora de componentes externos que a ajuda na execução das diversas operações.
A classe gestora de componentes externos ("*ComponentsManager*"), possui um conjunto de instâncias de "*ComponentInterface*", carregados com base na informação obtida na configuração (acesso através da classe "*Configuration*"). A interface "*ComponentInterface*" é a interface que todos os componentes externos, a desenvolver, terão de realizar. Na próxima secção será apresentada mais informação sobre este package "*components*".

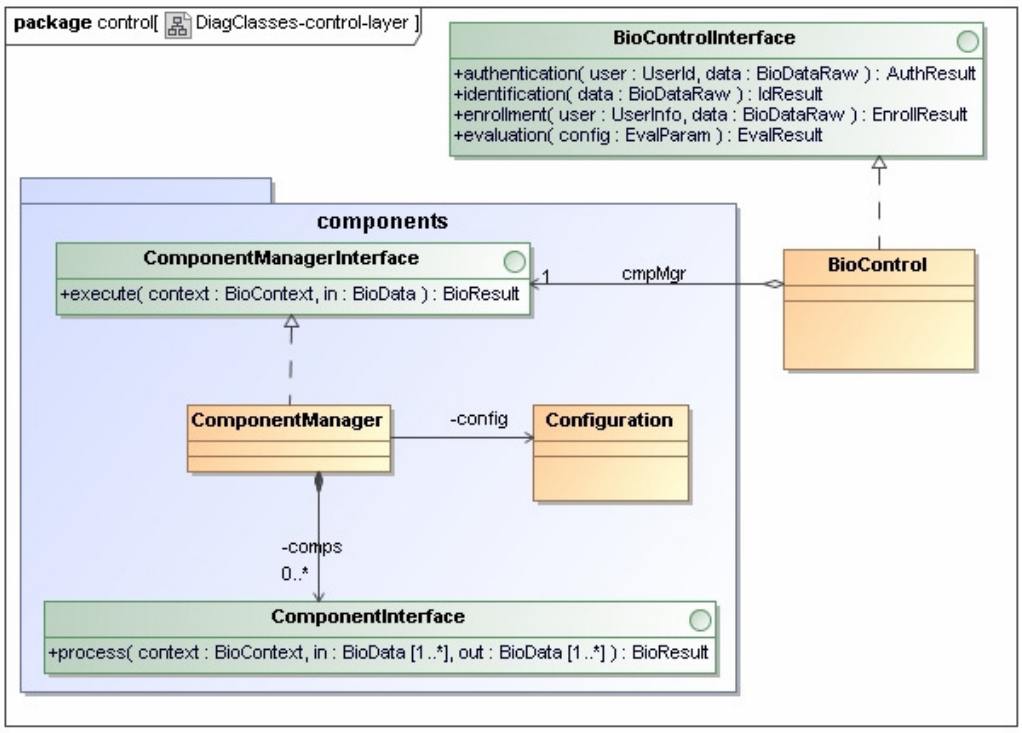


Figura 32 – Diagrama de classes da camada de controlo.

A camada de acesso a dados está presente no *package* "*data*" (ver Figura 33) e tem como objectivo garantir o acesso aos dados guardados na base de dados. Permite não só a consulta mas também a escrita de dados. É composta por uma interface ("*DataStoreInterface*") que define o conjunto de métodos acessíveis pela camada de controlo. A implementação dessa *interface* é realizada pela classe "*DataStore*".

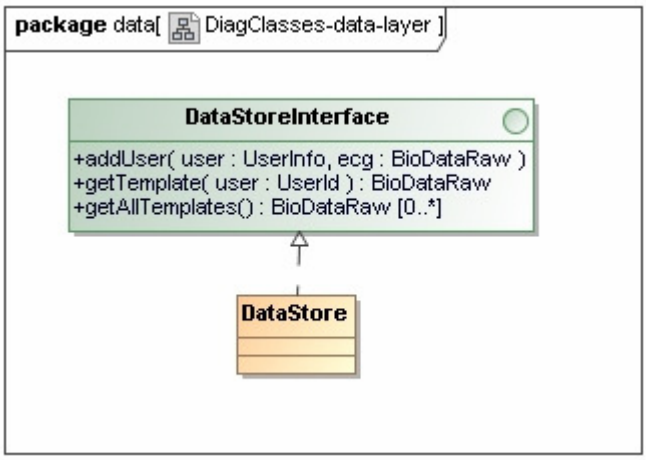


Figura 33 – Diagrama de classes da camada de acesso a dados

3.5.2 Componentes externos

Os componentes externos realizam todo o processamento envolvido num sistema clássico de reconhecimento de padrões, baseados numa abordagem em particular (*fiducial* ou *non-fiducial*). Esse processamento divide-se pelas etapas de pré-processamento, extracção de características e classificador.

Cada componente é implementado externamente, tendo necessariamente que cumprir o contrato definido pela *interface* dos componentes, a *ComponentInterface* (ver Figura 34). Após a implementação, o componente estará pronto para ser integrado na *framework,* e desempenhar o papel para o qual foi destinado para o funcionamento do sistema biométrico. Os componentes integrados pela *framework* deverão ser referidos no ficheiro de configuração. A definição desse ficheiro está detalhada na secção 3.5.3.

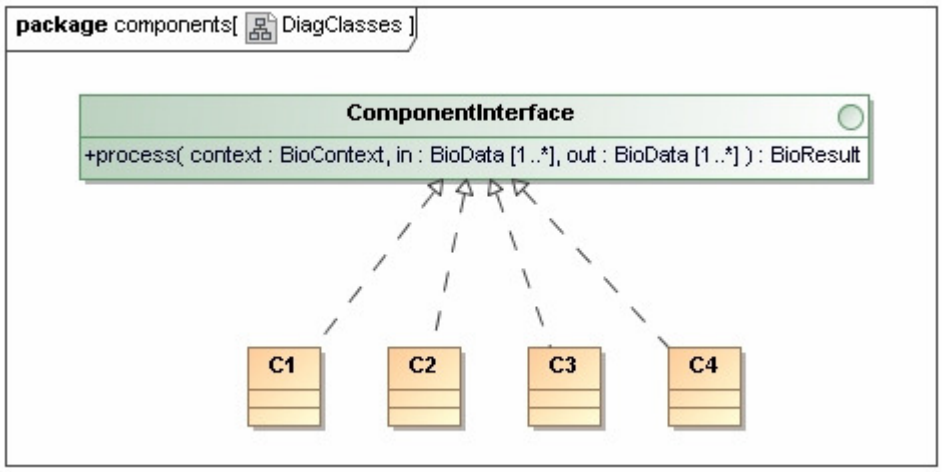


Figura 34 – Interface para os componentes externos.

A interface ComponentInterface, de obrigatória implementação por parte de cada componente, terá um único método chamado *process* que realiza o processamento destinado a esse componente. O método é composto por um conjunto de dados de entrada e um conjunto de dados de saída do tipo *BioData*. O tipo *BioData* corresponde a um *array* de *floats* que guardam as diversas amostras do sinal *ECG*. Caso exista necessidade, este pode ser alterado para suportar outros tipos através da extensão da classe BioData.
O parâmetro contexto permite a passagem de informação de contexto eventualmente necessária para a realização do processamento. Por exemplo, será através deste objecto que será passado a indicação do tipo de operação (autenticação, identificação, registo ou avaliação) a realizar pelo processamento. No entanto e na eventual necessidade de ser passado outro tipo de informação, esta classe poderá ser estendida de modo a suportar outras informações.
Como retorno do método *process* é usado um objecto do tipo *BioResult* que indica o resultado da execução do processamento realizado. Este resultado pode ser sucesso ou falha. Em caso de falha existem várias

causas possíveis que podem ser enumeradas nestes objectos de forma a descrever a causa com maior detalhe possível.
Cada elemento do *array in* e *out* corresponde a um porto (*port*) de entrada ou saída de informação, necessária para receber ou passar dados de ou para outro(s) componente(s). Essa passagem de informação entre componentes é definida também no ficheiro de configuração, através de ligações (*links*) entre os diversos componentes.

