

• U



C •



Diogo João Machado Garcia

*Modelação de expressões faciais
num avatar usando classificação
de bioassinais*

• U



C •



UNIVERSIDADE DE COIMBRA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES

Modelação de expressões faciais num avatar usando classificação de biossinais

Diogo João Machado Garcia

Coimbra, 2014

Modelação de expressões faciais num avatar
usando classificação de biossinais

Orientadores:

Professor Doutor Urbano José Carreira Nunes
Doutor Gabriel Pereira Pires

Júri:

Professor Doutor Vítor Manuel Mendes da Silva
Professor Doutor Paulo José Monteiro Peixoto
Professor Doutor Urbano José Carreira Nunes

Diogo João Machado Garcia

Departamento de Engenharia Eletrotécnica e de Computadores

Faculdade de Ciências e Tecnologia, Universidade de Coimbra

Setembro, 2014

Agradecimentos

Nesta mais recente etapa da minha formação, quero agradecer a todos os que me apoiaram, de alguma forma, para atingir os meus objetivos.

Um especial agradecimento aos meus orientadores, Professor Doutor Urbano Nunes e Doutor Gabriel Pires, pela qualidade da orientação científica, pela motivação e pela confiança que depositaram em mim para desenvolver este trabalho.

Um obrigado à aluna de Doutoramento Aniana Brito, por me ter ajudado na integração de assuntos de biomédica, pela ajuda no desenvolvimento dos algoritmos e pelos diversos testes dos sistemas.

Obrigado, também, às pacientes cobaias de testes Ivo Baptista e João Perdiz.

Por último, mas não menos importante, quero agradecer aos meus pais, por me terem proporcionado todas as condições para concluir o curso de engenharia, ao meu irmão e à minha namorada por todo o apoio.

Esta dissertação foi realizada no âmbito dos projetos "RECI/EEI-AUT/0181/2012 - AMS-HMI12: Apoio à Mobilidade Suportada por Controlo Partilhado e Interfaces Homem-Máquina Avançados" e "Centro-07-ST24-FEDER-002028: Diagnosis and Assisted Mobility for People with Special Needs" financiados pela Fundação para a Ciência e Tecnologia (FCT), FEDER, e programas QREN e COMPETE.

Resumo

O reconhecimento do estado afetivo ou emocional dos humanos, pelas máquinas, assunto que se explora já há muito tempo, pretende melhorar a interação entre máquinas/computadores e humanos, adaptando-se a diferentes situações. As expressões faciais são uma das formas que permite a inferência de estados emocionais. Existem várias técnicas de reconhecimento automático de expressões faciais, que passam por técnicas baseadas em processamento de imagem ou vídeo e por utilização de biossinais, mais especificamente sinais eletromiográficos.

Nesta tese, foi investigada a utilização de biossinais, nomeadamente sinais eletromiográficos e sinais eletro-oculográficos, para inferir expressões faciais e movimentos oculares. Os músculos faciais permitem detetar expressões, tais como “felicidade”, “tristeza”, “surpresa”, “raiva”, entre outras. A informação do olhar é também relevante, tanto para a criação de novas interfaces de interação homem-máquina como para a dedução de estados emocionais e sociais. Informações como: olhar para cima, olhar para baixo, desviar o olhar são úteis para este fim. Existem várias técnicas de seguimento ocular e também de deteção de piscar de olhos, nomeadamente as que se baseiam no processamento de imagem, na oculografia de infravermelhos, na técnica da bobina escleral e nos biossinais eletro-oculográficos.

A fim de mimetizar as expressões faciais efetuadas pelo utilizador, bem como movimentos oculares e movimentos da cabeça (através de um sensor de inércia) foi desenvolvido um avatar 3D. O sistema de deteção incorpora três subsistemas; o primeiro deteta entre quatro expressões faciais (neutra, felicidade, tristeza e raiva), o segundo sistema deteta entre cinco movimentos oculares (“cima”, “baixo”, “direita”, “esquerda” e “piscar de olhos”) e o movimento de “franzir”. Ainda foi desenvolvido um terceiro sistema de seguimento ocular de movimentos horizontais (*eye-tracking*) recorrendo também a sinais eletro-oculográficos. Os dois primeiros sistemas recorrem a um classificador discriminativo linear (classificador de Bayes). Todos estes sistemas foram implementados para funcionamento *online*. O sistema de reconhecimento de expressões faciais obteve desempenho bastante satisfatório, com uma média de exatidão, entre expressões, de 95,5%, sendo a expressão “tristeza” algumas vezes confundida pelo classificador com a expressão “raiva” (algo que se refletiu no desempenho global do sistema). Nos testes efetuados, o sistema de reconhecimento de movimentos oculares obteve uma taxa de sensibilidade de 84,7% e uma taxa de exatidão de 98,3%. O sistema de seguimento ocular, que foi desenvolvido, apresentou um ótimo desempenho, no seguimento de treze sacadas, e um erro de aproximadamente 3,0 graus na posição final.

Palavras-chave: eletro-oculografia, eletromiografia, expressão facial, emoção, seguimento ocular, avatar

Abstract

The recognition of the human affective and emotional state by machines is a subject that has been researched for a long time. This allows for better and more suitable interaction between machines/computers and humans. Facial expressions are one of the ways that allows the inference of emotional states. Various techniques are available to automate facial expressions recognition, ranging from techniques based on the image and video processing to the use of biosignals, more precisely electromyographic signals. In this thesis, the use of biosignals, in particular electromyographic signals and electrooculographic signals, was researched to deduce facial expressions and eye movements. The facial muscles allow the detection of expressions such as “happiness”, “sadness”, “surprise”, “anger”, among other expressions. Eye information is also relevant, not only for the creation of new human-machine interfaces, but also for depicting social and emotional states. Look in up, look in down or look away are useful information for this purpose. There are several techniques for eye tracking and detecting blinks of the eye: based on image processing, infrared oculography, scleral coil and based on electrooculographic biosignals.

In order to mimic the user facial expressions, as well as eye tracking and head movements (through a motion sensor), a 3D avatar was developed. The main detecting system incorporates three subsystems. The first subsystem detects four facial expressions (neutral, happiness, sadness and anger), the second subsystem detects five eye movements (up, down, right, left and blink of the eye) and frown movement. The third subsystem is an eye-tracking system that also uses electrooculographic signals. The first two subsystems use a linear discriminative classifier (Bayes classifier). All these subsystems were developed for online usage. The facial expressions recognition system had a positive performance, with an average accuracy of 95,5%, where the sad expression was sometimes confused with anger expression (that affected the system global performance). In the tests carried out, the eye-movement recognition system had a sensitivity of 84,7% and accuracy of 98,3%. The developed eye-tracking system showed great performance when following 13 saccades, and a final position error of approximately 3,0 degrees.

Key words: electrooculography, electromyography, facial expression, emotion, eye-tracking, avatar

Conteúdo

Agradecimentos	i
Resumo	iii
Abstract	v
Lista de Figuras	ix
Lista de Tabelas	xi
Abreviaturas	xiii
1 Introdução	1
1.1 Motivação e contexto	1
1.2 Implementações e contribuições chave	1
2 Estado da arte	3
2.1 Expressões faciais	3
2.1.1 Facial Action Coding System (FACS)	3
2.1.2 Reconhecimento automático de expressões faciais	3
2.1.3 Detecção de expressões faciais com base em processamento de imagem	4
2.1.4 O EMG no reconhecimento de expressões faciais	4
2.1.5 Caraterização de emoções a partir de expressões faciais	7
2.2 Movimentos oculares	9
2.2.1 Principais técnicas de seguimento ocular	10
2.2.2 Reconhecimento de movimentos oculares utilizando EOG	11
3 Conceitos neurofisiológicos e métodos	15
3.1 Conceitos neurofisiológicos	15
3.1.1 Eletro-oculografia (EOG)	15
3.1.1.1 Tipos de movimentos oculares	16
3.1.1.2 <i>Baseline drift</i>	16
3.1.2 Eletromiografia (EMG)	16
3.2 Classificador de Bayes	19
3.3 Extração de caraterísticas de biossinais EOG e EMG	21
3.3.1 As caraterísticas do sinal fEMG	22
3.3.2 As caraterísticas do sinal EOG	23
4 Arquitetura e <i>setup</i> experimental	25
4.1 Sistemas de aquisição de biossinais	25
4.2 <i>Inertial Motion Unit</i>	27

4.3	Framework de processamento e classificação de biossinais	28
4.4	Vizard	28
4.5	Autodesk Maya	30
5	Reconhecimento de expressões faciais	33
5.1	Classificador EOG	33
5.1.1	Treino do classificador	33
5.1.2	Pré-processamento	34
5.1.3	Deteção assíncrona de movimento: <i>sliding window</i>	34
5.1.4	Extração de caraterísticas	36
5.1.5	Classificação do movimento	36
5.2	Classificador EMG	38
5.2.1	Treino do classificador	38
5.2.2	Pré-processamento	40
5.2.3	Deteção de expressão facial	40
5.2.4	Extração de caraterísticas	40
5.2.5	Classificação	42
5.3	Seguimento ocular	42
5.3.1	Pré-processamento	42
5.3.2	Deteção de sacada	42
5.3.3	Cálculo da amplitude da sacada	43
6	Testes e resultados experimentais	45
6.1	Sistema de movimentos oculares	45
6.1.1	Funcionamento <i>offline</i>	45
6.1.2	Funcionamento <i>online</i>	45
6.2	Sistema de reconhecimento de expressões faciais	46
6.3	Sistema de seguimento ocular	47
7	Conclusões e trabalho futuro	49
	Bibliografia	51

Lista de Figuras

1.1	Diagrama do trabalho desenvolvido	2
2.1	Estrutura dos sistemas de reconhecimento de expressões faciais baseados em processamento de imagem.	4
2.2	Sinais EMG dos músculos masseter, levantador do lábio superior e corrugador do supercílio na execução das expressões: (a) “Felicidade”; (b) “Raiva”; (c) “Tristeza”.	5
2.3	Posicionamento dos elétrodos, no trabalho de [Gibert et al., 2009]. Músculos: 1) <i>Venter frontalis</i> ; 2) <i>Corrugator supercilii</i> ; 3) <i>Orbicularis oculi</i> ; 4) <i>Levator labii</i> ; 5) <i>Zygomaticus major</i> ; 6) <i>Masseter</i> ; 7) <i>Depressor anguli oris</i> ; 8) <i>Mentalis</i>	6
2.4	Gráfico representativo do espaço emocional (adaptado de [Stangor, 2014]).	8
2.5	Principais técnicas de seguimento ocular: (a) <i>Setup</i> do sistema recorrendo a processamento de vídeo [Ohno and Mukawa, 2003]; (b) Bobina escleral (retirado de [Haslwanter, 2012]); (c) Óculos com emissor de radiação infravermelha [Johns et al., 2007].	10
2.6	Cancelamento do <i>baseline drift</i> e remoção do ruído proposto em [Pettersson et al., 2013]. A) Sinal EOG original (EOG vertical a vermelho e EOG horizontal a azul); B) Remoção do <i>baseline drift</i> ; C) Remoção de ruído de alta frequência.	12
2.7	Método de análise de sinais EOG, proposto em [Barea et al., 2002].	12
2.8	Sinal vertical obtido com filtro passa-alto de 1 Hz. PAP, VAP, UL e LL são características extraídas do sinal, no trabalho de [Aungsakun et al., 2011].	12
2.9	<i>Goggles</i> para captação de sinais EOG [Bulling et al., 2009]	12
3.1	Rotação horizontal do olho de 45 graus	15
3.2	Montagem típica dos elétrodos, na recolha de sinais EOG [Barea et al., 2002].	17
3.3	Sinal EOG típico, obtido durante a leitura de texto (sacada para a direita, marcada com “r”, e sacada para a esquerda, marcada com “L”). [Bulling, 2010]	17
3.4	Sinal eletromiográfico do músculo triceps direito, no levantamento de um supino de 2,3 Kg [Sarver, 2001].	18
3.5	Ilustração dos músculos faciais [Boot, 2009]	18
3.6	Posicionamento dos elétrodos para leitura de sinais EMG dos músculos faciais	19
3.7	Estrutura do algoritmo do classificador de Bayes multi-classe	21
4.1	Diagrama dos módulos desenvolvidos	25
4.2	Fotografia do <i>setup</i> experimental do sistema EOG. (1) Elétrodos; (2) IMU; (3) Avatar; (4) g.MOBILab+.	26
4.3	Fotografia do <i>setup</i> experimental do sistema de seguimento ocular. (1) Elétrodos; (3) Avatar; (5) g.USBamp.	26
4.4	Disposição do sistema de aquisição de dados, utilizando o bioamplificador g.MOBILab+ com os canais vertical e horizontal.	27

4.5	Disposição do sistema de aquisição de dados com os canais do corrugador do supercílio e do zigomático maior.	27
4.6	Disposição do sistema de aquisição de dados, utilizando o bioamplificador g.USBamp com o canal horizontal.	28
4.7	Diagrama de blocos Simulink do sistema EOG: (a) Diagrama completo; (b) Bloco de pré-processamento.	29
4.8	Diagrama de blocos Simulink do sistema de seguimento ocular. (1) Bloco que devolve sinais adquiridos pelo g.USBamp; (2) Diferença entre os sinais dos dois elétrodos; (3) Bloco onde os biossinais são processados (algoritmo da secção 5.3); (4) Sinal da posição dos olhos.	30
4.9	Script do Vizard em execução	30
4.10	Constituição do avatar	32
4.11	<i>Morphs</i> das expressões faciais: (a) “Neutro”; (b) “Felicidade”; (c) “Tristeza”; (d) “Raiva”; (e) “Piscar de olhos”; (f) “Franzir”.	32
5.1	Estrutura do sistema de reconhecimento EOG	35
5.2	Sequência de movimentos de uma sessão de treino do classificador EOG	35
5.3	Sinal EOG vertical. Janela deslizante a detetar o movimento ocular para cima, com o máximo do sinal posicionado no 4 ^o sub-intervalo.	36
5.4	Sinais EOG (a) vertical e (b) horizontal para uma sequência de seis movimentos dos olhos, gravados durante uma sessão de treino: “piscar”, “direita”, “esquerda”, “franzir”, “cima” e “baixo”. Os espaços preenchidos a branco representam os períodos de repouso (não têm qualquer influência no treino do classificador)	37
5.5	Estrutura do sistema de reconhecimento EMG	38
5.6	Sequência de expressões faciais de uma sessão de treino do classificador EMG	39
5.7	Sinal EMG do músculo corrugador do supercílio seccionado em duas épocas	39
5.8	Sinal EMG do músculo zigomático maior. Este é um exemplo em que a janela deslizante não permite a classificação, porque está a acontecer a transição entre as expressões “smile” e “neutral”.	41
5.9	Sinais EMG: (a) do músculo corrugador do supercílio e (b) do músculo zigomático maior, para uma sequência de quatro expressões faciais gravadas durante uma sessão de treino: “neutro”, “feliz”, “triste” e “raiva”. Os espaços preenchidos a branco representam os períodos de repouso (não têm qualquer influência no treino do classificador)	41
5.10	Estrutura do algoritmo de seguimento ocular horizontal.	42
5.11	Sinal EOG com componente DC e respetiva derivada. Abaixo é apresentada a janela deslizante. Sinal sem filtro Savitzky-Golay, a azul; Sinal filtrado, a vermelho; Derivada do sinal, a preto.	44
6.1	Fotografia das expressões realizadas pelo autor da tese e mimetizações do avatar: (a) expressão neutra, (b) expressão de felicidade, (c) expressão de tristeza e (d) expressão de raiva.	46
6.2	Teste <i>online</i> do sistema de seguimento ocular: (a) Sinal EOG; (b) Posição angular obtida.	48

Lista de Tabelas

2.1	Exemplos de <i>Action Units</i> [Ekman and Friesen, 1978a]	6
2.2	Caraterização das seis expressões faciais básicas	7
2.3	Literatura sobre técnicas de deteção de expressões faciais e emoções	9
2.4	Literatura de técnicas de captura de movimentos oculares e seguimento ocular.	13
4.1	Identificação dos blocos do Simulink do sistema EOG	29
6.1	Matriz de confusão de classificação <i>offline</i> de movimentos oculares.	45
6.2	Desempenho <i>offline</i> do sistema de reconhecimento de movimentos oculares.	45
6.3	Performance da classificação <i>online</i> de cada sujeito.	46
6.4	Matriz de confusão do sistema de classificação de expressões faciais (em modo de funcionamento <i>offline</i>).	46
6.5	Matriz de confusão de um teste de classificação de expressões faciais.	47

Abreviaturas

3D	Tridimensional
A/D	Analog/Digital
API	Application Programming Interface
AU	Action Unit
DASDV	Difference Absolute Standard Deviation Value
DC	Direct Current
EEG	Eletroencefalograma/Eletroencefalografia
EMG	Eletromiografia/Eletromiográfico
EOG	Eletro-oculografia/Eletro-oculográfico
FACS	Facial Action Coding System
fEMG	Eletromiografia facial
HCI	Human Computer Interface
IMU	Inertial Motion Unit
ISR	Institute of Systems and Robotics
LDA	Linear Discriminant Analysis
MVA	Média dos Valores Absolutos
PPST	Porcentagem de pulsos superiores ao threshold
RMS	Root Mean Square
SMI	SensoMotoric Instruments
SNR	Signal Noise Ratio

Capítulo 1

Introdução

1.1 Motivação e contexto

As expressões faciais são um dos meios mais poderosos e imediatos de comunicação de emoções e de intenções entre seres humanos. A face pode expressar emoções ou estados afetivos antes das pessoas os verbalizarem ou mesmo antes de se aperceberem dos sentimentos [Tian et al., 2001]. As emoções podem ser expressas através de seis expressões faciais — felicidade, tristeza, raiva, repugnância, surpresa e medo —; do som da voz; da postura física corporal; do batimento cardíaco; da respiração e, também, da tensão muscular.

A possibilidade de um sistema inteligente reconhecer o estado emocional do utilizador (“*Machine emotional intelligence*”), e de interagir de acordo com ele, é um tópico interessante que tem potencialidade para melhorar a relação entre os humanos e a máquina. Este é um assunto que, há muito tempo tem sido foco de atenções, por parte de cientistas e de investigadores, uma vez que a população se aproxima cada vez mais de um futuro, rodeada por computadores, máquinas e robôs na interação diária.

O estudo apresentado, nesta tese, foca-se na área de reconhecimento automático de expressões faciais recorrendo a informação extraída dos bio-sinais eletromiográficos (EMG) e eletro-oculográficos (EOG), uma área de estudo pioneira no ISR (Institute of Systems and Robotics). Um dos objetivos propostos pelo ISR é o desenvolvimento de um robô interativo, que tenha a perceção do estado afetivo dos utilizadores, pelo que este trabalho visa também oferecer um contributo nessa tarefa.

1.2 Implementações e contribuições chave

Os objetivos propostos, no início da realização da tese, são, nomeadamente:

- Desenvolvimento de sistema de deteção de movimentos oculares;
- Desenvolvimento de sistema de reconhecimento de expressões faciais;
- Conceção de um avatar que mimetiza expressões faciais e movimentos oculares.

Para além destes módulos foi possível também implementar um sistema de seguimento ocular, usando sinais EOG, onde foi detetada a posição angular do olho.

O trabalho desenvolvido e descrito nos próximos capítulos, está apresentado, em forma de diagrama, na figura 1.1. Foi modelado um avatar 3D (que mimetiza várias expressões faciais e movimentos oculares) (secção 4.5), e desenvolvido um programa que controla o avatar (secção 4.4), que está representado a verde na figura (módulo designado por “Avatar”). Na área amarela (módulo designado por “Processamento de bio-sinais”) estão representados os algoritmos desenvolvidos de reconhecimento automático

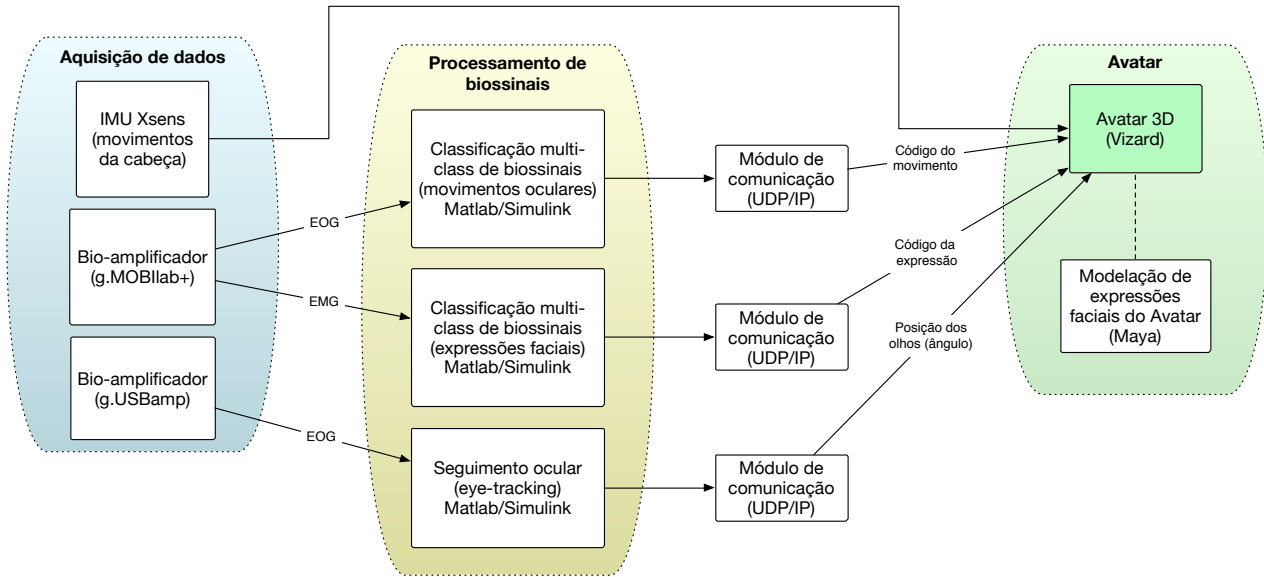


Figura 1.1: Diagrama do trabalho desenvolvido

de expressões faciais, movimentos oculares e seguimento ocular (secção 3.3 e capítulo 5). Ainda foi integrado um IMU (*Inertial Motion Unit*) no sistema do avatar, que capta a pose da cabeça do utilizador, permitindo a mimetização da pose (secção 4.2).