Foram definidas algumas limitações aos componentes e às ligações entre eles, de forma a garantir o bom funcionamento do sistema, evitando uma degradação de desempenho face a um possível número elevado de recursos.
A lista seguinte descreve os vários limites impostos:

- Número máximo de portas de entrada por componente: 5.
- Número máximo de portas de saída por componente: 5.
- Número máximo de componentes no sistema biométrico: 20
- Número máximo de ligações por sistema biométrico: 50
- Cada porto de entrada apenas suporta uma única ligação.
- Cada pode de saída pode suportar até um máximo 5 ligações.
- As ligações são unidireccionais e ligam sempre um porto de saída a um porto de entrada.
- Não é permitido qualquer tipo de realimentação. Ou seja, ligar (directamente ou indirectamente) uma saída a uma entrada do mesmo componente.

3.5.3 Configuração

Cada componente a usar pela *framework*, terá que ser configurado num ficheiro de configuração, juntamente com as diversas ligações existentes entre eles. Este deve estar presente numa directoria referida no *class path* definido para a execução do sistema biométrico.

O ficheiro de configuração está definido no formato *XML*, acompanhado por um ficheiro de *schema* que o valida sintacticamente.
Podemos dividir este ficheiro de configuração em quatro partes:

- **Componentes**: Define a lista de componentes externos.
- **Ligações**: Estabelece as ligações entre componentes e respectivos portos, incluindo a ligação com a *framework*.
- **Propriedades:** Define os parâmetros de configuração (nome e valor) a usar pelos diversos componentes externos.
- **Livrarias:** Define as livrarias necessárias na execução dos componentes externos.

Na tabela seguinte estão descritas as *tags XML* utilizados no ficheiro de configuração:

Elemento XML	Tipo / Valor	Descrição
Componentes		
component@id	*Integer*	Identificador único do componente dentro do ficheiro de configuração. Este valor será usado para as ligações.
component@Implement-class-name	*String*	Nome da classe Java que implementa o componente. Exemplo: "classificator.ClassificatorLZW"
Ligações		
link@from-cmp-id	*Integer*	Indicador do componente no qual a ligação terá origem.
link@to-cmp-id	*Integer*	Indicador do componente no qual a ligação irá terminar.
link@from-cmp-port-id	*Integer*	Índice da porta de saída do componente.
link@to-cmp-port-id	*Integer*	Índice da porta de entrada do componente.
Propriedades		
property@name	*String*	Nome da propriedade
property@value	*String*	Valor da propriedade
Livrarias		
jar-file@name	*String*	Nome do ficheiro a que corresponde a respectiva livraria. Exemplo: "lib-util.jar"
Outros		
bio-components-configuration@version	*String*	Identificador único para a configuração. Utilizada para distinguir diferentes configurações de componentes e os dados que eles produzem e guardem na base de dados.
description	*String*	Descrição do tipo de configuração usada na implementação desse sistema biométrico.

Tabela 2 – Elementos *XML* usados no ficheiro de configuração.

Este ficheiro de configuração é um exemplo prático da configuração de um sistema biométrico de base *non-fiducial*, constituído por três componentes e respectivas ligações:

```
<?xml version="1.0" encoding="UTF-8" ?>
<bio-components-configuration version="v2">

<description>Non fiducial approach: quantization 8 bits and classifier LZW</description>

<class-loader>
        <jar-file name="comp-libs/Non-FiducialComponentes.jar" />
        <jar-file name="comp-libs/jfreechart.jar" />
</class-loader>

<components>
        <component id="1" implement-class-
name="bioEcg.components.nonfiducial.preprocessor.PreProcessor" />
```

```
        <component id="2" implement-class-
name="bioEcg.components.nonfiducial.featuresExtractions.QuantizerUniformDistribution" />

        <component id="3" implement-class-
name="bioEcg.components.nonfiducial.classificator.ClassificatorLZW" />
</components>

<properties>
        <property name="bioEcg.components.nonfiducial.preprocessor.filter"
value="FILTER.TXT" />
        <property name="classificator.authentication.thresould" value="0.2" />
        <property name="classificator.identification.thresould" value="0.2" />
        <property name="classificator.enrollment.minNumberOfSegAllow" value="55" />
        <property name="classificator.audit.train_percent_size" value="30" />
        <property name="classificator.audit.test_percent_size" value="70" />
        <property name="classificator.audit.thresould_fraction_resolution" value="100" />
        <property name="classificator.audit.NMODEL" value="12" />
        <property name="classificator.audit.NTEST" value="1" />
        <property name="classificator.audit.NREP" value="30" />
        <property name="classificator.audit.identification.enable" value="true" />
        <property name="classificator.audit.authentication.enable" value="true" />
        <property name="classificator.audit.authentication.usertune.enable" value="true" />
</properties>

<links>
        <link from-cmp-id="f-in" from-cmp-port-id="f-in" to-cmp-id="1" to-cmp-port-id="1" />
        <link from-cmp-id="1" from-cmp-port-id="1" to-cmp-id="2" to-cmp-port-id="1" />
        <link from-cmp-id="2" from-cmp-port-id="1" to-cmp-id="3" to-cmp-port-id="1" />
        <link from-cmp-id="3" from-cmp-port-id="1" to-cmp-id="f-out" to-cmp-port-id="f-out" />
</links>

</bio-components-configuration>
```

3.5.4 Base de Dados

A informação necessária para o funcionamento do sistema biométrico será guardada, de forma persistente, numa base de dados relacional. Essa informação deve ser composta pelo registo de utilizadores e os respectivos dados biométricos. Os dados biométricos, correspondem ao sinal *ECG RAW* gerados directamente pela aquisição (*BioPLUX*). A partir deste sinal *ECG RAW* são gerados outros dados, produzidos pelos diversos componentes, que devem também ser mantidos na base de dados, de forma a melhorar os tempos de resposta do sistema.
São registados em base de dados o tempo total de execução de cada operação e também o tempo de execução de cada componente externo.

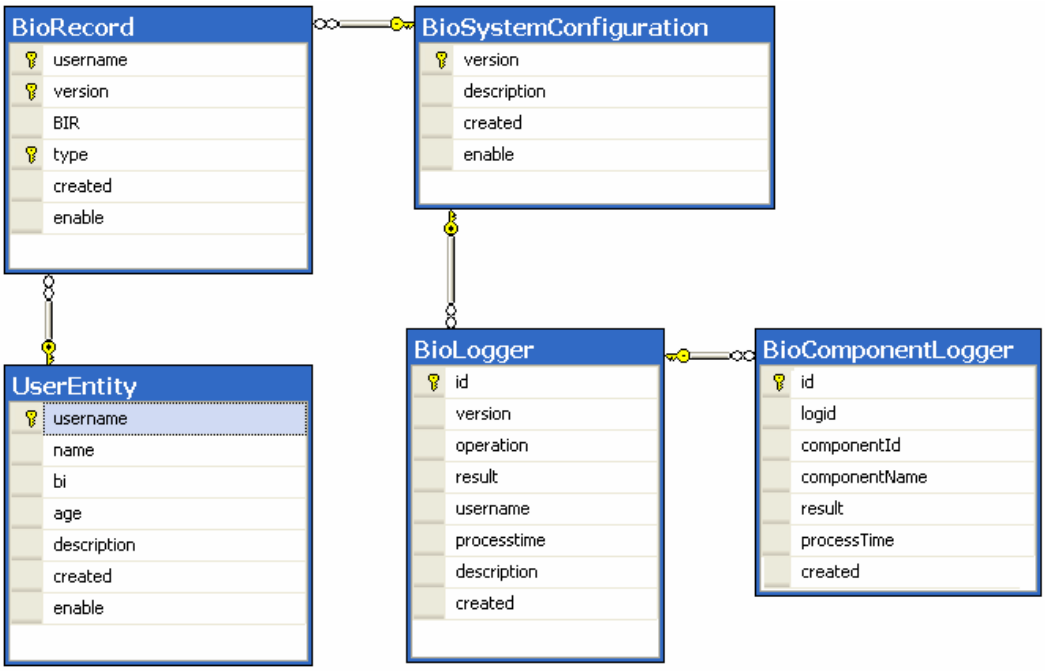


Figura 35 – Entidades a persistir na base de dados relacional.

A Figura 35 apresenta as entidades e os respectivos campos a guardar na base de dados relacional. Este modelo de entidades pode ser alterado no futuro de forma a contemplar a especificação definida para o *standard ISOIEC FCD 19785* (*Common Biometric Exchange Formats Framework*). Este define um formato comum para registo dos dados biométricos numa base de dados ou para a troca de entre sistemas biométricos.

3.5.5 WebService

O sistema biométrico, através da *framework*, tem um conjunto de operações (ver Figura 9) disponíveis para o utilizador. Essas operações poderão ser acedidas remotamente via *webservice*. Desta forma qualquer dispositivo, incluindo dispositivos móveis, pode de forma simples requer operações biometrias, para a realização de identificação pessoal.
Esta tecnologia permite às aplicações clientes serem independentes da linguagem de programação a que o sistema biométrico foi desenvolvido, e também da plataforma/sistema operativo a que esta assenta.

A interface *WebService* do sistema biométrico terá disponíveis os seguintes operações:

Método	Parâmetros		
	Nome	Tipo	Direcção
enrollment	*id*	*UserInfo*	*in*
	samples	*BioSamples*	*In*
		Boolean	*out*
authentication	*id*	*UserId*	*In*
	samples	*BioSamples*	*In*
		Boolean	*out*
identification	*samples*	*BioSamples*	*In*
		UserId	*out*
evaluation		*EvaluationResult*	*out*

Tabela 3 – Métodos e respectivos parâmetros propostos para o *webservice.*

Tipos utilizados pelos serviços:

Tipos	Atributos	
	Nome	Tipo
UserId	*Id*	*Integer*
	bi	*String*
	userName	*String*
UserInfo	*bi*	*String*
	Name	*String*
BioSamples	*raw*	*Float[ ]*
	source	*Integer*
EvaluationResult	*targetFile*	*String*

Tabela 4 – Tipos utilizados no *webservice*.

3.6 Implementação

A implementação da *framework* para sistemas biométricos foi implementada recorrendo à linguagem de programação *JAVA* (*Java SE 1.6*). Foram utilizadas também várias livrarias, tais como:

- Loj4j-1.2.16
- EclipseLink JPA-2.3.2
- Xerces-2.11.0
- JAX-WS-2.2.5

4 Verificação experimental

Neste capítulo apresentam-se resultados da verificação experimental da *framework,* para duas abordagens de base distintas:

I. *fiducial* – denominada de basic 10.
II. *non-fiducial* – denominada *LZ78.*

Os resultados apresentados baseiam-se no desempenho do sistema quanto aos erros de identificação e autenticação pessoal, bem como quanto ao tempo envolvido na execução de cada operação por parte do servidor.

4.1 *Setup* experimental

Na verificação experimental usou-se o *setup* experimental representado na Figura 36, em que o sinal *ECG* foi capturado por um conjunto de sensores, adquirido pelo dispositivo *BioPLUX* [9] e enviado para a aplicação cliente.

Figura 36 – Aquisição do sinal *ECG* e envio para a aplicação cliente.

Usou-se um dispositivo com um conjunto de sensores para a aquisição do sinal *ECG*, onde os eléctrodos ficam por baixo e em contacto com a palma das mãos, que permite ajustar a posição das superfícies de contacto do sensor para mãos de diferentes tamanhos e resolve igualmente a disposição dos sensores para facilitar a aquisições correctas do *ECG* dos dedos.
O próprio sensor já possui filtragem e amplificação perto das superfícies de contacto. A filtragem serve para minimizar o ruído fora do espectro de frequências úteis para o sinal *ECG* e a amplificação a montante aumenta a *SNR*. A Tabela 5 apresenta as características do sensor.

Característica	Valor
Ganho	1000 Vezes
Filtragem passa banda	1 – 30 *Hz*

Tabela 5 – Características do sensor *ECG*.

O sensor das mãos está ligado ao dispositivo electrónico *BioPLUX*, que recebe o sinal analógico filtrado e amplificado do sensor e tem como função digitaliza-lo para o enviar via *BlueTooth* [47] para o dispositivo onde reside a aplicação cliente. A Tabela 6 apresenta as características do dispositivo *BioPLUX*.

Característica	Valor
Alcance *BlueTooth*	Até 12 metros
Resolução	12 Bits
Amostragem	1000*Hz*

Tabela 6 – Características do dispositivo *BioPLUX*.

4.2 Configurações implementadas

Para a abordagem *fiducial* usaram-se os seguintes blocos:

- Pré-processamento – Filtragem do sinal com filtro passa banda (2-30Hz). Detecção dos picos *R* com a ajuda de uma derivada, segundo a abordagem do *Engelese e Zeelenberg* [17]. Segmentação do sinal *ECG* em amostras e por fim faz-se medias das amostras para tornar o sinal mais limpo de ruído.
- Extracção de características – Obtenção de 10 pontos de referência relativos à amplitude e posição relativa dos complexos *P-QRS-T*, como mostra a Figura 8. A forma para obter estes pontos é baseada na procura de máximos e mínimos dentro de um determinado *range*.
- Classificador – Cálculo das distâncias, através do método do vizinho mais próximo *1-NN* (ver Figura 18) com base na distância euclidiana.

Para a abordagem *non-fiducial*:

- Pré-processamento – o mesmo da abordagem *fiducial*.
- Extracção de características – quantização escalar uniforme a 8 *bits*, como se mostra a Figura 17.
- Classificador – cálculo da entropia relativa como medida de semelhança entre duas sequência, recorrendo à utilização de uma implementação do compressor *LZ78* (ver 2.4.3).

4.3 Base de dados

A verificação experimental baseou-se numa base de dados de sinais *ECG* previamente adquirida, sendo usados sinais sem qualquer processamento (*raw*).

Esta base de dados consiste em aquisições de sinais *ECG* de 44 indivíduos saudáveis. Durante as aquisições os indivíduos estavam em repouso ouvindo uma explicação do estudo em que estavam envolvidos.

Na aquisição foi utilizado o *setup* da Figura 37, em que apenas se usou dois dos quatro eléctrodos (os dois da parte inferior da imagem), colocados na palma das mãos. O sistema de aquisição usava uma frequência de amostragem de 1000*Hz*.

Figura 37 – Imagem do *setup* utilizado para a aquisição de sinais *ECG*.