No capítulo 2, Estado da arte, é feita uma revisão da literatura de sistemas de reconhecimento automático de expressões faciais, com foco especial nos sistemas que utilizam sinais EMG, e de sistemas de movimentos oculares (fazendo maior incidência nos sistemas que utilizam sinais EOG). O capítulo 3 resume os conceitos neurofisiológicos sobre sinais EOG e EMG, descreve os métodos utilizados na extração de características, bem como os métodos de classificação dos biossinais. No capítulo 4, é apresentada uma explicação geral dos sistemas implementados e do *framework* de desenvolvimento. Nos capítulos 5 e 6, é feita uma explicação detalhada dos algoritmos de reconhecimento implementados e uma análise dos resultados dos testes efetuados.

Do trabalho desenvolvido, resultaram contribuições incluídas num artigo intitulado “Facial expression recognition based on EOG toward emotion detection for human-robot interaction”, submetido a conferência.

Capítulo 2

Estado da arte

2.1 Expressões faciais

As expressões faciais são um modo de comunicação não verbal, gerado pelos movimentos dos músculos faciais, conhecidas por transmitir emoções. As expressões faciais podem ser voluntárias ou involuntárias, sendo as últimas são responsáveis por inferir o estado sentimental do ser humano. Darwin concluiu que as expressões faciais são iguais para todos os seres humanos, independentemente das diferenças culturais [Ekman, 1973]. Deste modo, as expressões faciais não são aprendidas, mas sim de origem biológica do ser humano [Ekman et al., 1999, Ekman, 1973].

Um estudo recente, feito pela Universidade do estado de Ohio, indica que existem vinte e uma expressões distintas [Du et al., 2014]. No entanto, a grande parte das investigações nesta área incide sobre seis expressões consideradas básicas: felicidade, tristeza, raiva, repugnância, surpresa e medo (*happiness, sadness, anger, disgust, surprise, fear*). Deve salientar-se que as seis expressões indicadas são, na verdade, estados emocionais, cada um deles com uma expressão facial associada.

2.1.1 Facial Action Coding System (FACS)

Desde os primeiros estudos sobre expressões faciais até ao ano de 1977, a maioria das investigações eram feitas a partir da observação dos sujeitos e da análise descritiva. Este tipo de métodos é ambíguo (varia entre os observadores) e, portanto, havia carência de um sistema que padronizasse a forma como são classificadas as expressões. Para descrever e classificar expressões faciais com exatidão, Ekman and Friesen desenvolveram o Facial Action Coding System (FACS) [Ekman and Friesen, 1978b]. Este sistema usa uma abordagem, baseada nas ações dos músculos da face, envolvendo a identificação dos vários músculos que, individualmente ou em grupos, manifestam alterações visíveis na face. Estas alterações ao nível facial, bem como os músculos (ou grupos de músculos) responsáveis por essas alterações, têm o nome de *Action Units* (AU). O FACS é constituído por várias *Action Units*. Alguns exemplos de AU's estão apresentados na tabela 2.1. Existem, também, *Action Units* que não são geradas a partir de músculos faciais, como os casos das AU's 19 e 33 (consultar tabela 2.1). A lista completa das AU's pode ser consultada em [Ekman et al., 1978].

2.1.2 Reconhecimento automático de expressões faciais

Com o avanço na área da robótica, mais especificamente nos robôs humanóides, o desenvolvimento de técnicas de reconhecimento de expressões faciais tornou-se evidente. Os robôs tendem a interagir cada vez mais com as pessoas, seja nos locais de trabalho ou nas habitações. Por esse motivo, os robôs necessitam de se tornar mais inteligentes, no que toca à perceção do estado emocional dos humanos que os rodeiam [Bettadapura, 2012].

O reconhecimento de expressões faciais tem também um papel importante na inferência de afetos para construir interfaces HCI (*Human Computer Interface*) verdadeiramente adaptadas ao utilizador. Esta é uma área de investigação chamada *Affective Computing* [Zeng et al., 2009].

Para além das áreas da robótica e das interfaces HCI, os sistemas de reconhecimento de expressões faciais têm aplicações em diversas áreas como os vídeo-jogos, animações (cinema), psiquiatria, segurança automóvel, *juke boxes* e televisões sensíveis ao estado emocional do utilizador, software de educação, entre outras [Bettadapura, 2012].

Existem imensas técnicas de reconhecimento de expressões faciais, mas que se organizam principalmente em dois grupos: técnicas baseadas em processamento de imagem e técnicas baseadas em sinais eletromiográficos faciais (fEMG).

2.1.3 Deteção de expressões faciais com base em processamento de imagem

Existem imensas técnicas desenvolvidas, no âmbito do reconhecimento de expressões, utilizando processamento de imagem. A maioria das técnicas segue um processo que começa com a deteção da face (*face tracking*), passando para a extração das características (*feature extraction*) para, posteriormente, proceder à classificação da expressão facial (ver figura 2.1).

O trabalho desenvolvido por Kotsia e Pitas, em 2007, demonstra um sistema automático (*online*) de reconhecimento de expressões com uma taxa de performance de 99,7%, na classificação das seis expressões faciais básicas, e uma taxa de 95,1% na classificação de 17 *Action Units* faciais [Kotsia and Pitas, 2007]. A maioria dos sistemas de reconhecimento de expressões, baseados em processamento de imagem, são testados com bases de dados de imagens em que a face é totalmente visível. A oclusão parcial da face constitui um problema neste tipo de sistemas. Em 2008, foi desenvolvido e testado um método de reconhecimento de expressões, que obteve resultados de performance de 91,4% na deteção das seis expressões faciais consideradas básicas, onde estão ocultas algumas áreas da face [Kotsia et al., 2008]. Numa abordagem um pouco diferente, Y. Tian, T. Kanade e J. F. Cohn apresentaram um sistema de reconhecimento de *Action Units*, ao invés de expressões faciais. No processo de classificação de 13 AU's superiores (zona dos olhos/testa) e de 21 AU's inferiores (zona da boca/queixo) foram obtidas performances de 96,4% e de 96,7%, respetivamente [Tian et al., 2001].

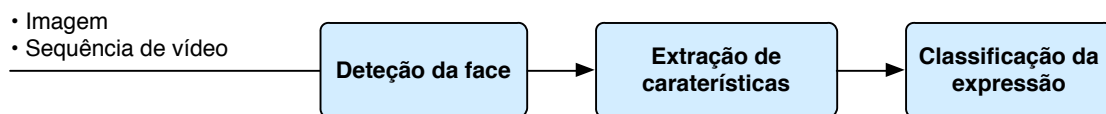


Figura 2.1: Estrutura dos sistemas de reconhecimento de expressões faciais baseados em processamento de imagem.

2.1.4 O EMG no reconhecimento de expressões faciais

Fridlund e Cacioppo tiveram um papel importante no estudo dos sinais EMG faciais. O artigo, escrito em 1986 [Fridlund and Cacioppo, 1986], é, ainda hoje, utilizado pelos investigadores no estudo dos sinais EMG. Fridlund e Cacioppo descreveram técnicas para uma leitura eficaz dos sinais, identificaram as principais fontes de ruído, as soluções para mitigar as mesmas, indicações para o posicionamento correto dos elétrodos e técnicas para análise dos sinais (ver figuras 3.5 e 3.6).

Existe uma vasta literatura no reconhecimento automático de expressões faciais a partir de fEMG. Em 2004, foi proposto um método de classificação automática (*offline*) de expressões, utilizando apenas três pares de elétrodos (nos músculos masseter, corrugador do supercílio, e levantador do lábio superior) onde se conseguiu uma taxa de sucesso na classificação de 94,44% [Ang et al., 2004]. As expressões classificadas foram: felicidade, raiva e tristeza. Neste processo foram extraídas quatro características do sinal EMG — média, desvio padrão, RMS e potência espectral. Na figura 2.2, podem ser vistos exemplos dos sinais obtidos de cada músculo para as três expressões faciais. A combinação de EMG com EEG foi também investigada, no trabalho [Chin et al., 2008], para inferir seis expressões faciais básicas, alcançando uma classificação de 86%. Utilizando oito pares de elétrodos, foi apresentado um método, apenas utilizando sinais EMG faciais, que obteve 92,19% de taxa de reconhecimento (*online*) de seis expressões faciais (neutro, surpresa, tristeza, felicidade, raiva e repugnância)[Gibert et al., 2009]. O posicionamento dos elétrodos deste trabalho pode ser visualizado na figura 2.3. As seis expressões consideradas básicas são analisadas, na tabela 2.2, em termos de músculos, normalmente utilizados em cada expressão, bem como as *Action Units* associadas aos mesmos e os efeitos visíveis na face das pessoas. Para consultar a localização de cada músculo pode recorrer à figura 3.5 na secção 3.1.2.

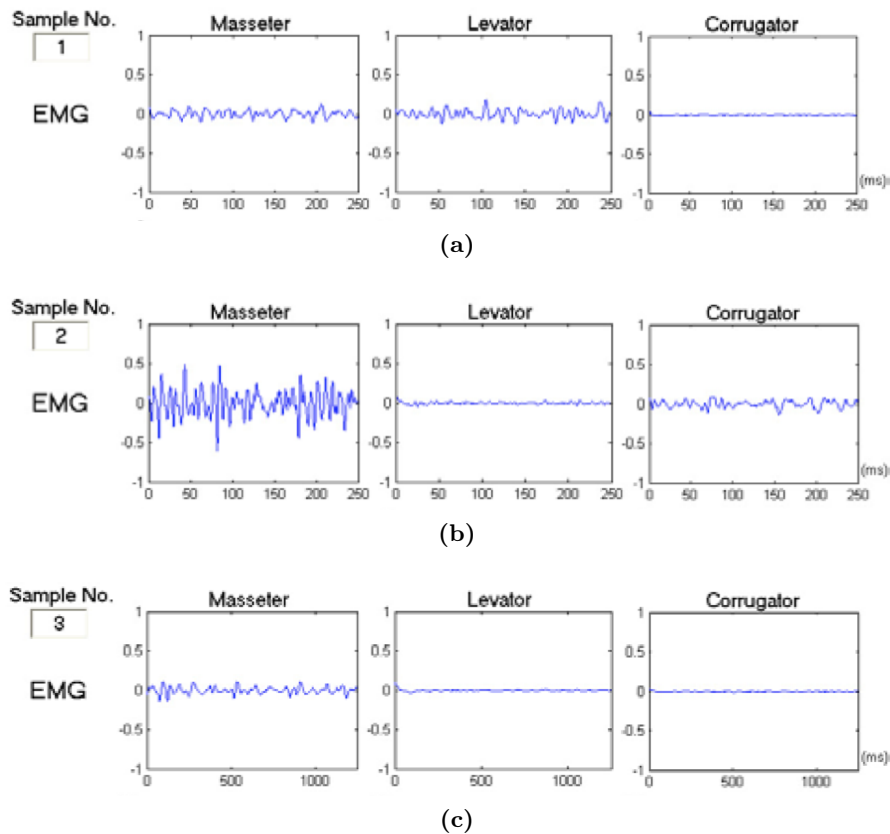


Figura 2.2: Sinais EMG dos músculos masseter, levantador do lábio superior e corrugador do supercílio na execução das expressões: (a) “Felicidade”; (b) “Raiva”; (c) “Tristeza”.

2.1. EXPRESSÕES FACIAIS

Tabela 2.1: Exemplos de *Action Units* [Ekman and Friesen, 1978a]




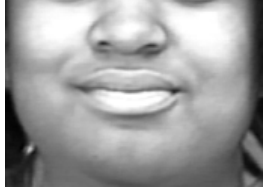
AU	Descrição	Músculos faciais	Imagem exemplar
1	Inner Brow Raiser	<i>Frontalis, pars medialis</i>	
2	Outer Brow Raiser	<i>Frontalis, pars lateralis</i>	
10	Upper Lip Raiser	<i>Levator labii superioris</i>	
12	Lip Corner Puller	<i>Zygomaticus major</i>	
19	Tongue Out	–	–
33	Cheek Blow	–	–



Figura 2.3: Posicionamento dos elétrodos, no trabalho de [Gibert et al., 2009]. Músculos: 1) *Venter frontalis*; 2) *Corrugator supercilii*; 3) *Orbicularis oculi*; 4) *Levator labii*; 5) *Zygomaticus major*; 6) *Masseter*; 7) *Depressor anguli oris*; 8) *Mentalis*.

Tabela 2.2: Caracterização das seis expressões faciais básicas

Expressão facial	Músculos envolvidos	Ação produzida	Action Units
Felicidade/ Sorriso	· Orbicular do olho · Zigomático maior	Fecho das pálpebras, deslocamento dos cantos da boca para cima e lateralmente e elevação das bochechas	6+7+12
Surpresa	· <i>Frontalis</i> · Levantador da pálpebra superior · Masseter	Levantamento das sobrancelhas e da pálpebra superior, queixo caído	1+2+5+26
Medo	· <i>Frontalis</i> · Corrugador do supercílio · Levantador da pálpebra superior	Levantamento das sobrancelhas (parte lateral) e da pálpebra superior, abaixamento da parte interior das sobrancelhas	2+4+5
Raiva	· Corrugador do supercílio · Levantador da pálpebra superior · Orbicular do olho · Orbicular da boca	Abaixamento das sobrancelhas, elevando a pálpebra superior, fechando as pálpebras e apertando os lábios	4+5+7+23
Tristeza	· <i>Frontalis</i> · Corrugador do supercílio · Abaixador do ângulo da boca	Levantamento das sobrancelhas (parte interior), juntando as duas sobrancelhas, e abaixamento dos cantos dos lábios	1+4+15
Repugnância	· Levantador do lábio superior · Levantador do lábio superior e da asa do nariz	Enrugamento da pele do nariz, depressão dos cantos da boca e elevação do lábio superior	9+10+15

2.1.5 Caracterização de emoções a partir de expressões faciais

O objetivo principal do reconhecimento de expressões faciais é a inferência de emoções nos humanos. O estado emocional pode ser caracterizado, utilizando dois fatores: a valência (*valence*) e a estimulação (*arousal*) [Partala et al., 2006]. O fator valência varia entre experiência emocional positiva e experiência emocional negativa, portanto no centro da dimensão valência está o estado emocional neutro. O fator estimulação varia entre o estado “pouco excitado” e “muito excitado”. A conjugação destes dois fatores chama-se espaço emocional (ver figura 2.4).

Com a leitura e a análise dos sinais eletromiográficos dos músculos zigomático maior e corrugador do supercílio, é possível inferir o estado afetivo de um indivíduo em tempo real com uma taxa de sucesso razoável [Partala et al., 2006, Schels et al., 2013]. O estado emocional de uma pessoa pode ser estimado, a partir de sistemas baseados em processamento de imagem, a partir do reconhecimento de expressões faciais. Não deve ser esquecido o facto das expressões faciais poderem ser voluntárias ou involuntárias, sendo que as últimas são as mais relevantes para a inferência dos estados emocionais. Este é um dos problemas principais dos dois tipos de sistemas apresentados. Na tabela 2.3 são apresentados alguns trabalhos relevantes de sistemas automáticos de reconhecimento de expressões faciais.

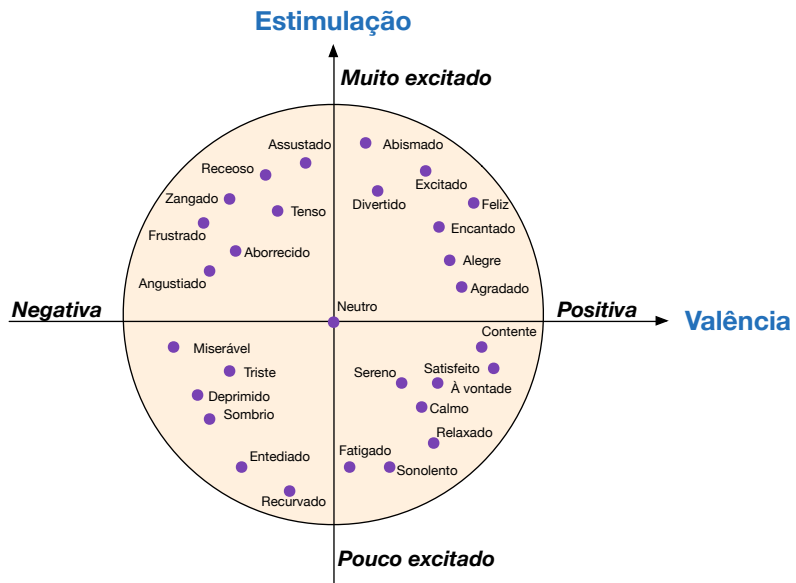


Figura 2.4: Gráfico representativo do espaço emocional (adaptado de [Stangor, 2014]).

Tabela 2.3: Literatura sobre técnicas de detecção de expressões faciais e emoções

Artigo	Tipo	Descrição
Face Expression Recognition and Analysis: The State of the Art [Bettadapura, 2012]	Processamento de imagem/vídeo	O artigo não retrata uma técnica em específico, mas faz a revisão de vários métodos de reconhecimento automático de expressões faciais com base em processamento de imagem. São comparados os trabalhos mais relevantes de reconhecimento de expressões desde o ano 2001. É um artigo muito completo e indispensável a novos investigadores no campo de reconhecimento de expressões faciais, com foco especial no processamento de imagem/vídeo.
Facial Expression Recognition through Pattern Analysis of Facial Muscle Movements Utilizing Electromyogram Sensors [Ang et al., 2004]	Eletromiografia (EMG)	Sistema que reconhece três expressões (“Happy”, “Sad” e “Angry”) a partir da leitura de sinais EMG dos músculos <i>masseter</i> , <i>corrugator supercilli</i> e <i>levator labii</i> . Utiliza o classificador do tipo <i>minimum-distance classifier</i> , com três características. A precisão atingida é de 92,44% (média dos cinco participantes).
Enhancement of Human Computer Interaction with facial Electromyographic sensors [Gibert et al., 2009]	Eletromiografia (EMG)	Utilizando oito sensores EMG (oito músculos faciais) atinge-se 92% de precisão no reconhecimento das seis expressões básicas. O trabalho destaca o reconhecimento mais rápido que os métodos baseados em imagem/vídeo, e a possibilidade de reconhecer expressões faciais mais subtis.
Real-time EEG-based Human Emotion Recognition and Visualization [Liu et al., 2010]	Eletroencefalografia (EEG)	Este trabalho retrata o reconhecimento online de seis emoções através de sinais de eletroencefalograma (EEG), reproduzindo as emoções num avatar 3D. As emoções são primeiro mapeadas no espaço emocional, em termos de valência e estimulação (figura 2.4) e depois traduzidas numa emoção propriamente dita.
Multiclass Voluntary Facial Expression Classification based on Filter Bank Common Spatial Pattern [Chin et al., 2008]	EMG/EEG	Este artigo investiga a classificação de seis expressões faciais voluntárias a partir de sinais EMG e EEG usando uma versão alterada do algoritmo <i>Filter Bank Common Spatial Pattern</i> , que permite a classificação multi-classe.
Real-time estimation of emotional experiences from facial expressions [Partala et al., 2006]	Eletromiografia (EMG)	É investigado um método que estima o estado emocional de um sujeito em tempo-real através da leitura de dois músculos: zigomático maior e corrugador do supercílio. Os participantes foram sujeitos a fotos e vídeos selecionados para induzir determinadas emoções e foram obtidos resultados de precisão de 70% e 80% para fotos e vídeos, respetivamente.

2.2 Movimentos oculares

O seguimento dos movimentos oculares e as expressões faciais, associadas aos olhos, podem ser úteis, tanto para a criação de novas interfaces de interação homem máquina, como para a deteção de estados emocionais e sociais. Movimentos como: piscar os olhos, olhar para cima, ou para baixo, fazer contacto visual com outra pessoa ou evitar o contacto visual, fixar o olhar, podem fornecer informação do estado emocional e afetivo de uma pessoa [Psychologist World, 2014].

2.2.1 Principais técnicas de seguimento ocular

Atualmente, existem quatro tipos de técnicas para monitorizar os movimentos oculares: através de processamento de imagem/vídeo, utilizando o método da “bobina escleral”, utilizando oculografia de infravermelhos e através do processamento de sinais eletro-oculográficos (EOG).

A deteção do seguimento ocular (*eye-tracking*) é, dos vários tipos de movimentos, aquele que tem merecido mais atenção. Tem, como principais aplicações, as interfaces homem-computador e, mais recentemente, o estudo do comportamento visual dos consumidores para marketing, a reação a múltiplas situações e na condução automóvel.

Os métodos de seguimento ocular, a partir de processamento de imagem/vídeo, apresentam atualmente um ótimo desempenho (a precisão, entre os melhores sistemas, ronda por volta de 1 grau). Estes sistemas não são intrusivos (geralmente o utilizador não precisa de nenhum tipo de objeto ligado ao corpo), permitem ao utilizador realizar movimentos com a cabeça de forma livre (ver figura 2.5a); podem, inclusivamente, detetar o piscar de olhos e estimar o diâmetro da pupila. As principais desvantagens deste tipo de sistema são a baixa taxa de frequência e a impossibilidade de funcionamento em ambientes com pouca luz.

O método da “bobina escleral” é o tipo de sistema com melhor desempenho ao nível de *eye-tracking*. Por outras palavras, é o sistema padrão para medir movimentos oculares [van der Geest and Frens, 2002]. O seu funcionamento baseia-se no posicionamento de uma bobina muito fina na esclera (parte branca do olho) e de pequenos ímanes à volta do olho. Essa bobina acompanha o movimento dos olhos e, com a medida da polaridade e a amplitude da corrente elétrica, gerada pela indução eletromagnética, pode ser estimado o ângulo e a direção da posição do olho (ver figura 2.5b). Este método é altamente intrusivo, podendo ser perigoso, devido a esse mesmo facto, pelo que o posicionamento da bobina deve ser extremamente cuidadoso.

Existe ainda um método de *eye-tracking*, que utiliza uma fonte de radiação infravermelha diretamente apontada ao olho, permitindo, através das reflexões da radiação, medir o deslocamento angular dos olhos com uma câmara de visão (ver figura 2.5a e 2.5c). Existem vários sistemas comerciais que usam esta técnica entre os quais se destacam os das empresas Tobii e SMI [Tobii, 2013, SMI,]. Utilizando este método, facilmente se consegue uma resolução a rondar 0.1 graus. Este tipo de técnica é amplamente utilizada em estudos comerciais e de marketing, ou seja, no estudo do comportamento visual dos consumidores e na indústria automóvel e médica.

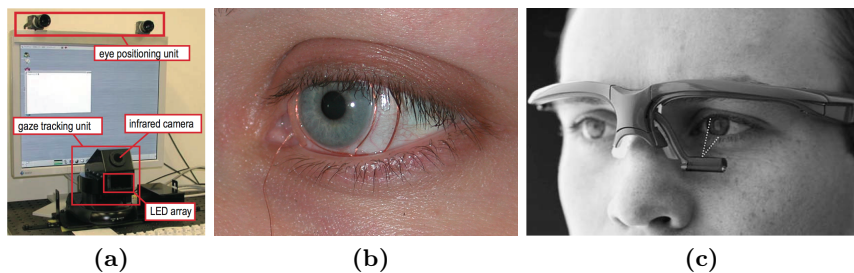


Figura 2.5: Principais técnicas de seguimento ocular: (a) *Setup* do sistema recorrendo a processamento de vídeo [Ohno and Mukawa, 2003]; (b) Bobina escleral (retirado de [Haslwanter, 2012]); (c) Óculos com emissor de radiação infravermelha [Johns et al., 2007].

2.2.2 Reconhecimento de movimentos oculares utilizando EOG

Uma característica, presente nos sinais EOG, é o *baseline drift* (consultar secção 3.1.1.2). As técnicas utilizadas, no reconhecimento de movimentos oculares, podem definir-se pela forma como tentam lidar com o *baseline drift* e, geralmente, seguem uma das três abordagens que são apresentadas de seguida.

A primeira abordagem consiste no tratamento do sinal sem remover a componente DC, utilizando técnicas de remoção do *baseline drift*. No trabalho desenvolvido por Pettersson et al., os sinais EOG vertical e horizontal são recolhidos dos respetivos canais, é feita a remoção da *baseline drift* e do ruído de alta frequência (ver figura 2.6)[Pettersson et al., 2013]. De seguida, é feita a derivada dos dois sinais, que permite a deteção de movimentos oculares aquando da existência de picos. Após a deteção de movimento, são extraídas as características dos sinais que permitem estimar o ângulo de deslocamento vertical e horizontal dos olhos.

A solução apresentada, por Barea et al., consiste na aplicação de um filtro passa-alto de 0,05 Hz aos sinais EOG vertical e horizontal, removendo assim o *baseline drift*, mas eliminando de igual forma a componente DC, característica dos sinais EOG [Barea et al., 2002]. Após esta fase, é efetuada a derivada do sinal, e uma remoção do ruído da mesma, podendo ser facilmente obtidos os movimentos oculares através da amplitude dos picos (figura 2.7).

A terceira abordagem é bastante diferente das duas anteriores. É aplicado um filtro passa-alto de 1 Hz aos sinais EOG vertical e horizontal, que remove completamente as componentes de baixa frequência dando origem a um sinal distinto dos apresentados anteriormente (figura 2.8). O processo de deteção de movimento é efetuado a partir de *thresholds* aplicados aos dois canais, que detetam movimentos verticais e/ou horizontais. Após a deteção de movimento ocular, é analisada uma janela temporal de 2 segundos de ambos os canais, onde são extraídas duas características de cada sinal. Por fim, o movimento é classificado num dos oito movimentos: “cima”, “baixo”, “esquerda”, “direita”, “cima+esquerda”, “cima+direita”, “baixo+esquerda” e “baixo+direita” [Aungsakun et al., 2011].

Algumas características, que a maioria dos métodos desenvolvidos para reconhecimento automático de movimentos oculares partilham, são a boa robustez no desempenho, na deteção de piscar de olho e na deteção de sacadas, no entanto o seu desempenho, na quantificação da posição angular vertical e horizontal do olho, é inferior quando comparado com as outras técnicas apresentadas no ponto 2.2.1.

A recolha de sinais EOG é feita a partir de elétrodos que têm de ser colocados e ajustados manualmente nos sítios corretos, sendo muitas vezes necessária a aplicação de um gel. Numa possível interface HCI, este processo é inviável, uma vez que existem restrições a nível da mobilidade. Para tentar colmatar este problema foram desenvolvidas algumas soluções, mais especificamente, foram desenvolvidos sistemas onde os elétrodos estão embutidos num tipo de óculos, chamados os *goggles* (figura 2.9), que se adaptam bastante bem a todo o tipo de rostos humanos [Barea et al., 2011, Bulling et al., 2009, Iáñez et al., 2013].

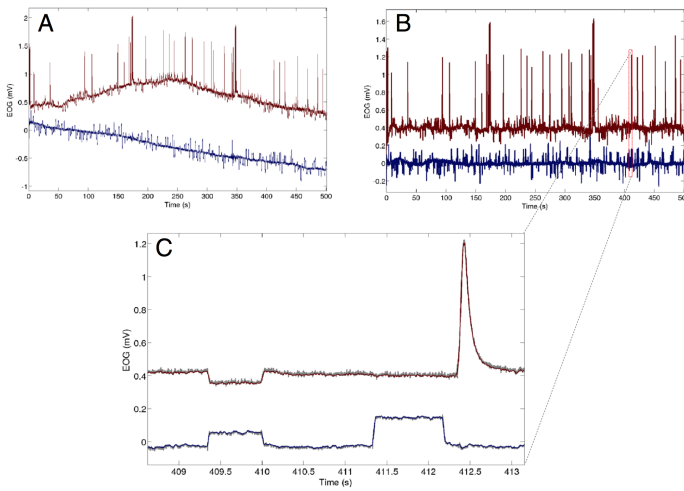


Figura 2.6: Cancelamento do *baseline drift* e remoção do ruído proposto em [Pettersson et al., 2013]. A) Sinal EOG original (EOG vertical a vermelho e EOG horizontal a azul); B) Remoção do *baseline drift*; C) Remoção de ruído de alta frequência.

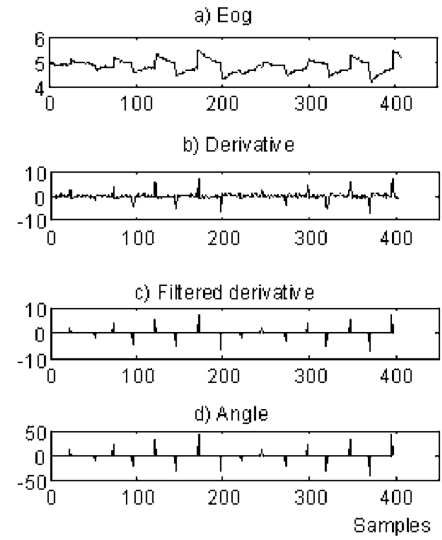


Figura 2.7: Método de análise de sinais EOG, proposto em [Barea et al., 2002].

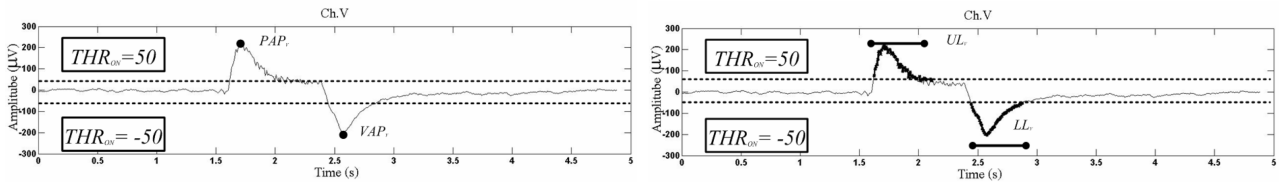


Figura 2.8: Sinal vertical obtido com filtro passa-alto de 1 Hz. PAP, VAP, UL e LL são características extraídas do sinal, no trabalho de [Aungsakun et al., 2011].

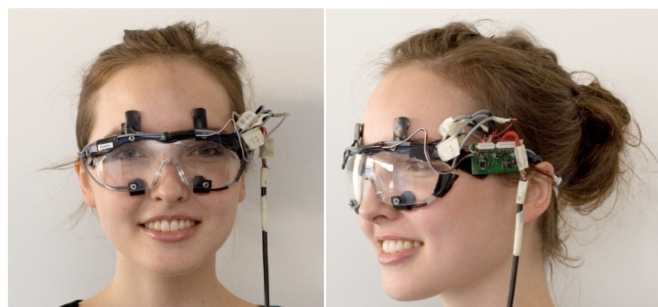


Figura 2.9: Goggles para captação de sinais EOG [Bulling et al., 2009]

Tabela 2.4: Literatura de técnicas de captura de movimentos oculares e seguimento ocular.

Artigo	Tipo	Descrição
A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction [Ohno and Mukawa, 2003]	Processamento de imagem/vídeo	Este artigo propõe um sistema de seguimento ocular, que consegue obter uma exatidão de 1,0 graus da posição dos olhos. Os pontos fortes são: calibração do sistema rápida e fácil, não exigência de acoplamento de qualquer dispositivo no utilizador e a possibilidade do utilizador se poder mover livremente, sem restrições.
A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field [Robinson, 1963]	Método da “bobina escleral”	Artigo que explica com elevado detalhe o método da bobina escleral — o <i>gold standard</i> das técnicas de seguimento ocular.
Evaluating Feature Extraction Methods of Electrooculography (EOG) Signal for Human-Computer Interface [Aungsakul et al., 2012]	Eletro-oculografia (EOG)	Neste trabalho são utilizados sinais EOG com filtro passa-banda de 1-500 Hz para remoção do <i>baseline drift</i> . São testadas performances de várias características — PAV (<i>peak amplitude value</i>), VAV (<i>valley amplitude value</i>), PAP (posição do PAV), VAP (posição do VAV), AUC (<i>area under curve</i>), TCV (<i>number of threshold crossing value</i>) e VAR (variância do sinal EOG)— na classificação de oito movimentos oculares: “cima”, “baixo”, “esquerda”, “direita”, “cima+esquerda”, “cima+direita”, “baixo+esquerda” e “baixo+direita”. Tem a vantagem de ser uma técnica com robustez elevada.
Wheelchair Guidance Strategies Using EOG [Barea et al., 2002]	Eletro-oculografia (EOG)	Barea fez várias contribuições na área do EOG. Neste artigo propõe um método para controlar uma cadeira de rodas, direcionado a pessoas com mobilidade reduzida. A técnica proposta passa por aplicar um filtro passa-alto de 0,05 Hz, para remover o <i>baseline drift</i> , e depois fazer a diferenciação do sinal obtido. A amplitude dos picos da derivada é medida, significando sacadas com determinado ângulo de deslocamento.
Automatic Artefact Compensation in EOG Signals [Bulling et al., 2007]	Eletro-oculografia (EOG)	Andreas Bulling apresenta neste artigo uma <i>framework</i> que visa a remoção dos vários artefactos que afetam o sinal EOG, para a obtenção da posição angular dos olhos. Os resultados apresentados na forma de gráfico são promissores.

Capítulo 3

Conceitos neurofisiológicos e métodos

Neste capítulo são apresentados os conceitos neurofisiológicos sobre a eletro-oculografia e a eletromiografia, bem como os métodos de classificação e extração de características usadas para a sua detecção.

3.1 Conceitos neurofisiológicos

3.1.1 Eletro-oculografia (EOG)

A eletro-oculografia é um método utilizado para detectar o movimento ocular. O sinal EOG é obtido através da medida da diferença de potencial córneo-retinal. O comportamento do olho pode ser modelado como um dipolo, em que a córnea é o polo positivo e a retina é o polo negativo. Na figura 3.1 está representado o olho em forma de dipolo efetuando uma rotação de 45 graus, traduzindo-se num sinal EOG ideal representado no gráfico da direita.

Para medir sinais EOG, são normalmente utilizados dois pares de elétrodos: um par para medir movimentos verticais e outro par para medir movimentos horizontais, que na figura 3.2 são os canais B e C — par de elétrodos dos movimentos verticais — e D e E — par correspondente para movimentos horizontais. Ao movimentar os olhos em diferentes direções a córnea (+) vai aproximar-se de um elétrodo (do par vertical ou horizontal), enquanto a retina (-) se aproxima do outro elétrodo desse par, sendo possível medir uma diferença de potencial e estimar, a partir de algoritmos, o deslocamento do olho, tanto vertical como horizontal.

O sinal EOG apresenta, tipicamente, amplitudes de $5 \mu\text{V}/\text{grau}$ a $20 \mu\text{V}/\text{grau}$ em que a potência útil permanece entre 0 Hz e 30 Hz [Bulling et al., 2011].

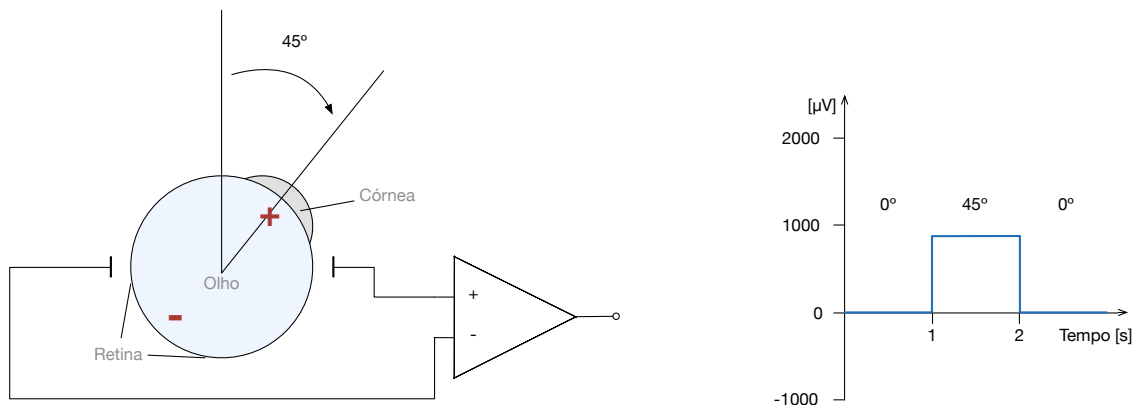


Figura 3.1: Rotação horizontal do olho de 45 graus

3.1.1.1 Tipos de movimentos oculares

Existem três tipos diferentes de movimentos oculares: as sacadas, as fixações e os piscar de olhos.

Sacadas

Quando estamos a observar um determinado cenário, os nossos olhos não estão fixos, estão em constante movimento, para tentarem visualizar com detalhe todos os pontos de interesse do cenário. Isto acontece, porque apenas uma parte dos nossos olhos (chamada fóvea) consegue ver com detalhe [Bulling et al., 2011]. Designa-se por sacada o movimento simultâneo dos dois olhos. A duração de uma sacada depende do ângulo efetuado pelos olhos, no entanto uma rotação de 20 graus tem uma duração entre 10 ms a 100 ms [Duchowski, 2007].

Fixações

As fixações são os momentos existentes entre sacadas, onde não existe movimento por parte dos olhos, ou seja, quando o olhar está fixado num determinado ponto do cenário.

Piscar de olhos

Por questões ambientais, designadamente a humidade, a temperatura ou a claridade e por factos inerentes a atividades diárias, tais como o cansaço ou a atividade física, é necessário que pisquemos os olhos, para cobrir a superfície do olho com um fluido protetor. Numa situação de relaxamento, piscamos com uma frequência de 12 a 19 vezes por minuto [Karson et al., 1981]; um piscar de olhos dura entre 100 ms e 400 ms [Schiffman, 2000].

3.1.1.2 *Baseline drift*

O sinal EOG está sujeito a várias fontes de ruído, tais como ruído proveniente da fonte de energia elétrica, do circuito de medida, dos elétrodos, de fontes fisiológicas (atividade muscular — EMG) e, também, através do contacto entre os elétrodos e a pele [Bulling et al., 2011].

O *baseline drift* consiste numa mudança lenta no potencial dos sinais EOG, que não é relacionada com o movimento dos olhos. Entre as possíveis causas do *baseline drift*, estão a interferência de sinais de fundo, a polarização dos elétrodos, a variação de pressão de contacto entre os elétrodos e a pele, entre outros fatores. Considerando a montagem típica de EOG, com quatro canais, o *baseline drift* é diferente entre os canais vertical e horizontal. Este impossibilita assim uma leitura direta dos sinais EOG, no entanto a amplitude das sacadas não é influenciada por este fator, visto que ocorrem em intervalos de tempo muito pequenos. O *baseline drift* não é considerado um tipo de ruído, pelos principais autores, mas sim um problema.

3.1.2 Eletromiografia (EMG)

Os sinais EMG provêm de potenciais elétricos, originados pelas células musculares, e são medidos no domínio temporal. Na maior parte das aplicações os sinais são lidos a partir de elétrodos colados na superfície da pele, no sítio do músculo, mas também podem ser obtidos diretamente do músculo, recorrendo a agulhas intra-musculares. Dependendo do músculo que se pretende analisar, são utilizados

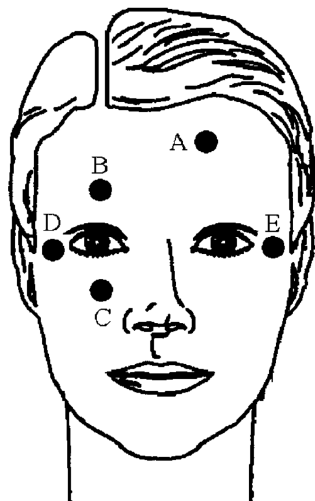


Figura 3.2: Montagem típica dos eléctrodos, na recolha de sinais EOG [Barea et al., 2002].