Todos os dados desta base de dados foram lidos pelo sistema biométrico, através de várias operações de registo, de forma a construir os vários modelos na base de dados do sistema biométrico, como apresenta a Figura 38.

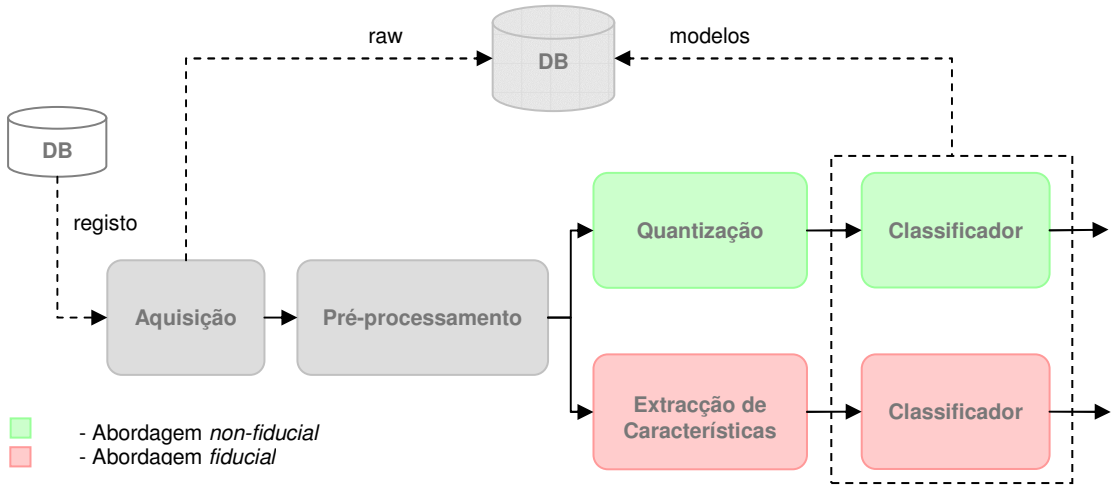


Figura 38 – Esquema de carregamento da base de dados no sistema biométrico para as duas abordagens.

Esta operação foi realizada para as duas abordagens em instantes distintos e após a correcta configuração da *framework* de acordo com a abordagem adoptada.

4.4 Metodologia de avaliação

A validação do sistema baseou-se na metodologia de *cross-validation* [63], que divide o conjunto de amostras recolhidas em dois grupos, conjunto de treino (30%), e conjunto de teste (70%), como mostra a Figura 39. O primeiro é usado para treinar o sistema enquanto que o

outro (normalmente distinto) é usado para o testar, obtendo-se um conjunto de resultados de desempenho do sistema biométrico.
Nesta metodologia realizaram-se 30 (*NREP*) repetições do teste, usando diferentes conjuntos de treino e teste, tentando reduzir a variabilidade nos resultados obtidos.

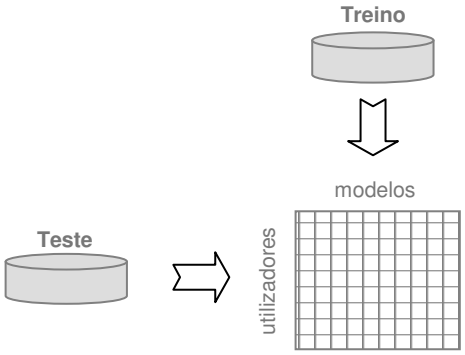


Figura 39 – Representação gráfica da implementação do classificador tendo como resultado a matriz de distribuição.

O conjunto de treino permite obter os modelos (*templates*), a usar como referência na classificação. E o conjunto de teste permite exercitar o sistema através da sua comparação com o modelo, ajudando a construir a matriz de distâncias necessária para a avaliação de desempenho do sistema biométrico (ver secção 2.5).
A comparação realizada pelo classificador entre o conjunto de treino e o de teste varia no número de segmentos utilizados em cada conjunto: define-se por *NMODEL* o número de modelos usados em teste; e define-se por *NTEST* o número de padrões usados para testar o modelo. Desta forma temos duas variáveis, uma para o conjunto de treino (*NMODEL*) e outra para o conjunto de teste (*NTEST*), que terão diferentes valores durante a execução dos testes. A escolha desses valores teve por base de referência de outros estudos realizados, de forma a possibilitar a comparação dos resultados.

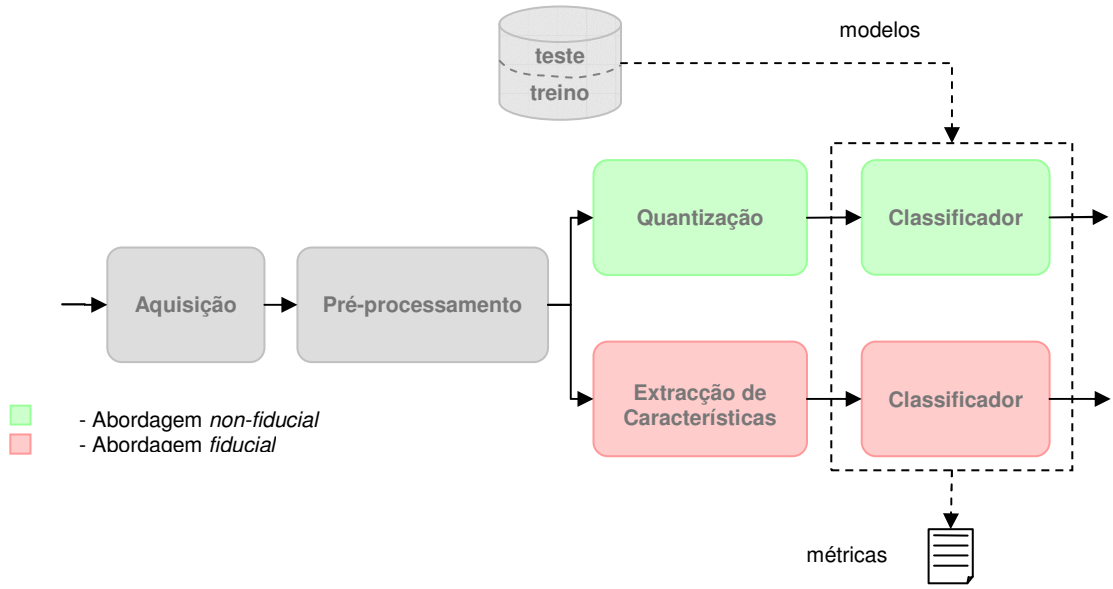


Figura 40 – Esquema de avaliação do sistema biométrico para duas abordagens.

Na Figura 40 é apresentado o esquema de avaliação do sistema biométrico implementado na *framework*, para as duas abordagens distintas.
A avaliação do sistema é realizada pelo componente classificador de cada abordagem, que consultando os diversos modelos exististes na base de dados (treino e teste), produz um ficheiro com os resultados das métricas.

O tempo envolvido na execução de cada operação por parte do sistema biométrico é registado em base e dados, permitindo uma análise dos tempos despendidos em cada operação pela *framework* e por cada componente externo envolvido nessa operação.
Esses tempos são relativizados com base no número de segmentos processados.

4.5 Resultados obtidos

A Tabela 7 e Tabela 8 apresentam os resultados dos testes de avaliação de desempenho do sistema biométrico, para a identificação e autenticação pessoal, baseado na abordagem *fiducial* e *non-fiducial* respectivamente. Estas apresentam a variação da taxa de erro de identificação e autenticação pessoal, para um conjunto de valores de *NMODEL* e *NTEST*. Trata-se de taxas médias obtidas com o cálculo da média aritmética dos vários valores intermédios obtidos em cada um dos *NREP* testes realizados.
As taxas de erro apresentadas têm o seguinte significado apresentado na seguinte lista, podendo se ser visto com maior detalhe no capítulo 2.5:

- *IdError* – taxa média de erro de identificação pessoal.
- *EER* – taxa média de erro da autenticação pessoal.
- *EER UT* – taxa média de erro para a autenticação pessoal dentro do domínio das amostras de teste de cada utilizador. O valor final

é a média aritmética de todos os resultados intermédios obtidos em cada utilizador.

NREP	NMODEL	NTEST	Identificação	Autenticação	
			IdError	EER	EER UT
30	4	1	57,67%	1.80%	1.92%
30	6	1	53,27%	1,71%	1,73%

Tabela 7 – Resultados do desempenho da abordagem *fiducial* na identificação e autenticação pessoal.

NREP	NMODEL	NTEST	Identificação	Autenticação	
			IdError	EER	EER UT
30	4	1	61,65%	10,60%	8,79%
30	4	2	53,69%	6,69%	10,21%
30	4	3	51,36%	7,20%	11,04%
30	6	1	59,80%	2,17%	8,35%
30	6	2	50,85%	7,59%	10,44%
30	6	3	43,18%	7,66%	1,11%

Tabela 8 – Resultados do desempenho da abordagem *non-fiducial* na identificação e autenticação pessoal.

Os resultados obtidos revelam valores de taxa de erro para a autenticação (*EER* e *EER UT*) substancialmente mais baixos que os valores obtidos para a identificação, independentemente da abordagem seguida.
Na autenticação pessoal a taxa de erro mínima obtida foi de 1,71 %, na abordagem *fiducial* para *NTEST* igual a 1 e *NMODEL* igual a 6. Por outro lado na abordagem *non-fiducial* esse mínimo é de 2,17% para *NTEST* igual a 1 e *NMODEL* igual a 6. Na autenticação aplicada ao domínio de cada utilizador, o valor mínimo da taxa de erro foi de 1,73% na abordagem *fiducial* (*NTEST* igual a 1 e *NMODEL* igual a 6) e de 1,11% na abordagem *non-fiducial* (*NTEST* igual a 3 e *NMODEL* igual a 6).

A Figura 41 apresenta duas curvas de *ROC* nos dois gráficos em baixo e duas curvas da taxa de erro *FAR* e *FRR* nos gráficos em cima. Estas apresentações a avaliação de desempenho do sistema biométrico para a abordagem *fiducial* (dois do lado esquerdo) e *non-fiducial* (dois do lado direito).

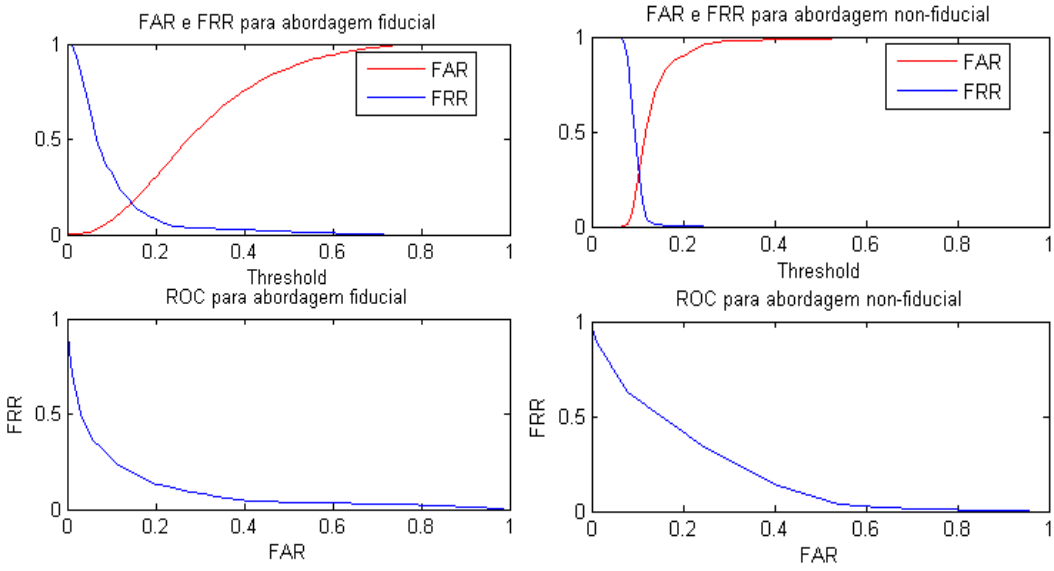


Figura 41 – Curvas da taxa de erro *FAR* e *FRR* (em cima) e curvas *ROC* (em baixo). Na esquerda foi usada a abordagem *fiducial* e na direita a abordagem *non-fiducial*.

A Figura 42 e a Figura 43 apresentam os valores médios de duração temporal despendidos por cada componente do sistema biométrico durante a execução das diversas operações para a abordagem *fiducial* e *non-fiducial* respectivamente.
Os valores obtidos traduzem uma evidente ocupação do tempo total utilizado durante as operações de registo, identificação e autenticação, por parte do componente de pré-processamento, com valores nunca inferiores a 91% do tempo total. Por outro lado na operação de avaliação a distribuição do tempo é claramente atribuída ao componente de classificação, com maior evidência na abordagem *non-fiducial.* Sendo esse tempo, na abordagem *fiducial,* partilhado pelo classificador (72,42%) e pela *framework* (27,57%).
Podemos ainda destacar uma pequena percentagem de tempo ocupado pela *framework* durante a operação de registo, referente à utilização da base de dados com a gravação dos dados biométricos dos utilizadores. Este valor é de 5,45% na abordagem *non-fiducial,* subindo ligeiramente para 5,97% na abordagem *fiducial*.

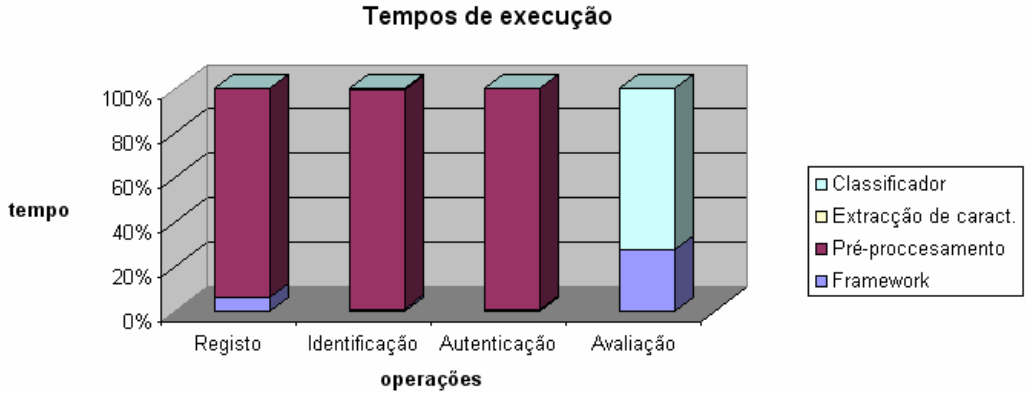


Figura 42 – Tempos médios de execução em cada componente para a abordagem *fiducial*.