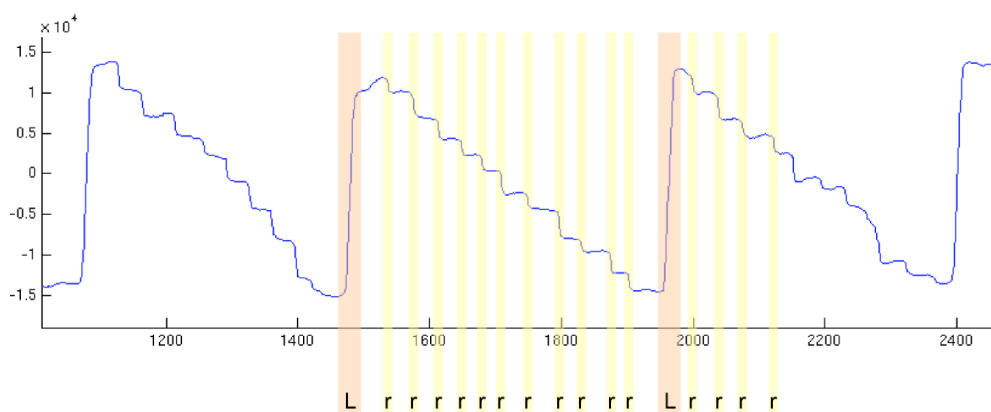


Figura 3.3: Sinal EOG típico, obtido durante a leitura de texto (sacada para a direita, marcada com “r”, e sacada para a esquerda, marcada com “L”). [Bulling, 2010]

dois elétrodos (um par), distanciados um do outro cerca de 1 cm ao longo da fibra muscular, para se poder medir a diferença de potencial. A força de contração de um músculo é proporcional ao número de células musculares que são contraídas simultaneamente, isto é, quanto mais células forem contraídas, maior é a força de contração de um músculo. O par de elétrodos mede a soma de potenciais das células musculares “ativadas”, portanto uma maior contração do músculo traduz-se em potenciais de maior amplitude medidos pelos elétrodos (figura 3.4) [Zweig, 2011]. O sinal EMG é caracterizado por uma gama de frequências que vai desde alguns Hz até mais de 2 kHz e amplitudes que variam entre frações de μV e várias centenas de μV na superfície da pele [Fridlund and Cacioppo, 1986].

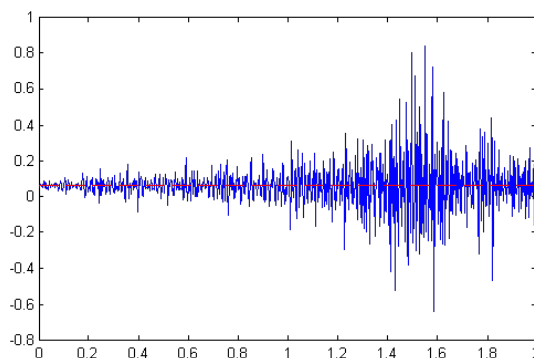


Figura 3.4: Sinal eletromiográfico do músculo tríceps direito, no levantamento de um supino de 2,3 Kg [Sarver, 2001].

Eletrromiografia facial (fEMG)

Tal como o nome indica, fEMG é o processo de leitura e de recolha de sinais eletromiográficos dos músculos da face. Os diferentes músculos faciais podem ser vistos na figura 3.5.

A leitura de sinais fEMG exige um posicionamento delicado e cuidado dos elétrodos para isolar os músculos, de forma eficiente, e evitar a leitura de sinais de músculos não-alvos. O artigo de Fridlund e Cacioppo ([Fridlund and Cacioppo, 1986]) pode servir como guia, no posicionamento dos elétrodos, para os diferentes músculos faciais. A figura 3.6 mostra o posicionamento dos elétrodos, sugerido nesse artigo, para os músculos faciais. Foi sugerido por Van Boxtel o uso de um filtro passa-alto de frequência 15 a 25 Hz, na análise de sinais fEMG, para eliminar artefactos de baixa frequência perdendo o mínimo de informação útil do sinal fEMG [Van Boxtel, 2001].

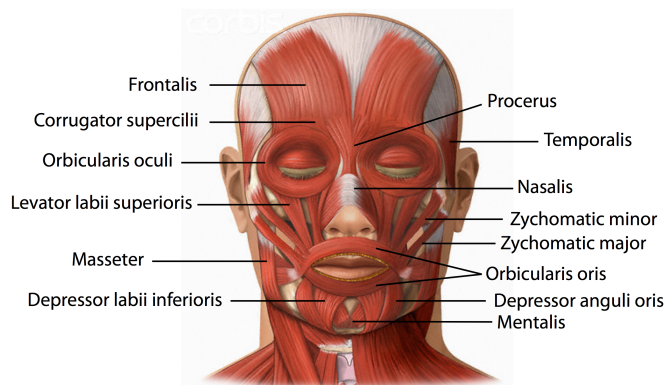


Figura 3.5: Ilustração dos músculos faciais [Boot, 2009]

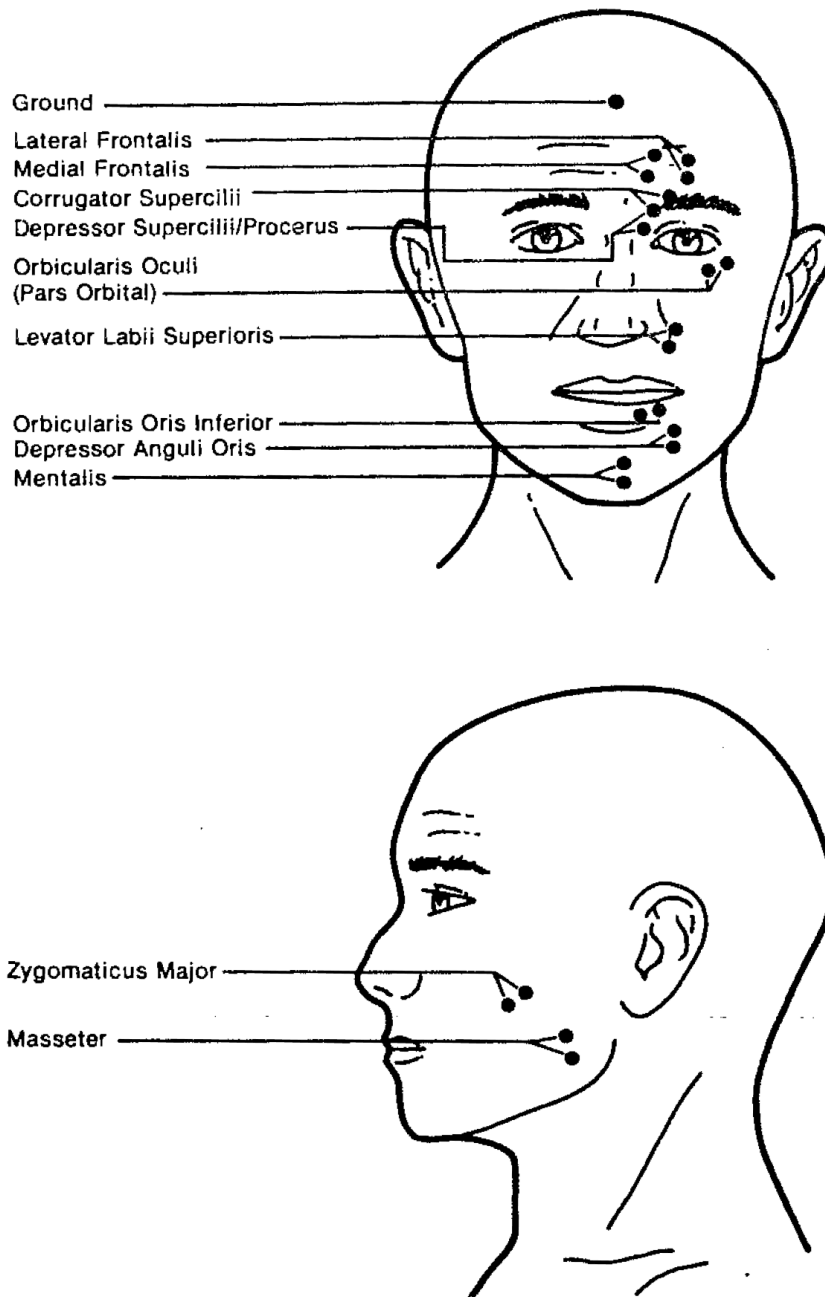


Figura 3.6: Posicionamento dos eletrodos, para leitura de sinais EMG dos músculos faciais [Fridlund and Cacioppo, 1986]

3.2 Classificador de Bayes

Um classificador discriminativo linear procura encontrar a melhor combinação linear, de um conjunto de características, que melhor discrimina as classes a partir de dados recolhidos nos treinos (*training sets*). A combinação linear é, posteriormente, usada para classificar as observações pertencentes a um teste (*test set*) [Boot, 2009].

O método de classificação, usado para a detecção dos sinais EMG e EOG, foi o classificador de Bayes a seguir descrito.

Notação:

- M — matriz de treino (N-por-p)
- N — n^o de amostras de matriz de treino
- p — n^o de caraterísticas
- x_n — amostra constituída por p caraterísticas (1-por- p) de índice n
- K — n^o de classes (ou grupos)

O classificador de Bayes pretende encontrar a solução para que o custo esperado de classificação seja mínimo. Matematicamente isto pode ser expresso como:

$$\hat{y} = \underset{y=1,\dots,K}{\operatorname{arg\,min}} \sum_{k=1}^K \hat{P}(k, x) C(y, k) \quad (3.1)$$

onde \hat{y} é a classe escolhida pelo classificador, $\hat{P}(k, x)$ é a probabilidade *a posteriori* da classe k para a amostra x , e $C(y, k)$ é o custo de classificar uma amostra como sendo da classe y quando a sua verdadeira classe é k . Seja π_k a probabilidade da classe k *a priori*. A probabilidade *a posteriori* de uma amostra x pertencer à classe k pode ser calculada da seguinte forma:

$$\hat{P}(k, x) = \frac{P(x, k)\pi_k}{\sum_{k=1}^K P(x, k)\pi_k} \quad (3.2)$$

A função de densidade da distribuição normal multivariada, $P(x, k)$, é dada pela expressão:

$$P(x, k) = \frac{1}{\sqrt{2\pi_k |\Sigma_k|}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)} \quad (3.3)$$

onde $|\Sigma_k|$ é o determinante de Σ_k e Σ_k^{-1} é a matriz inversa. A covariância é representada por Σ_k (equação 3.5) e a média representada por μ_k (equação 3.4). A estimação da média da classe k e a covariância é efetuada recorrendo às equações seguintes, em que $M_{nk} = 1$ se a amostra n pertencer à classe k e $M_{nk} = 0$ caso não pertença.

$$\hat{\mu}_k = \frac{\sum_{n=1}^N M_{nk} x_n}{\sum_{n=1}^N M_{nk}} \quad (3.4)$$

$$\hat{\Sigma}_k = \frac{\sum_{n=1}^N \sum_{k=1}^K M_{nk} (x_n - \hat{\mu}_k)(x_n - \hat{\mu}_k)^T}{N - K} \quad (3.5)$$

A probabilidade *a priori* da classe k , π_k , mencionada na expressão 3.2, é definida consoante a aplicação. No entanto, o mais comum é a probabilidade ser igual para todas as classes k . Portanto, pode ser obtida por $\pi_k = \frac{1}{K}$. O custo de classificação, $C(y, k)$, também pode ser escolhido consoante a aplicação. Geralmente, é definido da seguinte forma:

$$C(y, k) = \begin{cases} 0, & \text{se } y = k \\ 1, & \text{se } y \neq k \end{cases} \quad (3.6)$$

Ao classificador são, apenas, fornecidas a matriz de treino M e a amostra x , que queremos que seja classificada, para que o classificador nos devolva a classe \hat{y} , a partir do cálculo da expressão 3.1. Para uma visão geral do algoritmo, deve ser consultada a figura 3.7, bem como o pseudocódigo da função de classificação no algoritmo 3.1. O classificador de Bayes foi implementado com recurso à "Statistics Toolbox" do Matlab [MathWorks, 2014a]. Uma explicação mais exhaustiva do classificador de Bayes pode ser consultada em [Duda et al., 2001].

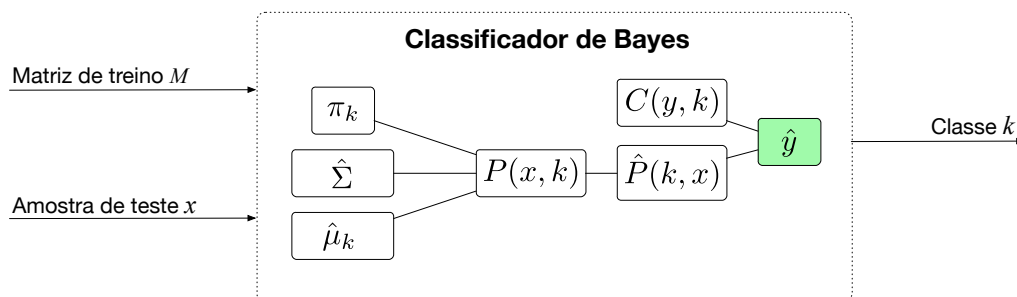


Figura 3.7: Estrutura do algoritmo do classificador de Bayes multi-classe

Algoritmo 3.1 Pseudocódigo da função de classificação

```

1: function CLASSIFICADOR( $M, x$ )
2:   for  $k = 1, \dots, K$  do
3:     Calcula  $\pi_k$ 
4:     Calcula  $\mu_k$  ▷ eq. 3.4
5:     Calcula  $\Sigma_k$  ▷ eq. 3.5
6:     Calcula  $P(x, k)$  ▷ eq. 3.3
7:     Calcula  $\hat{P}(k, x)$  ▷ eq. 3.2
8:   end for
9:   Calcula  $\hat{y}$  ▷ eq. 3.1
10:  return  $\hat{y}$ 
11: end function
    
```

3.3 Extração de características de biosinais EOG e EMG

Esta secção descreve os métodos de extração de características implementados para caracterizar os sinais EOG e EMG. Podem ser encontradas mais técnicas em [Aungsakul et al., 2012] e [Phinyomark et al., 2012]. Os artigos, também, fazem uma análise da relevância individual das características no desempenho de classificação.

3.3.1 As características do sinal fEMG

De acordo com A. Phinyomark, a extração de características de sinais eletromiográficos pode ser dividida em três tipos: domínio do tempo, domínio da frequência e domínio do tempo-frequência (denominado, também, por representação tempo-escala) [Phinyomark et al., 2012]. As técnicas implementadas, neste trabalho, são aplicadas no domínio do tempo. As características do domínio da frequência não foram utilizadas, pois destinam-se ao estudo da fadiga dos músculos.

As características do sinal EMG são extraídas de uma janela de amostragem com comprimento de 256 amostras (1 segundo). A notação utilizada é a seguinte: L é o número de valores da amostra analisada e x_i é o i -ésimo valor de amplitude da amostra.

Média dos valores absolutos (MVA)

O sinal EMG tem uma componente muito baixa, ou mesmo inexistente, de frequências de baixa gama. Praticamente, toda a potência do sinal EMG está situada na banda de frequências de 30 Hz a 200 Hz [Fridlund and Cacioppo, 1986], sendo por este motivo um sinal centrado em zero com bastantes oscilações. A média dos sinais EMG é quase nula, portanto foi descartada das características a extrair. Por outro lado uma característica que fornece uma boa identificação e distinção dos sinais é a média dos valores absolutos da amostra analisada, que pode ser definida como:

$$MVA = \frac{1}{L} \sum_{i=1}^L |x_i| \quad (3.7)$$

A Média dos valores absolutos é uma característica que pertence ao domínio do tempo.

Difference absolute standard deviation value (DASDV)

O DASDV é, também, uma característica do domínio temporal e foca-se no desvio padrão absoluto da diferença entre valores adjacentes do sinal EMG.

$$DASDV = \sqrt{\frac{1}{L-1} \sum_{i=1}^{L-1} (x_{i+1} - x_i)^2} \quad (3.8)$$

Porcentagem de pulsos superiores ao *threshold* (PPST)

A PPST é uma característica que permite avaliar a porcentagem do sinal absoluto que ultrapassa um determinado valor de *threshold* pré-definido. O valor para o *threshold* igual a 50, costuma apresentar bons resultados. Matematicamente, esta característica, é calculada da seguinte forma:

$$PPST = \frac{1}{L} \sum_{i=1}^L f(x_i) \quad (3.9)$$

$$f(x) = \begin{cases} 1, & \text{se } x \geq \textit{threshold} \\ 0, & \text{se } x < \textit{threshold} \end{cases}$$

3.3.2 As características do sinal EOG

Nesta secção, são descritas as características extraídas do sinal EOG, nomeadamente, o máximo, o mínimo, a média e a potência espectral relativa. As características são extraídas de uma janela de amostragem, $x[n]$, de 512 amostras (2 segundos) de comprimento.

Máximo, mínimo e média

Para cada janela de amostragem são extraídos os valores máximo, mínimo e médio. Estas três características foram escolhidas por serem extremamente eficientes no processo de classificação, pois facilmente distinguem os movimentos oculares cima, baixo, direita, esquerda e piscar de olhos.

Potência de banda relativa

A potência de banda relativa permite analisar o sinal EOG, ao nível da frequência, e pretende-se, nomeadamente, identificar uma maior potência relativa, nas frequências mais elevadas do sinal EOG medido. Por cada amostra de sinal EOG, é calculada a potência relativa para as bandas de frequência dos 10 Hz a 15 Hz, 15 Hz a 20 Hz, 20 Hz a 25 Hz e 25 Hz a 30 Hz. Considere P_{B_j} a potência de uma banda de frequências de índice j , e P_{Total} a potência total da amostra analisada. A potência relativa da banda de frequências j é dada por:

$$P_{rel_j} = \frac{P_{B_j}}{P_{Total}} \quad (3.10)$$

O cálculo da potência total e da potência de banda é explicado de seguida. Primeiramente, é calculada a potência associada às frequências $f = [0 : N - 1] \times (f_s/N)$ Hz, também chamada de potência espectral. Seja $f_s=256$ Hz a frequência de amostragem do sinal, $X(f)$ a DFT da amostra de sinal $x[n]$, e $\overline{X(f)}$ o conjugado de $X(f)$. A equação que segue define, matematicamente, a potência espectral:

$$P(f) = \frac{1}{f_s} X(f) \overline{X(f)} \quad (3.11)$$

A potência total é a soma da potência espectral para as frequências f , definida pela expressão matemática seguinte:

$$P_{Total} = \sum_{i=0}^{\frac{f_s}{2}-1} P(f_i) \quad (3.12)$$

onde f_i é a frequência de índice i das frequências f .

Seja P_{B_1} a potência da banda de frequência de a Hz a b Hz (sendo $a > b$):

$$P_{B_1} = \sum_{i=0}^{\frac{f_s}{2}-1} P(f_i)G(f_i) \quad (3.13)$$

$$G(f) = \begin{cases} 1, & \text{se } a \leq f \leq b \\ 0, & \text{outros casos} \end{cases}$$

Capítulo 4

Arquitetura e *setup* experimental

Neste trabalho, foram desenvolvidos três sistemas que funcionam de forma independente: um sistema de classificação de movimentos oculares, outro de classificação de expressões faciais e, por último, um sistema de seguimento ocular horizontal. No processo, foram utilizados vários equipamentos de aquisição de dados, foram desenvolvidos algoritmos de processamento de sinais que comunicam com o sistema do avatar desenvolvido através de comunicação UDP/IP. Nas secções seguintes são descritos cada um destes módulos.

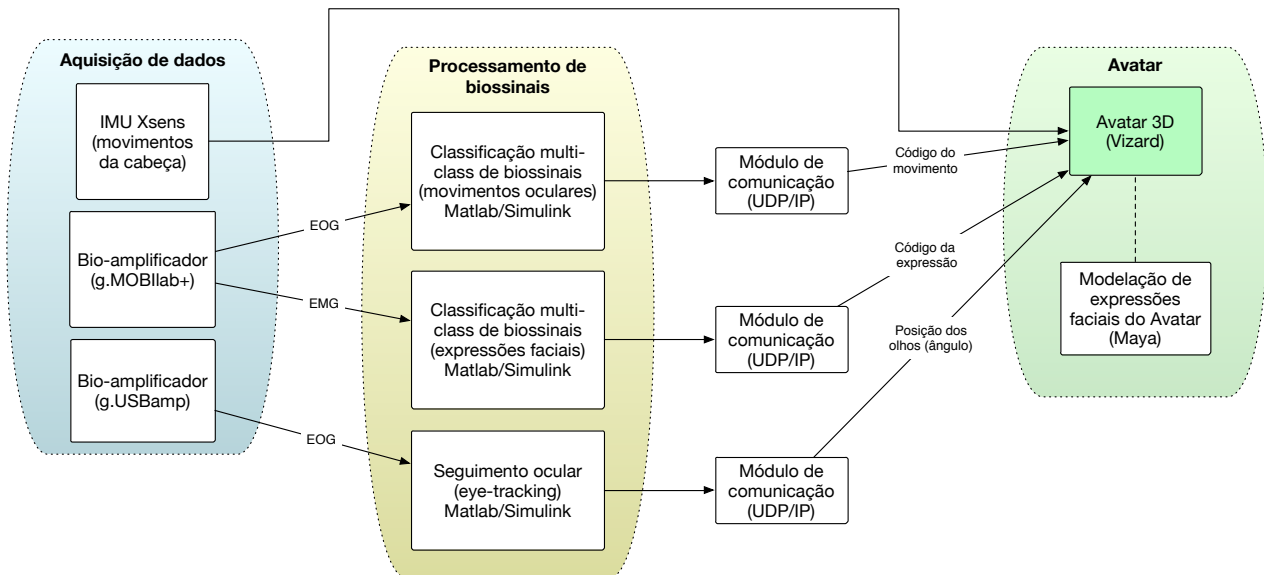


Figura 4.1: Diagrama dos módulos desenvolvidos

4.1 Sistemas de aquisição de biossinais

Foram usados dois sistemas de aquisição de biossinais: o g.MOBilab+[®] (figura 4.2, marcado com número 4) e o g.USBamp[®] (figura 4.3, marcado com número 5), que são dois amplificadores de biossinais fabricados pela g.tec. O g.MOBilab+ é um bioamplificador vocacionado ao estudo da atividade cerebral, muscular, do coração, dos movimentos oculares, da respiração, entre outros, adotando a mobilidade como principal característica. O equipamento é ligado a um computador através de ligação sem fios *Bluetooth* 2.0, tem um conversor A/D de 16 bits e uma taxa de amostragem máxima de 256 Hz. Foi utilizado um conector para ligação de elétrodos, que tem um filtro passa-banda, não-removível, de 0,5 a 100 Hz, sendo que os canais ligados têm sensibilidade de 5 mV. O g.USBamp é o sistema com melhor desempenho e mais precisão produzido pela g.tec, sendo o equipamento *standard* para investigação nas áreas que envolvem processamento de biossinais. O sistema possui 16 canais, onde se podem ligar



Figura 4.2: Fotografia do *setup* experimental do sistema EOG. (1) Eléctrodos; (2) IMU; (3) Avatar; (4) g.MOBIlab+.

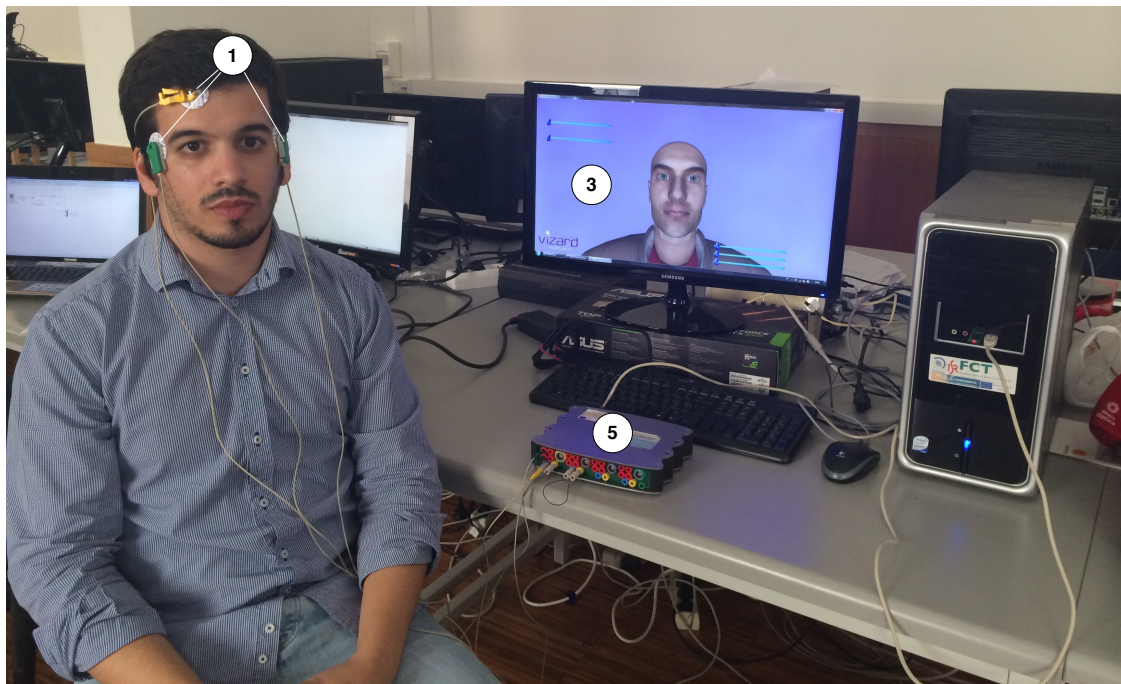


Figura 4.3: Fotografia do *setup* experimental do sistema de seguimento ocular. (1) Eléctrodos; (3) Avatar; (5) g.USBamp.

vários elétrodos, com 4 *grounds* independentes. O conversor A/D é de 24 bits de resolução, possuindo a taxa de amostragem máxima de 38,4 kHz. A ligação ao computador é feita através de uma porta USB 2.0. Ambos os amplificadores são acompanhados por filtros digitais, desenvolvidos pela g.tec, uma configuração fácil para as plataformas Simulink e LabVIEW, e por API's de desenvolvimento para ambientes Matlab e C [g.tec medical engineering GmbH, 2013].

O g.MOBilab+ foi utilizado na aquisição de sinais EOG e EMG, que são posteriormente descritos nas secções 5.1 e 5.2. O EOG permite medir a direção, vertical e horizontal, colocando os elétrodos em posições específicas (figura 4.4). Quatro elétrodos foram montados numa configuração bipolar: os elétrodos da direita e da esquerda, na zona exterior do canto dos olhos, para detetar os movimentos horizontais, e os elétrodos abaixo e acima do olho para medir movimentos verticais. No processo de aquisição dos sinais EMG foram utilizados quatro elétrodos, também em configuração bipolar, colocados na superfície da pele sobre os músculos corrugador do supercílio e zigomático maior. A colocação dos elétrodos foi efetuada, de acordo com o artigo [Fridlund and Cacioppo, 1986] que pode ser vista em detalhe na figura 4.5. O distanciamento entre dois elétrodos de cada par é de aproximadamente 1 cm.

O bioamplificador g.USBamp foi utilizado na aquisição de sinais EOG para o sistema descrito na secção 5.3. A montagem dos elétrodos segue a montagem típica dos sistemas EOG (utilizada com o g.MOBilab+), mas apenas com o par de elétrodos referente ao sinal EOG horizontal (figura 4.6).

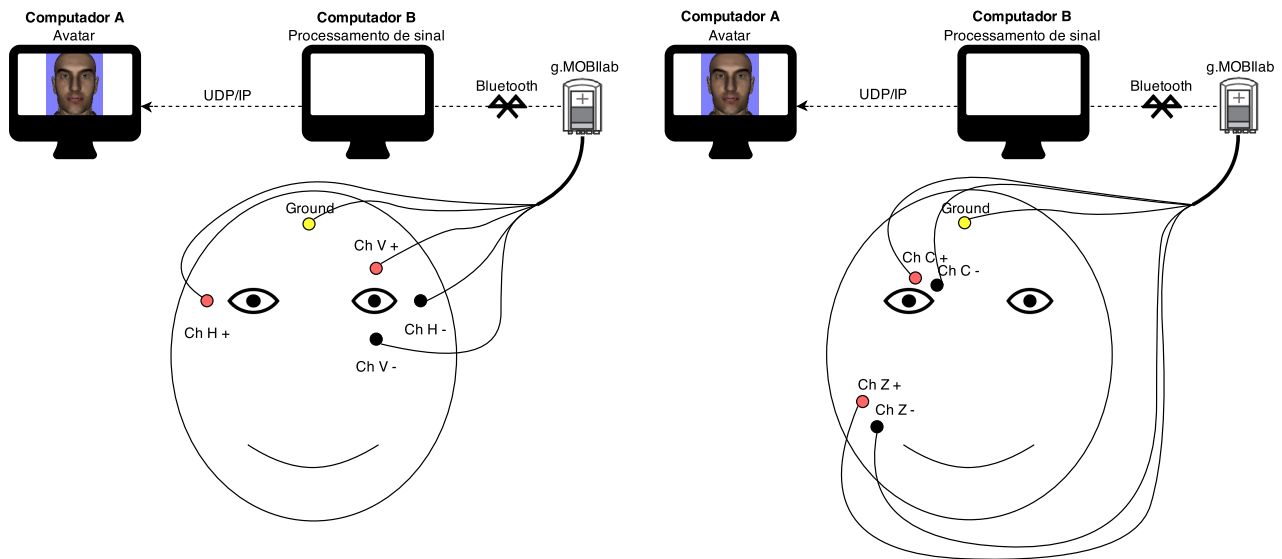


Figura 4.4: Disposição do sistema de aquisição de dados, utilizando o bioamplificador g.MOBilab+ com os canais vertical e horizontal.

Figura 4.5: Disposição do sistema de aquisição de dados com os canais do corrugador do supercílio e do zigomático maior.

4.2 *Inertial Motion Unit*

Par este trabalho, foi usado o sistema Xsens[©] MTi-100 (3^a geração) , um sensor IMU, fabricado pela Xsens, e utilizado em diversas áreas, como a robótica móvel, as animações de jogos e de filmes para captura de movimentos de diversos membros do corpo. Este IMU tem incorporado giroscópios, acelerómetros e magnetómetros 3D.

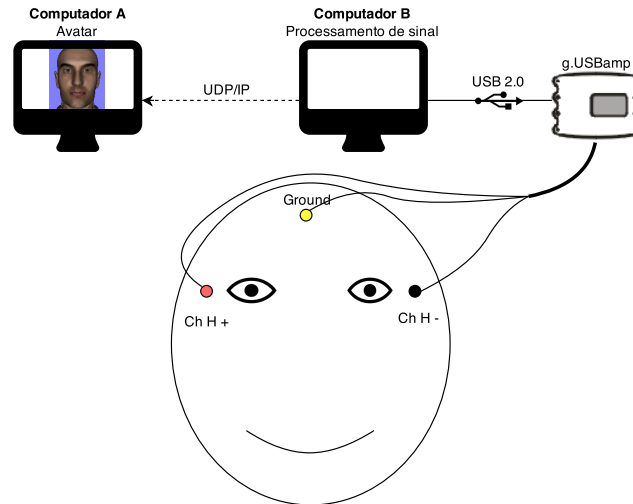


Figura 4.6: Disposição do sistema de aquisição de dados, utilizando o bioamplificador g.USBamp com o canal horizontal.

O sensor Xsens foi utilizado, neste trabalho, para medir a pose da cabeça, que se pode revelar importante na inferência de estados emocionais. Este sensor mede os ângulos *roll*, *pitch* e *yaw* efetuados após inicialização da medição. Por este motivo, o utilizador deve permanecer com uma posição da cabeça frontal e equilibrada, aquando da inicialização do sistema. Os eixos x e y do sensor também precisam de obedecer a um posicionamento específico, portanto foi embutido na parte de trás de um boné (figura 4.2, marcado com número 2), permitindo a colocação correta de forma bastante rápida.

A informação recolhida por este sensor é enviada diretamente ao programa que controla o avatar, o Vizard, através de ligação USB 2.0.

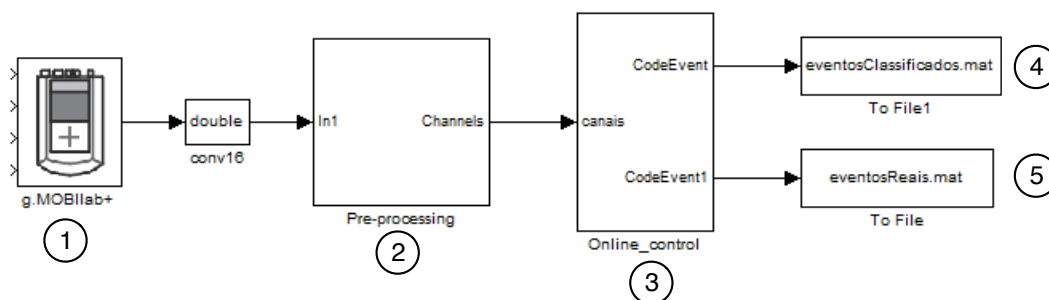
4.3 Framework de processamento e classificação de bioassinais

Todo o processamento e classificação de bioassinais, deste trabalho, foi feito no Simulink[®] e no Matlab. Na figura 4.7, pode ser visualizado o esquema do Simulink do sistema de EOG (o esquema do sistema EMG é similar), seguido da tabela 4.1 que explica o significado dos blocos do Simulink usados. O bloco 3, onde é feito o processamento de bioassinais em tempo-real, tem embutido um pequeno módulo de comunicação UDP/IP, que envia as informações para o Vizard. O diagrama do Simulink, da figura 4.8, foi o utilizado no sistema de seguimento ocular.

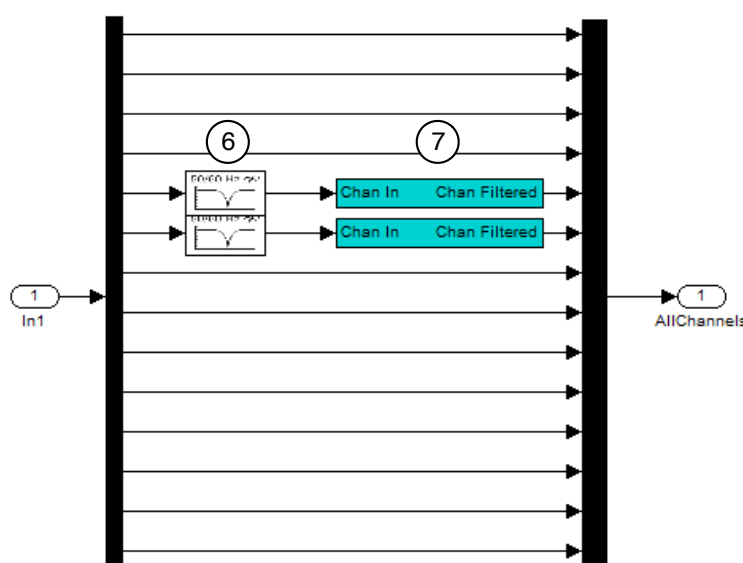
4.4 Vizard

O Vizard[™] é um software desenvolvido, pela empresa WorldViz, que proporciona aos utilizadores a construção de conteúdos e ambientes tridimensionais interativos de forma rápida. As principais características deste software são: a rapidez de execução de grandes projetos, a possibilidade de importação de objetos 3D, concebidos noutros programas, o seu vasto suporte de hardware, bem como de sensores e o seu excelente desempenho em aplicações de tempo-real. A linguagem utilizada, na programação do Vizard, é Python 2.7.

Na implementação prática, foi utilizada a versão Vizard 4.09 Lite, onde foi desenvolvido um pro-



(a)



(b)

Figura 4.7: Diagrama de blocos Simulink do sistema EOG: (a) Diagrama completo; (b) Bloco de pré-processamento.

Tabela 4.1: Identificação dos blocos do Simulink do sistema EOG

Bloco	Funcionalidade
1	Desenvolvido pela g.tec, este bloco devolve os sinais recolhidos pelo bioamplificador g.MOBllab+
2	Pré-processamento dos canais (filtragem)
3	Bloco onde os biosinais são processados, e onde os algoritmos, explicados no capítulo 5, são aplicados. Este bloco é uma S-Function de nível 2 que tem vários parâmetros de entrada, sendo o tamanho da janela deslizante (“win”) e o tempo de amostragem (“Ts”) os parâmetros comuns aos dois sistemas, EOG e EMG. O <i>buffering</i> do sinal é efetuado neste bloco, guardando as últimas L amostras temporais, onde L é o tamanho da janela deslizante.
4	Bloco que grava num ficheiro os resultados do funcionamento <i>online</i> para efeitos de testes, guardando um código que identifica um movimento/expressão facial e o instante de tempo em que foi classificado
5	Bloco que grava num ficheiro os movimentos que se pretendem fazer num teste para que possam ser comparados com os dados do ficheiro anterior
6	Filtros <i>notch</i>
7	Filtros <i>Butterworth</i> desenvolvidos pela g.tec

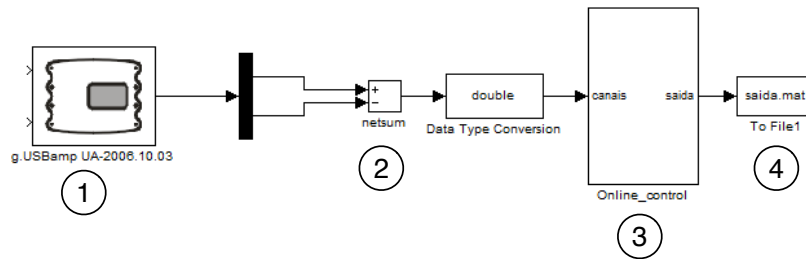


Figura 4.8: Diagrama de blocos Simulink do sistema de seguimento ocular. (1) Bloco que devolve sinais adquiridos pelo g.USBamp; (2) Diferença entre os sinais dos dois elétrodos; (3) Bloco onde os biossinais são processados (algoritmo da secção 5.3); (4) Sinal da posição dos olhos.

grama (pseudocódigo apresentado no algoritmo 4.1) que recebe os dados dos subsistemas EOG e EMG em forma de eventos codificados por um número inteiro e executa esses eventos no avatar apresentado na figura 4.9. O avatar é constituído por um corpo (que vem com o programa Vizard) e uma cabeça, que foi concebida à parte (consulte a figura 4.10). A cabeça do avatar é, por sua vez, constituída por duas partes — a cabeça e os olhos. Este tipo de independência das diferentes partes é imperial para o funcionamento do sistema do avatar, pois só assim é permitido controlar cada uma das partes individualmente.

Para construir a cabeça, e prepará-la para o funcionamento dentro do Vizard, recorreu-se ao programa PeopleMaker, que acompanha o Vizard. Neste programa, procedeu-se à configuração das *morphs* do avatar, ao escalonamento da cabeça e à seleção dos pontos do pescoço, para que pudesse ser usada no corpo fornecido pelo Vizard. As *morphs* são o nome dado às transformações da cabeça/face, por forma a criar expressões faciais diferentes, bem como movimentos — “pisar de olhos” e “franzir”. Neste caso foram consideradas seis *morphs*.

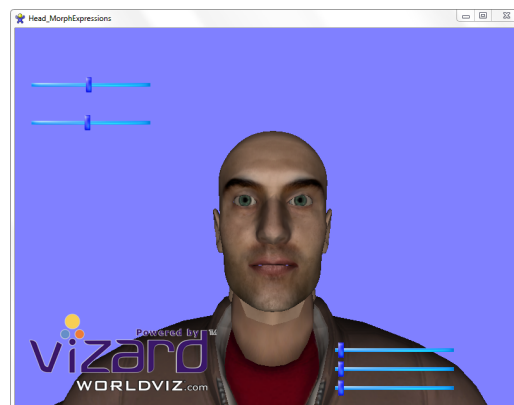


Figura 4.9: Script do Vizard em execução

4.5 Autodesk Maya

O Autodesk Maya[®] 2014 é um software muito completo de criação e de edição de objetos 3D, desenvolvido pela Autodesk. Foi o software utilizado para a criação das *morphs*, mencionadas na secção anterior.