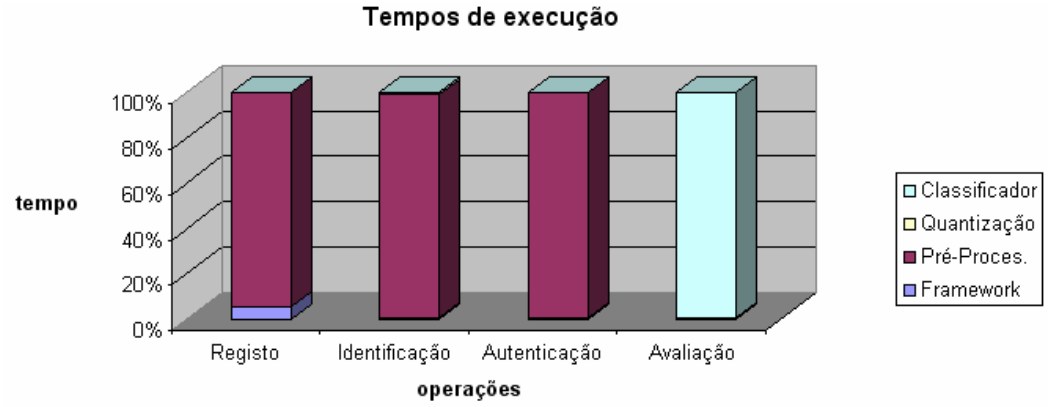


Figura 43 – Tempos médios de execução em cada componente para a abordagem *non-fiducial*.

5 Conclusão

A biometria baseada em características comportamentais apresenta-se como um dos campos de investigação em desenvolvimento na actualidade. O electrocardiograma (*ECG*) anteriormente usado apenas em aplicações clínicas, tem sido usado agora como indicador biométrico. As abordagens ao problema da biometria baseada no *ECG* consistem em retirar características únicas do sinal de *ECG* que permitem distinguir o utilizador dos demais e ao mesmo tempo assegurar a não variação dessas características no mesmo utilizador.

A plataforma de *software* (*framework)* proposta neste trabalho permite de uma forma simplificada, criar sistemas biométricos de identificação e autenticação pessoal usando sinais *ECG*. Uma das características desta *framework* é permitir a integração de novas funcionalidades através da instalação de novos componentes, que constituam os blocos de pré-processamento, extracção de características e o classificador. Desta forma a *framework* permite criar sistemas biométricos que usam diferentes abordagens para o problema da identificação e autenticação pessoal baseada no sinal *ECG*.

A *framework* possui uma arquitectura do tipo cliente servidor, baseada em *webservices*, permitindo o acesso de múltiplos e diferentes clientes a partir de qualquer localização. O desenvolvimento desta foi baseado no modelo de desenho de *software MVC* com três camadas: *user interface*, *controlo* e modelo de dados. A vantagem da utilização deste padrão de desenho é a separação das diferentes áreas de forma a diminuir o acoplamento entre camadas aumentando a coesão.

Foram desenvolvidos componentes, que passaram a fazer parte da *framework*, e permitiram avaliar o desempenho do sistema biométrico implementado, em duas abordagens distintas. Uma das abordagens (*fiducial*) assenta em pormenores dos diferentes segmentos da forma de onda do sinal *ECG*, enquanto que a outra abordagem (*non-fiducial*) tem a vantagem de não depender criticamente desses pormenores.

Na abordagem *fiducial*, usou-se um bloco de pré-processamento com filtragem, detecção dos picos *R*, segmentação e médias de sinal *ECG*. Para o bloco de extracção de características, foram usadas 10 características relativas à amplitude e posição relativa dos complexos *P-QRS-T*. Finalmente no classificador foi utilizada a distância baseada no vizinho mais próximo *1-NN*.

Foi proposta uma nova abordagem *non-fiducial* baseada no estudo em [8] que implementa uma ferramenta para a classificação de texto baseada numa versão modificado do algoritmo *Lempel-Ziv* (*LZ78*).

Para testar o sistema foi usada uma base de dados de sinais *ECG* de 44 indivíduos. Durante as aquisições os indivíduos estavam em repouso ouvindo uma explicação do estudo em que estavam envolvidos. Na aquisição foi utilizado o *setup* em que apenas se usou dois dos quatro eléctrodos (os dois da parte inferior da imagem), colocados na palma das mãos. O sistema de aquisição usava uma frequência de amostragem de 1000*Hz* e após segmentação foram obtidos para cada indivíduo 76 *single-heart-beats.* Cada *single-heart-beat* é constituído por 600 amostras.

Os resultados obtidos mostram que a abordagem *fiducial* obteve os melhores resultados na autenticação pessoal com uma taxa de erro de 1,71% (*NMODEL*=6 e *NTEST*=1). Enquanto que na identificação pessoal, os melhores resultados foram obtidos na abordagem *non-fiducial* com a taxa de erro de 43,18% (*NMODEL*=6 e *NTEST*=3).

A *framework* desenvolvida facilitou a realização destes testes e a obtenção dos resultados na medida em que possibilitou a troca de abordagens de forma prática através da substituição dos componentes externos na configuração.

6 Futuros trabalhos

As funcionalidades de base implementadas permitem explorar cenários de aplicação desafiantes:

- Aplicações clientes em telemóveis que façam a aquisição do sinal *ECG* e o apresentem no respectivo visor em tempo real.
- Integração com sistemas de pagamento, usando a tecnologias como o *Near Field Communication* (*NFC*).

Quanto às funcionalidades implementadas, seria interessante explorar alternativas:

- Outras bases de dados não relacionais que permitem outros níveis de *performance* de acesso.
- Novas soluções de desenho que permitam atingir outros níveis de resposta e de desempenho.

Extensão de funcionalidades a considerar:

- Interface gráfico para gestão e configuração dos componentes instalados na *framework*, incluindo as ligações entre eles.
- Soluções de segurança que permitam a confidencialidade da informação, tanto ao nível do armazenamento dos dados em base de dados como na comunicação entre as diversas entidades envolvidas no sistema biométrico.
- Tornar a *framework* proposta neste trabalho, compatível com o *standard* internacional *BioAPI 2.0*.

Referências

[1] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition.” IEEE Trans. Circuits Syst. Video Techn., vol. 14, no. 1, pp. 4–20, 2004.

[2] Jorge Salvador Marques, “Reconhecimentos de padrões”, IST Press, 1999.

[3] L. A. S. Medina and A. L. N. Fred, “Genetic algorithm fo clustering temporal data – application to the detection of stress from ecg signals,” in Proc 2nd International Conference on Agents and Artificial Intelligence (ICAART), 2010, pp. 135–142.

[4] L. Biel, O. Pettersson, L. Philipson, and P. Wide, “ECG analysis – a new approach in human identification,” IEEE Transactions on Instrumentation and Measurement, vol. 50, no. 3, pp. 808–812, June 2001.

[5] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, “ECG to identify individuals.” Pattern Recognition, vol. 38, no. 1, pp. 133–142, 2005.

[6] A. Chan, M. Hamdy, A. Badre, and V. Badee, “Wavelet distance measure for person identification using electrocardiograms,” Instrumentation and Measurement, IEEE Transactions on, vol. 57, no. 2, pp. 248–253, Feb. 2008.

[7] C.-C. Chiu, C.-M. Chuang, and C.-Y. Hsu, “A novel personal identity verification approach using a discrete wavelet transform of the ECG signal,” in MUE ’08: Proceedings of the 2008 International Conference on Multimedia and Ubiquitous Engineering. Washington, DC, USA: IEEE Computer Society, 2008, pp. 201–206.