Algoritmo 4.1 Pseudocódigo executado no Vizard

```

1: loop
2:   data ← recebe pacote UDP
3:   if data é uma expressão/movimento then
4:     if data == expressão neutra then
5:       desativar todas as morphs
6:       aplicar morph neutra
7:     else if data == expressão felicidade then
8:       desativar todas as morphs
9:       aplicar morph felicidade
10:    else if data == expressão tristeza then
11:      desativar todas as morphs
12:      aplicar morph tristeza
13:    else if data == expressão raiva then
14:      desativar todas as morphs
15:      aplicar morph raiva
16:    else if data == expressão felicidade then
17:      desativar todas as morphs
18:      aplicar morph felicidade
19:    else if data == movimento ocular cima then
20:      posição vertical dos olhos = 30 graus
21:      posição horizontal dos olhos = 0 graus
22:    else if data == movimento ocular baixo then
23:      posição vertical dos olhos = -30 graus
24:      posição horizontal dos olhos = 0 graus
25:    else if data == movimento ocular direita then
26:      posição vertical dos olhos = 0 graus
27:      posição horizontal dos olhos = 50 graus
28:    else if data == movimento ocular esquerda then
29:      posição vertical dos olhos = 0 graus
30:      posição horizontal dos olhos = -50 graus
31:    else if data == piscar de olhos then
32:      PISCAROLHOS
33:      posição vertical dos olhos = 0 graus
34:      posição horizontal dos olhos = 0 graus
35:    else if data == franzir then
36:      aplicar morph franzir
37:      posição vertical dos olhos = 0 graus
38:      posição horizontal dos olhos = 0 graus
39:    end if
40:
41:    else if data é um ângulo ocular then
42:      posição vertical dos olhos = 0 graus
43:      posição horizontal dos olhos = data graus
44:    end if
45:
46:    function PISCAROLHOS
47:      aplicar morph olhos fechados
48:      aplicar morph olhos abertos
49:    end function
50:
51:    poseRollPitchYaw ← leitura do IMU
52:    Pose da cabeça = poseRollPitchYaw
53:  end loop

```

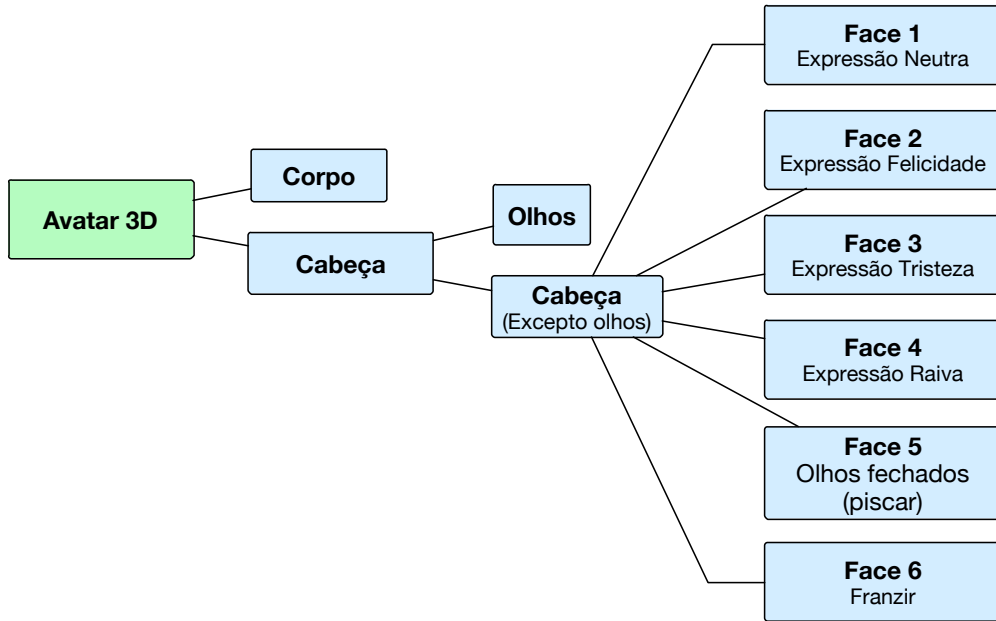


Figura 4.10: Constituição do avatar

Este processo iniciou-se com a escolha de um modelo 3D de uma cabeça humana, já concebido. Uma tarefa bastante morosa, tendo em conta que o modelo deveria ser, preferencialmente, de licença livre e que tivesse em conta o fenómeno do *uncanny valley* (este fenómeno é um hipótese no campo da robótica e das animações 3D que afirma que existe um comportamento de repulsa em nós, humanos, quando vemos réplicas humanas muito parecidas com humanos reais, mas não iguais [Masahiro et al., 2012]). O modelo da figura 4.9 foi seleccionado do site TurboSquid, cuja licença permite o uso livre para diversos trabalhos, incluindo este.

Usando o Maya, a primeira tarefa consistiu em reduzir consideravelmente o número de vértices (pontos) do modelo 3D, de forma a facilitar a edição e o processamento gráfico. De seguida, foram recriadas as várias expressões faciais, e também uma *morph* para o movimento “piscar de olhos” e “franzir”, a partir do modelo de base neutro, através da translação manual dos vértices na zona dos olhos/testa e boca. A figura 4.11 apresenta os produtos finais da edição.

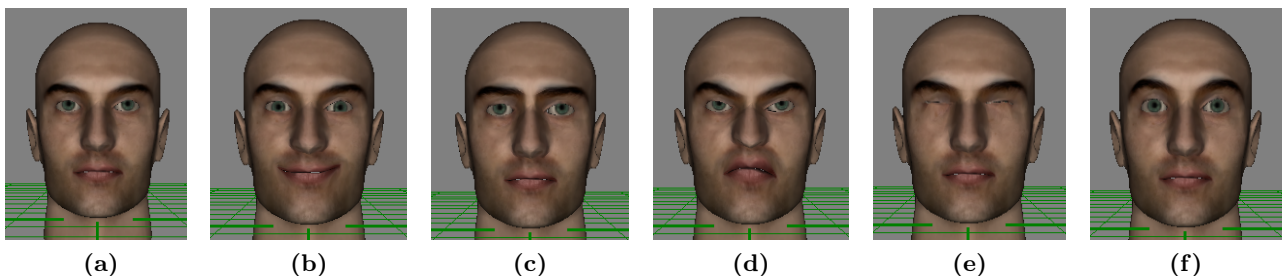


Figura 4.11: *Morphs* das expressões faciais: (a) “Neutro”; (b) “Felicidade”; (c) “Tristeza”; (d) “Raiva”; (e) “Piscar de olhos”; (f) “Franzir”.

Capítulo 5

Reconhecimento de expressões faciais

Este capítulo explica com detalhe o funcionamento dos algoritmos de reconhecimento automático implementados.

5.1 Classificador EOG

Foi desenvolvido um sistema de reconhecimento automático de expressões faciais na zona dos olhos e da testa baseado em sinais eletro-oculográficos (EOG), de acordo com a montagem da figura 4.4. São detetados seis movimentos dos olhos, nomeadamente, “cima”, “baixo”, “direita”, “esquerda”, “pisca” e “franzir”.

A figura 5.1 tem esquematizado o funcionamento global do sistema de reconhecimento online (setas contínuas) e treino (setas a tracejado). O sistema começa por receber dois sinais EOG dos canais vertical e horizontal, que são pré-processados (filtragem na banda de interesse). No funcionamento online, um algoritmo deteta assincronamente a ocorrência de movimentos (método baseado numa janela deslizante). Sempre que um movimento é detetado, são extraídas características dos dois sinais que servem para classificar o movimento num dos seis mencionados anteriormente. Para o classificador funcionar, é necessário obter previamente modelos de classificação, ajustados ao utilizador, através do processo de treino/calibração, como descrito na secção seguinte.

5.1.1 Treino do classificador

Antes da utilização do sistema de classificação online, é necessária uma sessão de treino, que consiste na repetição dos seis movimentos referidos: cima, baixo, direita, esquerda, piscar e franzir. Os participantes sentam-se em frente a um computador, numa divisão com ambiente calmo e luminosidade moderada, e seguem uma bola em movimento que se posiciona de forma a simular quatro direções: direita, esquerda, cima e baixo. Os restantes movimentos são indicados através de mensagens exibidas no centro do ecrã, nomeadamente, “franzir” e “pisca”. Cada movimento é realizado durante 2 segundos com intervalos de repouso de 2 segundos, tendo uma sequência de movimentos a duração de 20 segundos (figuras 5.2 e 5.4). A sequência de movimentos pode ser consultada na figura 5.2. Um *dataset* de uma sessão de treino com 20 repetições do mesmo movimento (20×6 épocas) (considera-se uma época o intervalo de tempo referente a um movimento ocular) é usado para treinar o classificador para o funcionamento online. Para cada movimento são extraídas as características (máximo, mínimo, média e potência relativa de quatro bandas de frequência distintas) dos sinais EOG, vertical e horizontal, permitindo construir uma matriz de treino M .

A matriz M (equivalente à matriz M mencionada na descrição do classificador de Bayes, secção 3.2) é constituída por 120 linhas (N amostras) e 14 colunas (p características). Cada linha representa as características de cada movimento repetido, sendo constituído pelas sete características dos sinais EOG

vertical e horizontal (vetor de características [max_v min_v ...] apresentado na figura 5.1), nesta ordem.

5.1.2 Pré-processamento

Os sinais EOG, são frequentemente afetados por ruído, proveniente do contacto entre o elétrodo e a pele, por artefactos musculares e pela própria fonte de energia elétrica. Para eliminar essas interferências, os sinais de EOG são filtrados, utilizando um filtro *notch* de 50 Hz e um filtro *Butterworth* (passa-banda) de quarta ordem com frequência de corte mínima de 1 Hz e frequência de corte máxima de 30 Hz. Note-se que a frequência de corte mínima elimina o efeito do *base-line drift*, mencionado na secção 3.1.1.2. A frequência de corte máxima de 30 Hz pretende eliminar principalmente artefactos musculares, aproveitando o máximo de sinal EOG útil, como é mencionado no capítulo do background, secção 3.1.1.

5.1.3 Detecção assíncrona de movimento: *sliding window*

Nas sessões de treino, indica-se ao utilizador para que realize um movimento específico num determinado intervalo de tempo (ou seja, a ordem para realizar o movimento é despoletada pelo sistema). No entanto, durante o funcionamento on-line, o início do movimento deve ser detetado automaticamente e de forma assíncrona, para posteriormente serem extraídas as características. Esta deteção é realizada através de uma janela deslizante (*sliding window*) que armazena as últimas 256 amostras (1 segundo) do sinal EOG. É impreterível que a janela seja bem posicionada para captar todo o movimento. Para isso, a janela deslizante é dividida em 8 sub-intervalos iguais (ver figura 5.3), e os valores máximos absolutos são calculados para cada sub-intervalo. De seguida, o pico máximo da amostra deve ser posicionado no 4^o sub-intervalo do canal vertical ou do canal horizontal, e deve ser maior do que um certo *threshold* (calculado após a sessão de treino de cada indivíduo). Caso contrário, nenhuma classificação é efetuada e a janela deslizante é deslocada 8 amostras para a frente no tempo.

O *threshold* vertical (equação 5.1) é 70 por cento da média do valor máximo e mínimo absoluto do movimento ocular “cima” e “baixo” do canal vertical. O *threshold* horizontal (equação 5.2) é 80 por cento da média do valor máximo e mínimo absoluto do movimento ocular “direita” e “esquerda” do canal horizontal.

$$T_v = 0.7 \times \left[\frac{1}{N_{vm}} \sum_{i=1}^{N_{vm}} \max(|Xv_i|) \right] \quad (5.1)$$

$$T_h = 0.8 \times \left[\frac{1}{N_{hm}} \sum_{i=1}^{N_{hm}} \max(|Xh_i|) \right] \quad (5.2)$$

Onde T_v é o *threshold* vertical, T_h é o *threshold* horizontal, N_{vm} é o número de movimentos “cima” e “baixo”, N_{hm} o número de movimentos “direita” e “esquerda”, realizados durante a sessão de treino, e Xv_i e Xh_i são janelas de amostragem do sinal EOG vertical e horizontal, respetivamente, correspondentes aos movimentos mencionados anteriormente.

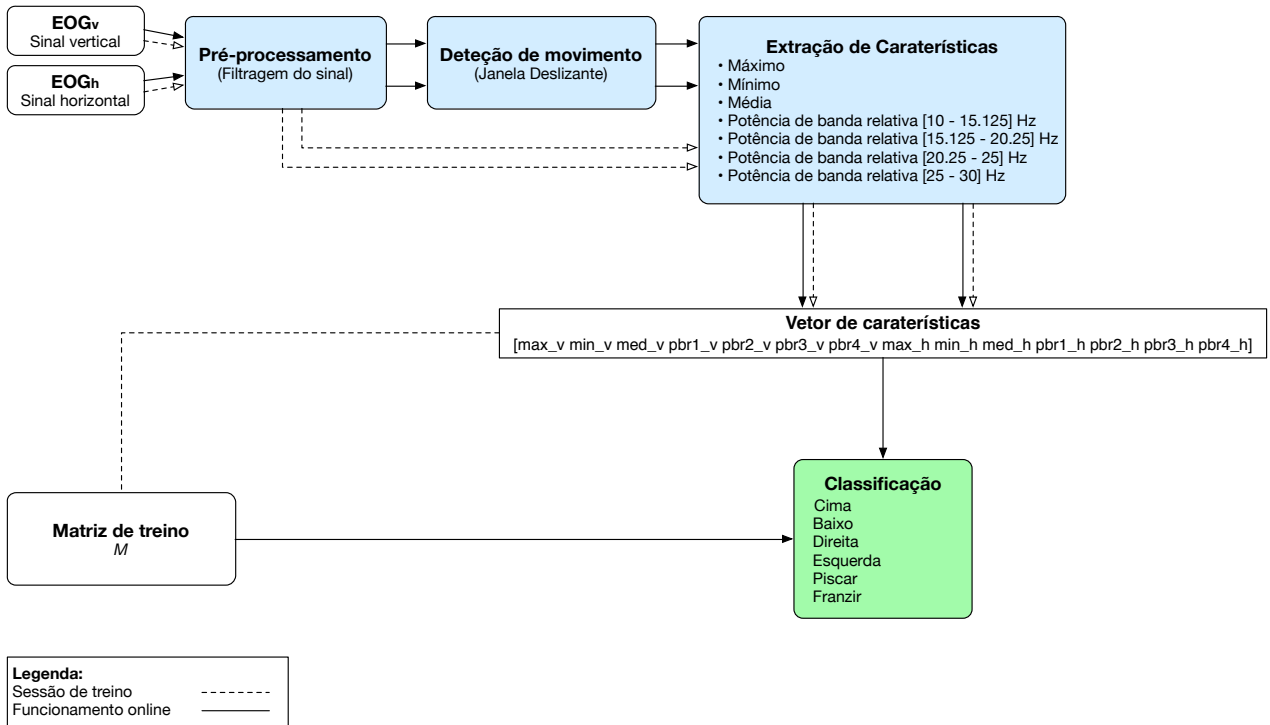


Figura 5.1: Estrutura do sistema de reconhecimento EOG

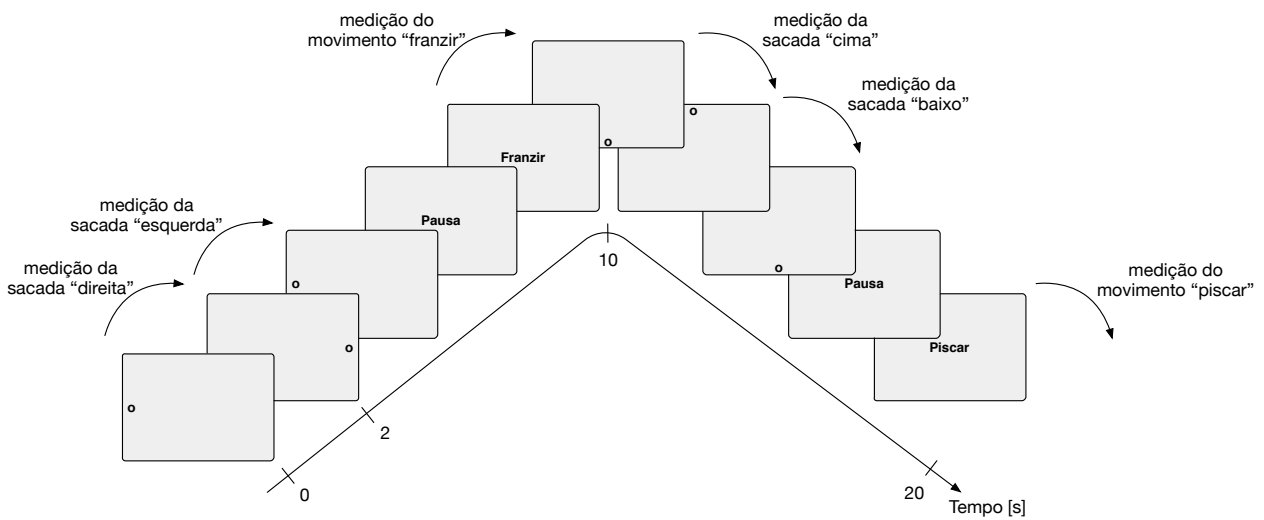


Figura 5.2: Sequência de movimentos de uma sessão de treino do classificador EOG

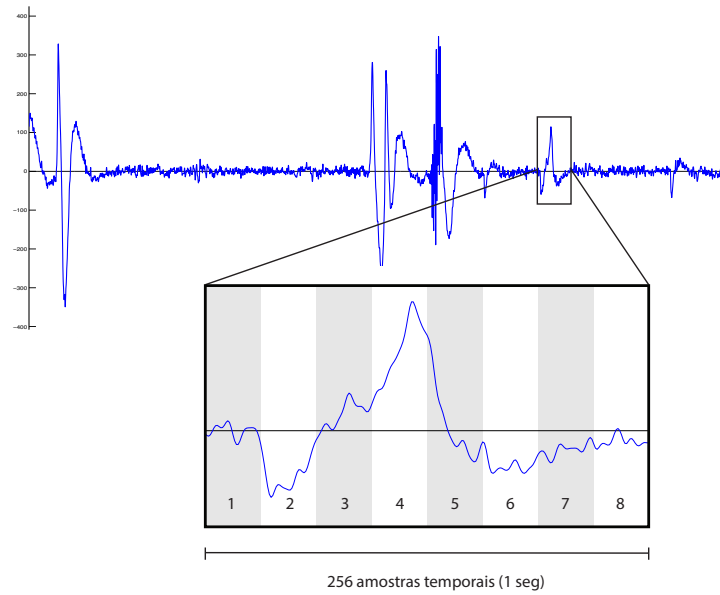


Figura 5.3: Sinal EOG vertical. Janela deslizante a detetar o movimento ocular para cima, com o máximo do sinal posicionado no 4º sub-intervalo.

5.1.4 Extração de caraterísticas

As caraterísticas são extraídas no domínio temporal, através do cálculo do máximo, mínimo e média e no domínio da frequência através do cálculo da potência de banda relativa nas frequências de banda 10 Hz a 15,125 Hz, 15,125 Hz a 20,25 Hz, 20,25 Hz a 25 Hz e 25 Hz a 30 Hz.

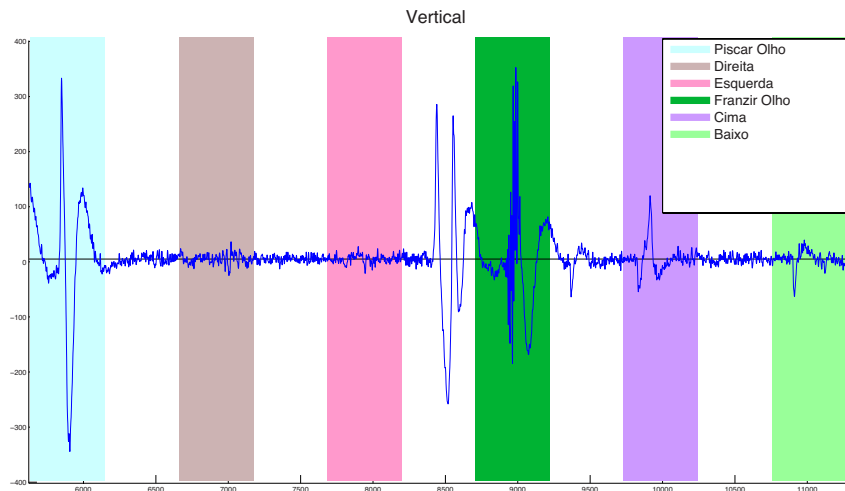
Experimentalmente, as caraterísticas máximo, mínimo e média, usadas numa primeira fase, forneceram um desempenho de classificação aquém do desejado. O movimento “franzir”, por vezes, era confundido com os movimentos oculares “cima” e “pisar”. Não é habitual o movimento “franzir” ser identificado a partir de sinais eletro-oculográficos, pois deriva da ação de músculos faciais (*frontalis*). Por este motivo, foi adicionada a caraterística potência de banda relativa.

A potência de banda relativa de cada um dos intervalos mencionados é adicionada ao vetor de caraterísticas que, juntamente com o máximo, mínimo e média do sinal, formam um vetor com sete caraterísticas para cada sinal EOG vertical e horizontal. Na banda de frequências dos 25 Hz a 30 Hz verifica-se um valor de potência relativa maior, aquando da análise de um movimento “franzir”, o que se traduz numa classificação com desempenho superior.

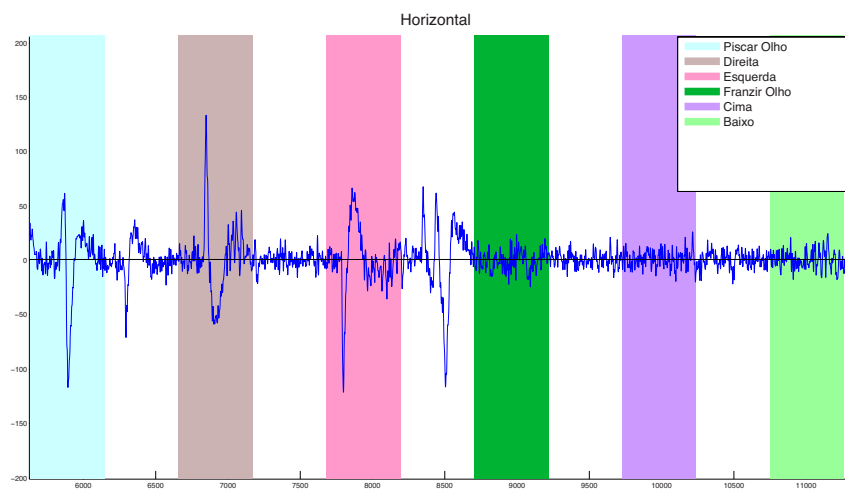
Após a extração de caraterísticas dos sinais EOG vertical e horizontal, as mesmas são agregadas num único vetor com 14 valores, referido na figura 5.1 como “Vetor de caraterísticas”.