[8] Coutinho, D.P.; Fred, A.L.N.; Figueiredo, M.A.T.; , “One-Lead ECG-based Personal Identification Using Ziv-Merhav Cross Parsing,” Pattern Recognition (ICPR), 2010 20th International Conference on , vol., no., pp.3858-3861, 23-26 Aug. 2010 DOI: 10.1109/ICPR.2010.940

[9] Sistema BioPLUX, acedido em Janeiro de 2011, http://www.bioplux.com/pt-pt/home

[10] Comissão Nacional de Protecção de Dados, acedido em Maio de 2011 http://www.cnpd.pt/bin/orientacoes/principiosbiometricos.htm

[11] Wikipedia contributors, 'Quantization (signal processing)', *Wikipedia, The Free Encyclopedia,* 21 February

2012, 06:27 UTC, <http://en.wikipedia.org/w/index.php?title=Quantization_(signal_processing)&oldid=478027839> [accessed 11 March 2012]

[12] A. Oppenheim, A. Willsky, and S. Nawab. Signals and Systems, 2nd Ed. Prentice-Hall, 1996.

[13] S. Suppappola and Y. Sun. A comparison of three qrs detection algorithms using the aha ecg database. IEEE Engineering in Medicine and Biology Society, 13:586-587, 1991.

[14] T. Wrublewski, Y. Sun, and J. Beyer. Real-time early detection of r waves of the ecg signals. IEEE Engineering in Medicine and Biology Society, 1:38-39,1989.

[15] Michael Dipperstein. Lempel-Ziv-Welch (LZW) Encoding Discussion and Implementation. Acedido em 15 Abril de 2011. http://michael.dipperstein.com/lzw/index.html

[16] Hugo Silva, Hugo Gamboa, and Ana Fred. HiMotion Project ECG Data Preprocessing. March 29, 2011.

[17] G.M. Friesen, T.C. Jannett, M.A. Jadallah, S.L. Yates, S.R. Quint, and H.T. Nagle, “A comparison of the noise sensitivity of nine QRS detection algorithms,” IEEE Trans. Biomed. Eng., vol. 37, pp. 85-98, 1990.

[18] K. Esbensen, S. Schönkopf, and T. Midtgaard, Multivariate Anal. In Practice, 1st ed. Trondheim, Norway: Camo, 1994, vol. 1, p. 361.

[19] C. M. Bishop, Neural Networks for Pattern Recognition. New York Oxford Univ. Press, 1995, p. 482.

[20] R. G. Gonzales and R. E. Woods, Digital Image Processing. Reading, MA: Addison-Wesley, 1992, p. 716.

[21] J. Lu, Discriminant learning for face recognition, Ph.D. thesis, University of Toronto, Toronto, Ontario, Canada, 2004.

[22] W. A. H. Engelse and C. Zeelenberg, “A single scan algorithm for QRS-detection and feature extraction,” IEEE Comput. Card., Long Beach: IEEE Computer Society, 1979, pp. 37-42.

[23] Kyoso, M., & Uchiyama, A. (2001). Development of an ECG identification system. Proceedings of the 23rd IEEE Engineering in Medicine and Biology Confonerence, 4, 3721-3723.

[24] Shen, T. W., Tompkins, W. J., & Hu, Y. H. (2002). One-lead ECG for identity verification. Proceedings of the 2nd Joint EMBS/BMES Conference, 62–63.

[25] Saechia, S., Koseeyaporn, J., & Wardkein, P. (2005). Human Identification System Based ECG Signal. TENCON 2005, 2005 IEEE Region 10.

[26] Zhang, Z., & Wei, D. (2006). A new ECG identification method using Bayes' Theorem. TENCON IEEE Region 10 Conference, Hong Kong.

[27] Singh, Y. N., & Gupta, P. (2008). ECG to Individual Identification. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, 2008.

[28] Boumbarov, O., Velchev, Y., & Sokolov, S. (2009). ECG personal identification in subspaces using radial basis neural networks. IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2009.

[29] Plataniotis, K. N., Hatzinakos, D., & Lee, J. K. M. (2006). ECG Biometric Recognition Without Fiducial Detection. 2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference.

[30] Fatemian, S. Z., & Hatzinakos, D. (2009). A new ECG feature extractor for biometric recognition. 16th International Conference on Digital Signal Processing, 2009.

[31] Palaniappan, R. & Krishnan, S. (2004). Identifying individuals using ECG beats, International Conference on Signal Processing and Communications, pp. 569 – 572.

[32] Ting, C. M. & Salleh, S. H. (2010). ECG based personal identification using extended kalman filter, 10th International Conference on Information Sciences Signal Processing and their Applications, pp. 774 –777.

[33] Venkatesh, N. & Jayaraman, S. (2010). Human electrocardiogram for biometrics using DTW and FLDA, 20th International Conference on Pattern Recognition (ICPR), pp. 3838 –3841.

[34] Tawfik, M., Selim, H. & Kamal, T. (2010). Human identification using time normalized QT signal and the QRS complex of the ECG, 7th International Symposium on Communication Systems Networks and Digital Signal Processing, pp. 755 –759.

[35] Wübbeler, G., Stavridis, M., Kreiseler, D., Bousseljot, R. & Elster, C. (2007). Verification of humans using the electrocardiogram, Pattern Recogn. Lett. 28(10): 1172–1175.

[36] Molina, G. G., Bruekers, F., Presura, C., Damstra,M. & van der Veen,M. (2007). Morphological sythesis of ECG signals for person authentication, 15th European Signal Proc. Conf., Poland.

[37] Odinaka, I., Lai, P.-H., Kaplan, A., O'Sullivan, J., Sirevaag, E., Kristjansson, S., Sheffield, A. & Rohrbaugh, J. (2010). Ecg biometrics: A robust short-time frequency analysis, IEEE International Workshop on Information Forensics and Security, pp. 1 –6.

[38] Ye, C., Coimbra, M. & Kumar, B. (2010). Investigation of human identification using two-lead electrocardiogram (ECG) signals, 4th Int. Conf. on Biometrics: Theory Applications and Systems, pp. 1 – 8.

[39] Ghofrani, N. & Bostani, R. (2010). Reliable features for an ECG-based biometric system, 17th Iranian Conference of Biomedical Engineering, pp. 1 –5.

[40] Li, M. & Narayanan, S. (2010). Robust ECG biometrics by fusing temporal and cepstral information, 20th International Conference on Pattern Recognition, pp. 1326 –1329.

[41] Mahalanobis, P C (1936). "On the generalised distância in statistics". Proceedings of the National Institute of Sciences of Índia 2 (1): 49–55.

[42] "Distributed Application Architecture". Sun Microsystem. Acedido em 7 Maio de 2011. http://java.sun.com/developer/Books/jdbc/ch07.pdf

[43] Web Services Architecture - W3C Working Group Note 11 February 2004. http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/

[44] Web Services Description Language (WSDL) Version 2.0 Part 0: Primer. http://www.w3.org/TR/2007/REC-wsdl20-20070626/

[45] SOAP Version 1.2 Part 1: Messaging Framework (Second Edition). http://www.w3.org/TR/2007/REC-soap12-part1-20070427/

[46] HTTPS – Wikipedia, acedido em Junho de 2011, http://en.wikipedia.org/wiki/HTTP_Secure.

[47] BlueTooth – Wikipedia, acedido em Junho de 2011, http://pt.wikipedia.org/wiki/Bluetooth.