5.1.5 Classificação do movimento

Os padrões EOG, representando cada uma das 6 classes (movimentos), são classificados com o classificador de Bayes multi-classe, descrito na secção 3.2. O classificador, com a informação de um treino prévio (Matriz de treino), recebe os dados de um novo movimento (Vetor de teste) e classifica-o num destes movimentos: “cima”, “baixo”, “direita”, “esquerda”, “pisar” ou “franzir”.



(a)



(b)

Figura 5.4: Sinais EOG (a) vertical e (b) horizontal para uma sequência de seis movimentos dos olhos, gravados durante uma sessão de treino: “piscar”, “direita”, “esquerda”, “franzir”, “cima” e “baixo”. Os espaços preenchidos a branco representam os períodos de repouso (não têm qualquer influência no treino do classificador)

5.2 Classificador EMG

De igual modo, foi desenvolvido um sistema de reconhecimento de expressões faciais baseado nos sinais eletromiográficos (EMG). São detetadas 4 expressões faciais básicas a partir dos músculos corrugador do supercílio (*corrugator supercilii*) e zigomático maior (*zygomaticus major*): “neutro”, “felicidade”, “tristeza” e “raiva”. Neste caso, deve-se realçar que se considera “neutro” como uma expressão facial.

Considere a figura 5.5, onde é apresentado o diagrama do funcionamento do sistema de reconhecimento EOG implementado, com foco especial nas setas contínuas. A estrutura do sistema é semelhante à utilizada pelo sistema de reconhecimento EOG anterior, começando por se fazer uma filtragem dos dois sinais EMG recebidos (consulte a figura 4.5 para ver o posicionamento dos elétrodos). De seguida, no bloco “Detecção de expressão facial”, é feita também uma análise a partir de uma janela deslizante, seguida da extração de características. Por fim, é efetuada a classificação da expressão facial.

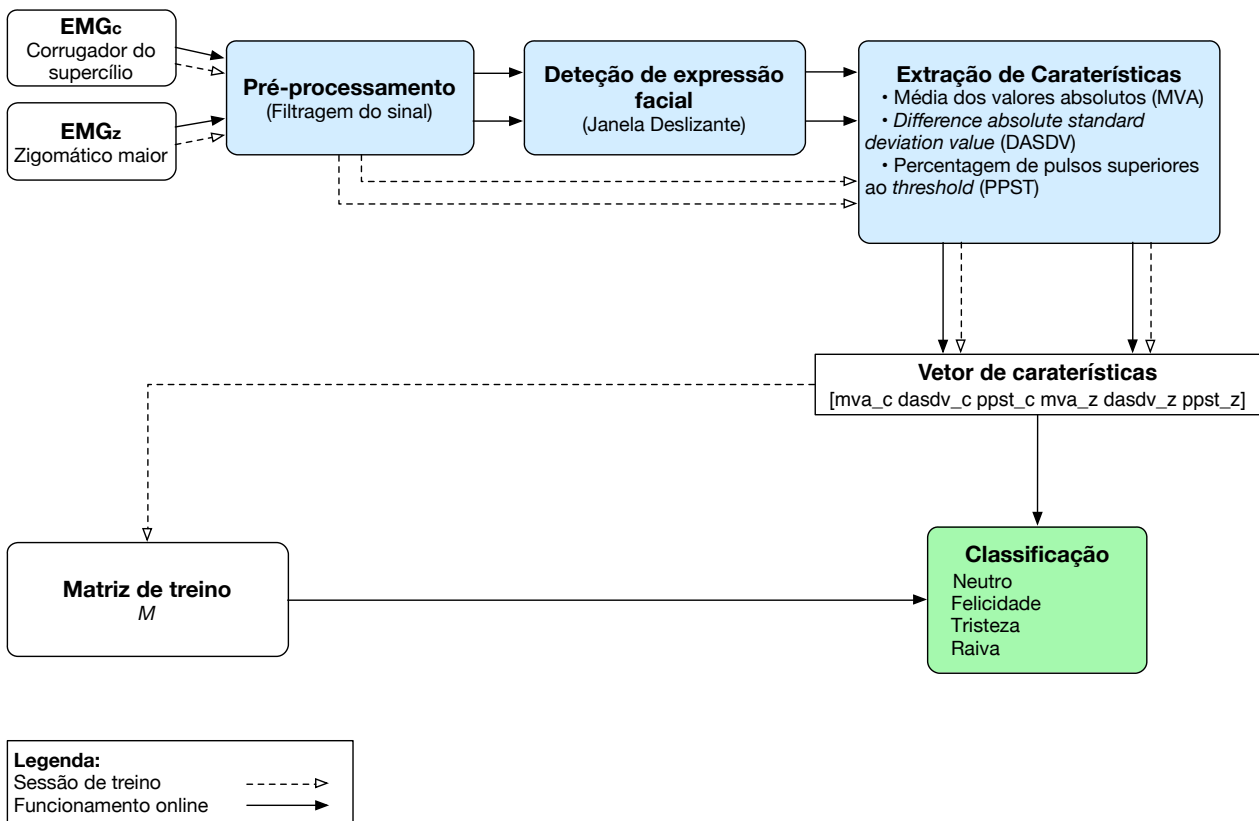


Figura 5.5: Estrutura do sistema de reconhecimento EMG

5.2.1 Treino do classificador

A sessão de treino é semelhante ao caso do EOG, explicado na secção anterior, e consiste na repetição das quatro expressões. Os participantes sentam-se em frente a um computador e são indicadas, no centro do ecrã, as expressões faciais pretendidas em forma de texto: “Neutro”, “Feliz”, “Triste” e “Raiva”. Cada expressão facial é realizada durante 4 segundos com um intervalo de repouso de 4 segundos entre cada expressão, tendo uma sequência de expressões a duração de 48 segundos. A sequência pode ser

vista na figura 5.6. Um *dataset* de uma sessão de treino tem 10 repetições da sequência, onde as expressões “neutro” e “feliz” são repetidas 10 vezes, ao invés das expressões “triste” e “raiva”, que são repetidas 20 vezes cada uma. Optou-se por gravar mais épocas das expressões “triste” e “raiva” por se tratarem de expressões que se distinguem pouco ao nível dos sinais gravados (como se pode comprovar pela figura 5.9).

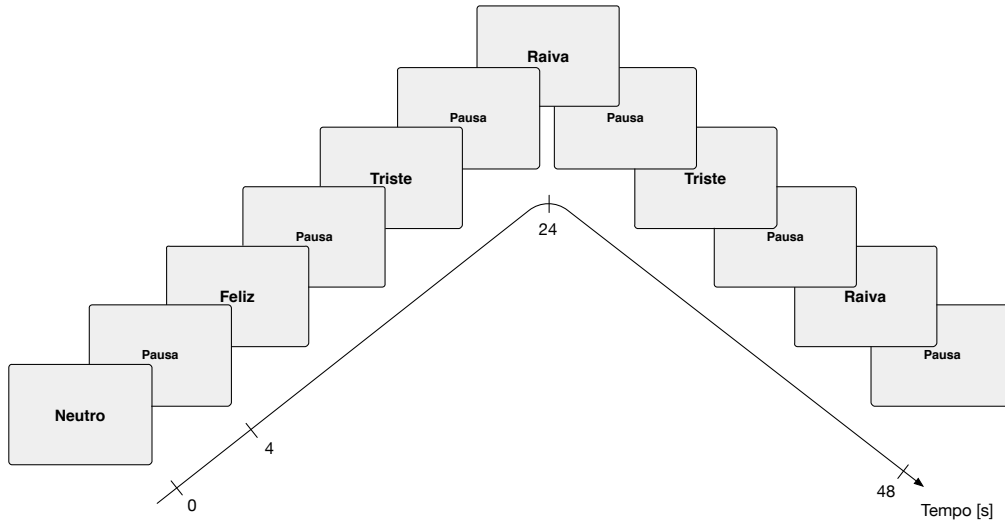


Figura 5.6: Sequência de expressões faciais de uma sessão de treino do classificador EMG

Para cada expressão realizada são extraídas as características, mencionadas no capítulo anterior, ponto 3.3.1. No caso do classificador EOG, a duração de uma época (2 segundos) é usada por completo no cálculo das características. No caso do classificador EMG, nem todo o sinal de uma época é útil. Considere a figura 5.7 onde é apresentada uma época da expressão facial “triste”, seguida de uma pausa. Aos 0 segundos, é apresentada a mensagem no ecrã “Triste”, aos 4 segundos é apresentada a mensagem ”Pausa”. É natural que, após a apresentação das mensagens, exista um pequeno atraso até ao momento em que a pessoa realiza a expressão. Portanto, no cálculo das características, apenas se considera a parte da época considerada útil, que corresponde ao último segundo de cada época (dos 3 aos 4 segundos no caso da figura 5.7).

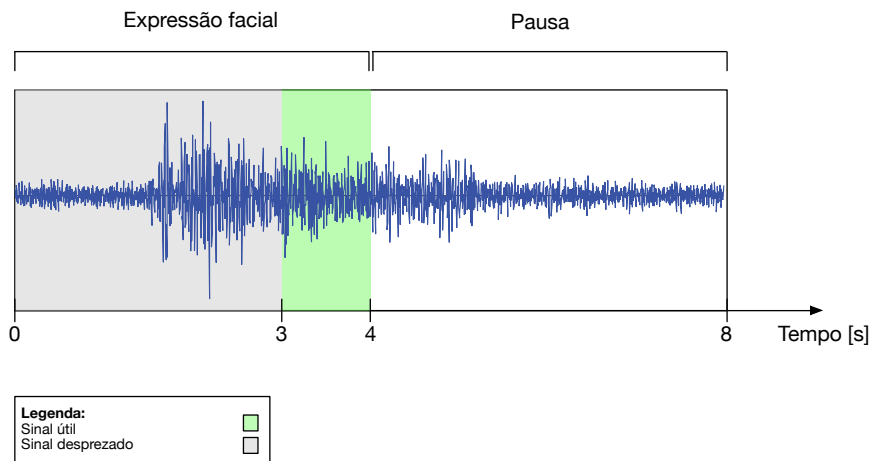


Figura 5.7: Sinal EMG do músculo corrugador do supercílio seccionado em duas épocas

Para concluir, é construída uma matriz de treino M , constituída por 60 linhas (repetição de movimentos em que cada linha corresponde ao vetor de características de uma expressão) com 6 colunas (três características por cada sinal EMG).

5.2.2 Pré-processamento

Os sinais EMG estão sujeitos aos ruídos que afetam também os sinais EOG que precisam de ser filtrados. No entanto, no caso do EOG, o sinal proveniente dos músculos, que é considerado ruído (e considerado um artefacto), no caso do EMG é o sinal que importa gravar. Do mesmo modo, o sinal do EOG proveniente da diferença de potencial entre a córnea e a retina do olho, é considerado ruído, e deve ser filtrado.

Para mitigar estas interferências foram aplicados um filtro *notch* de 50 Hz e um filtro *Butterworth* passa-banda de quarta ordem com frequência de corte mínima de 30 Hz e frequência de corte máxima de 100 Hz. Note-se que, agora, as frequências mais baixas (correspondentes ao movimento ocular) são filtradas.

5.2.3 Detecção de expressão facial

No caso do classificador EOG foi desenvolvido um método para detetar o início de um movimento, no sentido de ajustar o posicionamento da janela temporal, para que a extração de características fosse otimizada. Neste caso, foi utilizada uma abordagem diferente no desenvolvimento da janela deslizante. O indivíduo, durante o funcionamento online, nunca deixa de realizar uma determinada expressão facial (em último caso realiza a expressão “neutro”). Por este motivo, a janela deslizante deveria permitir a classificação, sem restrições, durante o funcionamento online. No entanto, existem alguns problemas aquando da transição entre uma expressão facial e outra, leia-se, a janela deslizante captura a transição entre duas expressões faciais, levando a uma extração de características errónea. Para resolver este problema, a janela deslizante é dividida em 8 sub-intervalos iguais e a média dos valores absolutos (equação 3.7) é calculada para cada sub-intervalo. A janela deslizante permite apenas a classificação quando a MVA dos 1º e 8º sub-intervalos de ambos os canais é idêntica, mais especificamente, quando a diferença entre as duas médias é menor que um determinado *threshold* (equação 5.3).

$$T = 0.2 \times \frac{MVA_1 + MVA_8}{2} \quad (5.3)$$

Onde MVA_1 e MVA_8 são a média dos valores absolutos nos sub-intervalos 1 e 8, respetivamente. A janela deslizante armazena os últimos 256 valores (1 segundo) do sinal EMG e desloca-se 8 amostras, para a frente no tempo, no final de cada avaliação do sinal EMG.

5.2.4 Extração de características

Após a seleção da janela temporal a testar, prossegue-se para o cálculo das características. Foram escolhidas três características pertencentes ao domínio do tempo, sendo estas a MVA, a DASDV, e a PPST, explicadas em detalhe na secção 3.3.1. Pode observar-se na figura 5.9 (b) que, quando o utilizador executa a expressão facial “felicidade”, existem picos com amplitude muito elevada, na ordem

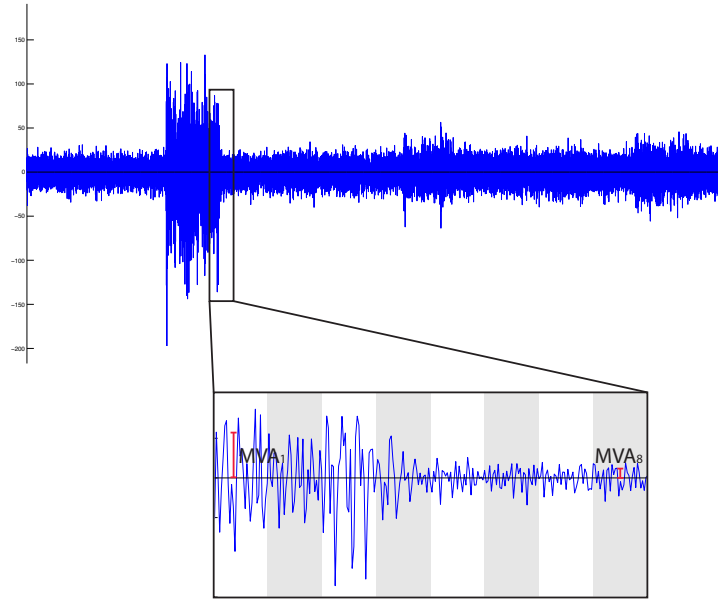


Figura 5.8: Sinal EMG do músculo zigomático maior. Este é um exemplo em que a janela deslizante não permite a classificação, porque está a acontecer a transição entre as expressões “smile” e “neutral”.

dos $100 \mu V$, que são facilmente distinguíveis do resto do sinal EMG do zigomático maior (associado ao sorriso e emoção positiva).

O vetor de teste (figura 5.5) é formado por seis características, três do sinal EMG, associado ao músculo corrugador do supercílio, e as restantes três do sinal do músculo zigomático maior, por esta ordem.

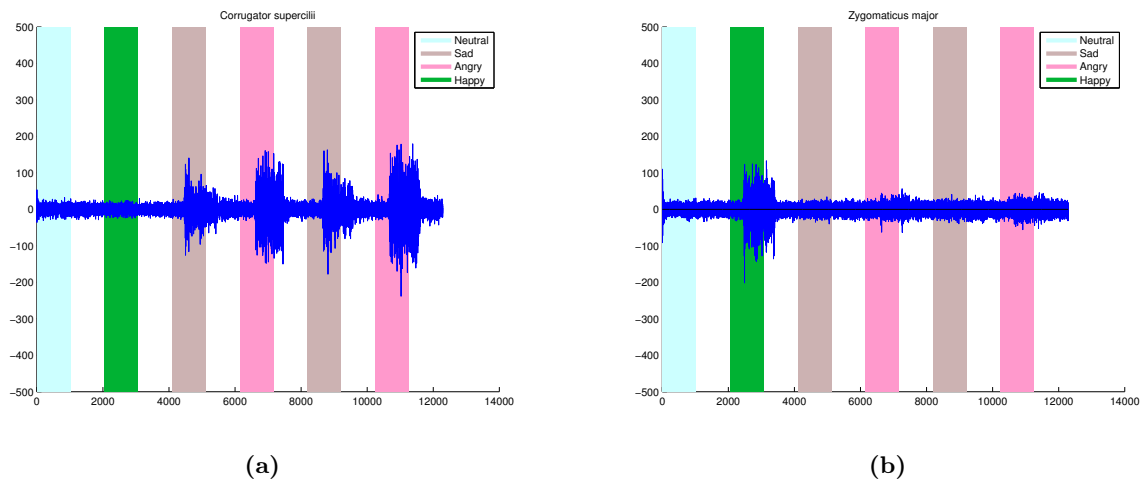


Figura 5.9: Sinais EMG: (a) do músculo corrugador do supercílio e (b) do músculo zigomático maior, para uma sequência de quatro expressões faciais gravadas durante uma sessão de treino: “neutro”, “feliz”, “triste” e “raiva”. Os espaços preenchidos a branco representam os períodos de repouso (não têm qualquer influência no treino do classificador)

5.2.5 Classificação

Por fim, é recebido o vetor de teste, concebido no ponto anterior, e classificado, numa das quatro expressões (“neutro”, “feliz”, “triste” ou “raiva”), por um classificador de Bayes multi-classe.

5.3 Seguimento ocular

Foi construído um sistema de seguimento ocular que determina a posição angular horizontal dos olhos de acordo com a montagem da figura 4.6. A estrutura do algoritmo desenvolvido é apresentada na figura 5.10. O sistema recebe o sinal EOG_h do canal horizontal que é sujeito a um pré-processamento (filtragem na banda de interesse e remoção de ruído). De seguida, é feita a diferenciação do sinal, onde, através da janela deslizante, é detetada a existência de sacadas. Após este processo, é calculada a amplitude da sacada (bem como a direção da mesma) e somada ao ângulo dos olhos antes da sacada. Pressupondo que o utilizador está a olhar “para a frente”, o ângulo dos olhos inicial é nulo.

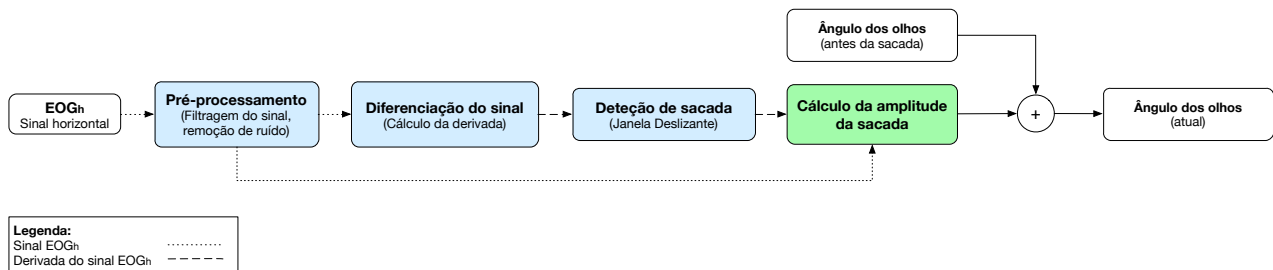


Figura 5.10: Estrutura do algoritmo de seguimento ocular horizontal.

5.3.1 Pré-processamento

Como já foi referido, o sinal EOG é suscetível a várias fontes de ruído, tendo também o conhecido *baseline drift*. Foi aplicado um filtro passa-baixo de 30 Hz e um filtro *notch* em 50 Hz, mantendo o sinal com componente DC (incluindo o *baseline drift*). Aquando da diferenciação de um sinal, o ruído de alta frequência é “amplificado”, pelo que é imperativo filtrá-lo, de forma a obter SNR (*Signal Noise Ratio*) elevado. O sinal foi filtrado com um filtro de Savitzky-Golay [MathWorks, 2014b] — filtro digital polinomial de 3^a ordem, com comprimento de 51 amostras — visto ter apresentado melhor desempenho, entre outros métodos (entre eles, o filtro de mediana). Na figura 5.11, pode ser visualizada a diferença do sinal EOG filtrado e o original.

5.3.2 Deteção de sacada

Na deteção de sacadas recorreu-se a uma janela deslizante que armazena as últimas 256 amostras do sinal, avançando 16 amostras, para a frente no tempo, a cada iteração. É feita a diferenciação do sinal da janela deslizante, dando origem a um sinal com picos (ver figura 5.11). São obtidos os valores absolutos dos picos, obedecendo a um conjunto de regras: têm de ser superiores a um *threshold* (foi utilizado o *threshold* igual a 1) e não podem existir dois ou mais picos num espaço de 20 amostras seguidas.

Tendo o conjunto de picos existentes na janela deslizante, e sabendo que cada pico representa uma sacada, à medida que a janela avança no tempo, o pico situado entre a amostra 120 e 139 (aproximadamente no centro da janela) é analisado para calcular a amplitude da sacada.