[48] W.W. Eckerson: Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications, Open Information Systems, vol. 10, no. 1, January 1995

[49] Channu Kambalyal. 3-Tier Architecture. http://channukambalyal.tripod.com/NTierArchitecture.pdf

[50] "BioAPI Specification Version 1.1". BioAPI Consortium. ANSI/INCITS 358-2002 (March 2001). http://www.bioapi.org/Version_1.1_Description.asp

[51] X. Yuan, S.C. Hui, M.H.K. Leung, Y. Gao, Towards a BioAPI compliant face verification system, Computer Standards & Interfaces 26 (2004) 289–299.

[52] BioAPI 2.0, BioAPI Consortium: (2005). http://www.bioapi.org/Version_2.0_Description.asp

[53] BioAPI for win32 by Intel, SAFLINK, IriScan, and Mytec Technologies Inc. (September 2000). http://www.bioapi.org/Downloads/bioapi.zip

[54] BioAPI for Linux/Unix by National Institute of Standards and Technology (NIST) International Biometric Group (IBG) (January 2003). http://www.bioapi.org/Downloads/bioapi_unix_1.2.tar.gz

[55] BioAPI Framework for Windows CE 1.1 by National Biometrics Security Project (NBSP) (January 2005). http://www.bioapi.org/devtools3.asp

[56] JNI BioAPI wrapper for Win32 platform by Gens Software (March 2003). http://www.genssoft.com.

[57] Open-source project JBioAPI (JNI BioAPI wrapper for Linux/Unix) (13 June, 2005). http://freshmeat.net/projects/jbioapi.

[58] C# wrapped Biometric API by H. Kaiser Yang. http://sourceforge.net/project/downloading.php?groupname=boiapi-dt&filename=BioAPI_CSHARP_Wrap-1.0.zip&use_mirror=superb-east

[59] WBF - Windows Biometric Framework API by Microsoft (July 2011). http://msdn.microsoft.com/en-us/library/windows/desktop/dd401509(v=vs.85).aspx

[60] Ziv, J., Merhav, N.: A measure of relative entropy between individual sequences with application to universal classification. IEEE Transactions on Information Theory 39 (1993) 1270–1279.

[61] Ziv, J., and Lempel, A. A universal algorithm for sequential data compression. IEEE Transactions on Information Theory 23 (1977), 337–343.

[62] Ziv, J., and Lempel, A. Compression of individual sequences via variable-rate coding. IEEE Transactions on Information Theory 24 (1978), 530–536.

[63] R. Kohavi, “A study of cross validation and bootstrap for accuracy estimation and model selection”, In Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1137-1143, 1995.

Anexos

Anexo A

Lista de requisitos não funcionais da *framework* biométrica.

- **RNF-1:** O sistema recorre a *webservices* para permitir o acesso remoto.
- **RNF-2:** A informação biométrica de cada utilizador é mantida numa base de dados relacional.
- **RNF-3:** A configuração do sistema biométrico é efectuada num ficheiro de configuração.
- **RNF-4:** A operação de avaliação do sistema biométrico gera um ficheiro com um conjunto de informação (matriz de distâncias) que permitirá a avaliação do sistema.
- **RNF-5:** Número máximo de portas de entrada por componente: 5.
- **RNF-6:** Número máximo de portas de saída por componente: 5.
- **RNF-7:** Número máximo de componentes no sistema biométrico: 20
- **RNF-8:** Número máximo de ligações por sistema biométrico: 50
- **RNF-9:** Cada porto de entrada apenas suporta uma única ligação.
- **RNF-10:** Cada porto de saída pode suportar até um máximo 5 ligações.
- **RNF-11:** As ligações são unidireccionais e ligam sempre um porto de saída a um porto de entrada.
- **RNF-12:** Não é permitido qualquer tipo de realimentação. Ou seja, ligar (directamente ou indirectamente) uma saída a uma entrada do mesmo componente.

Anexo B

Os quadros seguintes apresentam a descrição dos casos de utilização do sistema biométrico para as várias operações.

Verificar Autenticação

Nome:	**Verificar Autenticação**
Descrição:	Este caso de utilização permite verificar a autenticação pessoal.
Pré-condições:	
Actores:	Utilizador
Cenário principal:	
1.	O caso de utilização inicia-se quando o utilizador pretende verificar a sua autenticação.
2.	O utilizador preenche os seus dados de identificação.
3.	O utilizador selecciona o botão “Autenticação” e coloca as mãos no respectivo setup para a recolha do sinal ECG.
4.	O sistema apresenta uma indicação do inicio e do fim da aquisição do sinal ECG.
5.	O sistema, após realizar o processo de autenticação, indica ao utilizador o resultado dessa operação.
6.	O caso de utilização termina.
Cenário alternativo:	A
1.	No passo 3 do cenário principal, a aquisição do sinal ECG falha.
2.	O sistema apresenta uma mensagem informativa ao utilizador.
3.	O caso de utilização continua no passo 3 do cenário principal.

Verificar Identificação

Nome:	**Verificar Identificação**
Descrição:	Este caso de utilização permite efectuar a identificação pessoal.
Pré-condições:	
Actores:	Utilizador
Cenário principal:	
1.	O caso de utilização inicia-se quando o utilizador pretende efectuar a identificação pessoal.
2.	O utilizador selecciona o botão “Identificação” e coloca as mãos no respectivo setup para a recolha do sinal ECG.

3.	O sistema apresenta uma indicação do início e do fim da aquisição do sinal ECG.
4.	O sistema, após realizar o processo de identificação, indica ao utilizador o resultado dessa operação.
5.	O caso de utilização termina.
Cenário alternativo:	A
1.	No passo 2 do cenário principal, a aquisição do sinal ECG falha.
2.	O sistema apresenta uma mensagem informativa ao utilizador.
3.	O caso de utilização continua no passo 2 do cenário principal.

Registar Utilizador

Nome:	**Registar Utilizador**
Descrição:	Este caso de utilização permite registar novo utilizador ao sistema.
Pré-condições:	
Actores:	Utilizador e Administrador
Cenário principal:	
1.	O caso de utilização inicia-se quando o utilizador pretende inserir o seu registo.
2.	O utilizador preenche o formulário com os seus dados pessoais.
3.	O utilizador selecciona o botão “Registo” e coloca as mãos no respectivo setup para a recolha do sinal ECG. O Administrador deve garantir que os dados biométricos correspondem à identidade inserida no sistema.
4.	O sistema, após realizar o processo de registo, indica ao utilizador o resultado dessa operação.
5.	O caso de utilização termina.
Cenário alternativo:	A
1.	No passo 3 do cenário principal, a aquisição do sinal ECG falha.
2.	O sistema apresenta uma mensagem informativa ao utilizador.
3.	O caso de utilização continua no passo 3 do cenário principal.

Solicitar Avaliação

Nome:	**Solicitar Avaliação**
Descrição:	Este caso de utilização permite solicitar uma avaliação do sistema.
Pré-condições:	
Actores:	Administrador
Cenário principal:	
1.	O caso de utilização inicia-se quando o administrador pretende avaliar o

	sistema.
2.	O administrador selecciona o botão “Avaliação”.
3.	O sistema apresenta na interface a indicação do resultado da operação e o nome do ficheiro no qual os resultados serão colocados.
4.	O caso de utilização termina.

Alterar Configuração

Nome:	**Alterar Configuração**
Descrição:	Este caso de utilização permite alterar a configuração do sistema.
Pré-condições:	
Actores:	Configurador
Cenário principal:	
1.	O caso de utilização inicia-se quando o configurador pretende alterar a configuração do sistema.
2.	O configurador abre o ficheiro de configuração e altera a configuração deseja. Grava as alterações e fecha o ficheiro.
3.	O configurador deve reiniciar o serviço do sistema biométrico.
4.	O caso de utilização termina.