5.3.3 Cálculo da amplitude da sacada

A localização do pico é conhecida. Portanto, para calcular a amplitude da sacada, recorre-se ao sinal EOG filtrado e faz-se a diferença entre a média dos 50 valores antes e depois da ocorrência do pico da derivada (ver figura 5.11).

Este método apresenta uma abordagem com uma semelhança, em relação ao método apresentado na secção 2.2.2, em [Barea et al., 2002], onde é utilizada a derivada do sinal EOG, para calcular a amplitude da sacada (através da medição da amplitude dos picos). No método apresentado, nesta secção, usa-se apenas a derivada para detetar a ocorrência de sacadas.

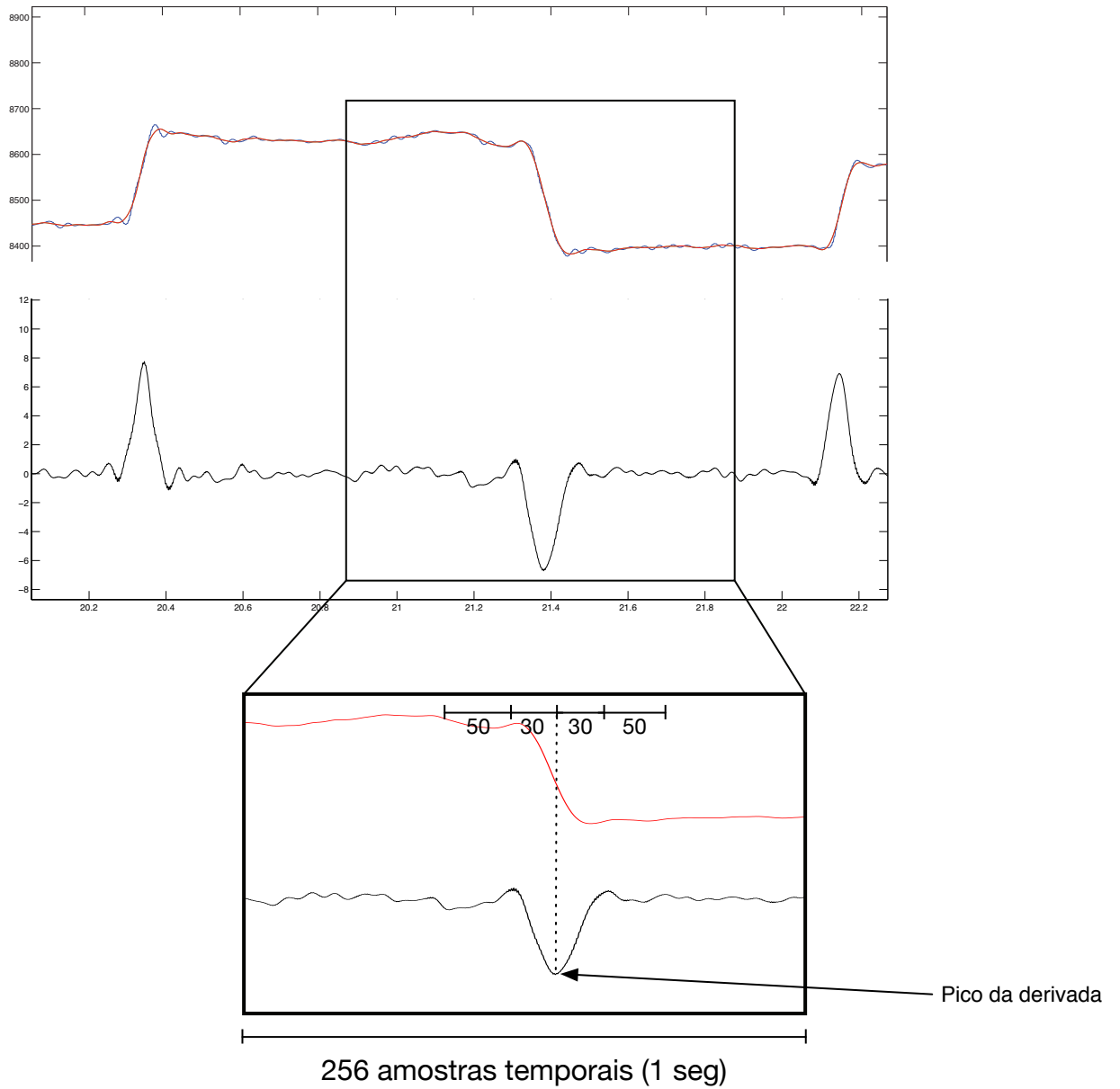


Figura 5.11: Sinal EOG com componente DC e respetiva derivada. Abaixo é apresentada a janela deslizante. Sinal sem filtro Savitzky-Golay, a azul; Sinal filtrado, a vermelho; Derivada do sinal, a preto.

Capítulo 6

Testes e resultados experimentais

Este capítulo apresenta resultados de alguns testes, com o objetivo de caracterizar a performance dos sistemas implementados. Todos os testes foram realizados em condições semelhantes, isto é, num ambiente calmo, com luminosidade média que não influenciasse as expressões faciais efetuadas.

6.1 Sistema de movimentos oculares

Os testes foram realizados por três pessoas saudáveis com idades compreendidas entre 23 e 28 anos.

6.1.1 Funcionamento *offline*

A tabela 6.1 apresenta uma matriz de confusão do classificador com as oito características utilizadas (domínio do tempo e da frequência) durante o funcionamento *offline*. Neste modo de funcionamento, a janela deslizante não tem qualquer influência nos resultados, porque se sabe *a priori* em que intervalo temporal o classificador deve agir.

Tabela 6.1: Matriz de confusão de classificação *offline* de movimentos oculares.

		Movimento classificado					
		Piscar	Franzir	Direita	Esquerda	Cima	Baixo
Mov. efetuado	Piscar	191	7	0	0	1	1
	Franzir	11	176	4	1	5	3
	Direita	2	4	192	0	1	1
	Esquerda	2	7	1	189	1	0
	Cima	1	10	1	0	176	12
	Baixo	2	7	0	1	7	183

Tabela 6.2: Desempenho *offline* do sistema de reconhecimento de movimentos oculares.

	Piscar	Franzir	Direita	Esquerda	Cima	Baixo	Média
Sensibilidade	95.5	88.0	96.0	94.5	88.0	91.5	92.3
Especificidade	99.1	97.7	99.2	98.9	97.7	98.3	98.5
Exatidão	97.8	95.1	98.8	98.8	96.8	97.2	97.4

6.1.2 Funcionamento *online*

No funcionamento *online*, o primeiro passo do sistema de reconhecimento de movimentos é a deteção de movimento, através da janela deslizante. Durante os testes, a taxa de falsos positivos e de falsos negativos foi de 4,9% e 15,4%, respetivamente.

A performance do sistema de classificação *online* é apresentada na tabela 6.3. O sujeito 3 obteve o melhor desempenho, com uma taxa de sensibilidade de 88,1%. Estes resultados refletem o desempenho da janela deslizante, como era esperado, baixando a performance geral do sistema.

Tabela 6.3: Performance da classificação *online* de cada sujeito.

	Sensibilidade	Especificidade	Exatidão
Sujeito 1	86.9	97.5	98.0
Sujeito 2	77.7	97.5	97.9
Sujeito 3	88.1	98.8	98.9
Média	84.7	97.9	98.3

6.2 Sistema de reconhecimento de expressões faciais

Para testar o desempenho deste sistema, foram realizadas, de forma arbitrária, várias expressões faciais por um participante (ver figura 6.1), cujos os resultados foram registados da seguinte maneira: realizava-se uma expressão voluntariamente e, de seguida, registava-se a expressão, classificada pelo sistema. Os resultados do teste do sistema de classificação *offline* são apresentados, em forma de matriz de confusão, na tabela 6.4, e os resultados, em funcionamento *online*, são apresentados na tabela 6.5.

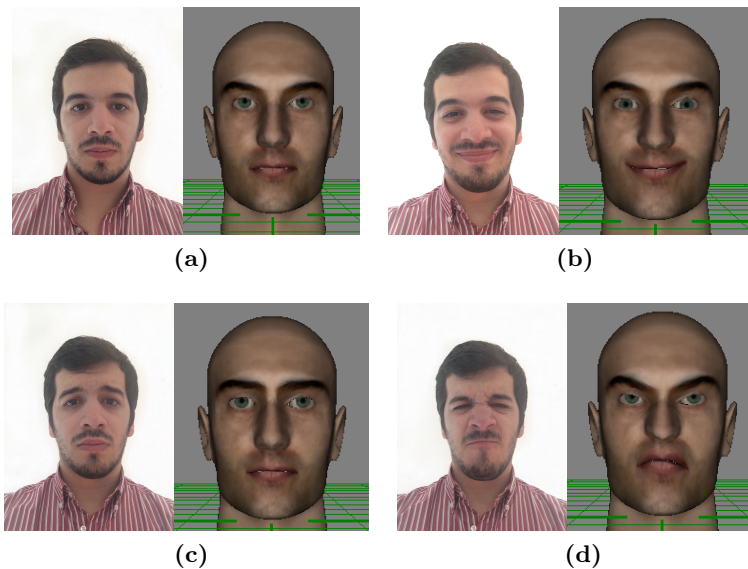


Figura 6.1: Fotografia das expressões realizadas pelo autor da tese e mimetizações do avatar: (a) expressão neutra, (b) expressão de felicidade, (c) expressão de tristeza e (d) expressão de raiva.

Tabela 6.4: Matriz de confusão do sistema de classificação de expressões faciais (em modo de funcionamento *offline*).

	Expressão classificada				Exatidão	
	Neutra	Felicidade	Tristeza	Raiva		
Exp. efetuada	Neutra	10	0	0	0	10/10 = 100%
Felicidade	0	10	0	0	10/10 = 100%	
Tristeza	0	0	20	0	20/20 = 100%	
Raiva	0	0	0	20	20/20 = 100%	

Tabela 6.5: Matriz de confusão de um teste de classificação de expressões faciais.

	Expressão classificada				Exatidão	
		Neutra	Felicidade	Tristeza		Raiva
Exp. efetuada	Neutra	100	0	0	0	100/100 = 100%
	Felicidade	0	50	0	0	50/50 = 100%
	Tristeza	0	0	41	9	41/50 = 82%
	Raiva	0	0	0	50	50/50 = 100%

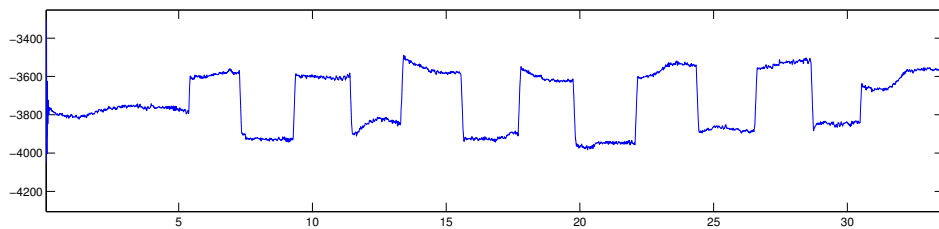
Todas as expressões são identificadas com bom desempenho pelo classificador, utilizando apenas três características para cada sinal EMG (dos músculos corrugador do supercílio e zigomático maior). Como seria de esperar, existe um déficit de exatidão na classificação da expressão “tristeza”, que é confundida com a expressão de “raiva”, no modo de funcionamento *online* (tabela 6.5). Isto deve-se à grande semelhança entre os sinais dos dois músculos. Existe uma grande atividade do músculo corrugador do supercílio (sobrancelha) e muito pouca do zigomático maior. Em relação ao sistema de classificação *offline*, obteve-se exatidão de 100% em todas as expressões. Neste caso, o classificador só classifica de 4 em 4 segundos, intervalo em que são extraídas as características no ato de calibração do sistema (consultar secção 5.1.1). Por este motivo, são esperados excelentes resultados de classificação.

O desempenho da janela deslizante deste sistema foi, também, avaliado. Foi dito, na secção 5.2.3, que a função da janela deslizante seria a de rejeitar a classificação entre transições de expressões faciais. No decorrer dos testes, observa-se que a janela deslizante desempenhou a sua função com sucesso.

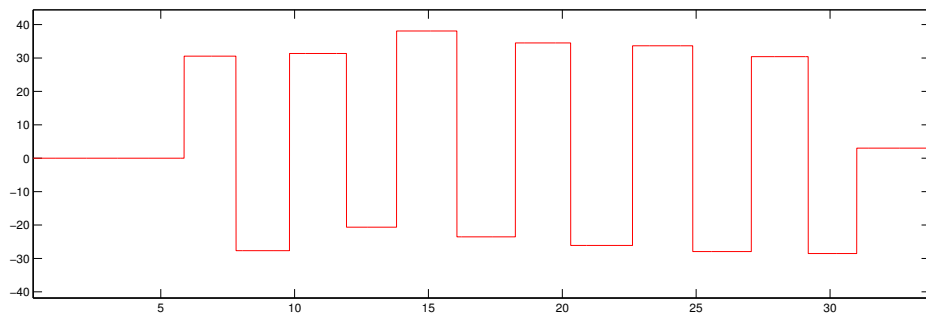
6.3 Sistema de seguimento ocular

O sistema de seguimento ocular é complexo de avaliar em termos de performance. O método adotado foi: o sujeito olha inicialmente para o centro do ecrã de um computador (corresponde a um ângulo horizontal nulo); de seguida, olha para a direita e para a esquerda do ecrã, várias vezes periodicamente (do centro para a direita corresponde a 30 graus aproximadamente; da direita para a esquerda e vice-versa, corresponde a uma rotação de 60 graus); finalmente, o sujeito olha para o centro do ecrã, finalizando, assim, o teste. Na figura 6.2a, é apresentado o sinal EOG horizontal gravado durante o teste apenas com um filtro passa-baixo de 30 Hz e um *notch* de 50 Hz. As sacadas são bem visíveis, bem como o *baseline drift* que ocorre entre sacadas. Na figura 6.2b, é visível o gráfico da posição horizontal dos olhos, ao longo do tempo. Pode observar-se algumas oscilações acima e abaixo dos 30 graus e -30 graus, que podem ser índice do erro do método aplicado, ou ap facto do sujeito não ter olhado para o local desejado.

A posição final dos olhos é um dado importante, neste teste. Começou-se na posição de 0 graus e finalizou-se em 3,033 graus. Estes resultados são encorajadores, uma vez que foram feitas 13 sacadas.



(a)



(b)

Figura 6.2: Teste *online* do sistema de seguimento ocular: (a) Sinal EOG; (b) Posição angular obtida.

Capítulo 7

Conclusões e trabalho futuro

Na fase inicial desta tese, foram propostos objetivos, designadamente o desenvolvimento de dois sistemas: um de reconhecimento automático de expressões faciais, utilizando sinais eletromiográficos faciais, e outro de reconhecimento de movimentos oculares, recorrendo aos sinais eletro-oculográficos. A estes, foi possível acrescentar um módulo para seguimento ocular, testado apenas para movimentos horizontais. Também foi proposto que se juntassem os dois sistemas acima mencionados, para que funcionassem em simultâneo. Estes subsistemas fazem parte de um sistema maior. Para mimetizar as expressões faciais e os movimentos oculares reconhecidos pelos subsistemas, bem como a reprodução dos movimentos da cabeça do utilizador, apresentou-se como objetivo primordial o desenvolvimento de um avatar 3D controlável.

Todos os objetivos propostos foram alcançados. Um dos movimentos que é reconhecido, pelo sistema de reconhecimento de movimentos oculares, é o “franzir”. A execução deste movimento induz um sinal eletromiográfico, detetado também no sistema de reconhecimento de expressões faciais, facto que dificultou a junção dos dois sistemas, impedindo este último subsistema de funcionar corretamente. Por isso, os dois subsistemas foram implementados de forma independente.

O sistema de reconhecimento de expressões faciais foi implementado, recorrendo a sinais EMG de dois músculos. O seu desempenho e a sua robustez são bastante satisfatórios, no reconhecimento de quatro expressões faciais. Quanto ao sistema de reconhecimento de movimentos oculares, são detetados, com boa robustez, quatro movimentos oculares, bem como o “piscar de olhos” e “franzir”. Este último movimento, nunca foi classificado nos sistemas de reconhecimento, baseados em EOG. O algoritmo de seguimento ocular apresenta boa robustez, na deteção de sacadas horizontais e no cálculo da sua amplitude, sendo viável para monitorizar a posição angular dos olhos. Este sistema apresenta, ainda, algumas limitações: não ter sido, ainda, implementado para sinais EOG verticais (o “piscar de olhos” introduz artefactos no sinal) e não detetar sacadas de amplitude muito baixa (normalmente confundem-se com ruído). No entanto, o algoritmo apresentado corresponde a uma abordagem que ainda não foi apresentada em artigos, portanto representa uma abordagem inovadora.

O avatar 3D, desenvolvido para funcionamento *online* no Vizard, tem o comportamento esperado, recebendo a informação dos outros subsistemas e mimetizando, rapidamente, a expressão ou o movimento ocular. A operação do IMU, aplicado na cabeça do utilizador, apresenta uma resposta rápida e realista dos movimentos da cabeça do mesmo.

Apresentam-se como propostas, para um trabalho futuro:

- Testar o sistema de expressões faciais com mais de quatro expressões (talvez as seis expressões básicas);
- Melhorar o desempenho do sistema de expressões faciais, através da utilização de mais pares de elétrodos em diferentes músculos;

-
- Desenvolver um sistema ultra-portátil, e pouco intrusivo, para a leitura de biossinais;
 - Reconhecer visemas (expressão bucal quando falamos), através da utilização de sinais eletromiográficos de músculos na zona bucal;
 - Juntar a informação recolhida pelos sistemas desenvolvidos na tese para inferir emoções e estados afetivos.

Bibliografia

- [Ang et al., 2004] Ang, L. B. P., Belen, E., Bernardo, R., Boongaling, E., Briones, G., and Coronel, J. (2004). Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors. *2004 IEEE Region 10 Conference TENCON 2004.*, C:600–603. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1414843>.
- [Aungsakul et al., 2012] Aungsakul, S., Phinyomark, a., Phukpattaranont, P., and Limsakul, C. (2012). Evaluating Feature Extraction Methods of Electrooculography (EOG) Signal for Human-Computer Interface. *Procedia Engineering*, 32:246–252. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1877705812012957>.
- [Aungsakun et al., 2011] Aungsakun, S., Angkoon, P., Phukpattaranont, P., and Chusak, L. (2011). Robust Eye Movement Recognition Using EOG Signal for Human-Computer Interface. In *Software Engineering and Computer Systems*, volume 180, pages 714–723.
- [Barea et al., 2002] Barea, R., Boquete, L., Mazo, M., and López, E. (2002). Wheelchair Guidance Strategies Using EOG. pages 279–299.
- [Barea et al., 2011] Barea, R., Boquete, L., Rodriguez-Ascariz, J. M., Ortega, S., and López, E. (2011). Sensory system for implementing a human-computer interface based on electrooculography. *Sensors (Basel, Switzerland)*, 11(1):310–28. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3274094&tool=pmcentrez&rendertype=abstract>.
- [Bettadapura, 2012] Bettadapura, V. (2012). Face expression recognition and analysis: the state of the art. pages 1–27. Available from: <http://arxiv.org/abs/1203.6722>.
- [Boot, 2009] Boot, L. (2009). *Facial expressions in EEG/EMG recordings*. PhD thesis, University of Twente. Available from: <http://essay.utwente.nl/58633/>.
- [Bulling, 2010] Bulling, A. (2010). *Eye Movement Analysis for Context Inference and Cognitive-Awareness: Wearable Sensing and Activity Recognition Using Electrooculography*. PhD thesis, Technical University of Karlsruhe.
- [Bulling et al., 2007] Bulling, A., Herter, P., Wirz, M., and Troster, G. (2007). Automatic artefact compensation in EOG signals. pages 3–4. Available from: https://www.andreas-bulling.de/fileadmin/docs/bulling07_eurossc.pdf.
- [Bulling et al., 2009] Bulling, A., Roggen, D., and Tröster, G. (2009). Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments. *Ambient Intelligence and Smart Environments*. Available from: <http://iospress.metapress.com/index/b338217432901653.pdf>.
- [Bulling et al., 2011] Bulling, A., Ward, J. a., Gellersen, H., and Tröster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE transactions on pattern analysis and machine intelligence*, 33(4):741–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20421675>.

- [Chin et al., 2008] Chin, Z. Y., Ang, K. K., and Guan, C. (2008). Multiclass voluntary facial expression classification based on Filter Bank Common Spatial Pattern. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2008:1005–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19162828>.
- [Du et al., 2014] Du, S., Tao, Y., and Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15):E1454–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24706770>.
- [Duchowski, 2007] Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [Duda et al., 2001] Duda, R., Hart, P., and Stork, D. (2001). *Pattern classification*. Pattern Classification and Scene Analysis: Pattern Classification. Wiley. Available from: <http://books.google.pt/books?id=YoxQAAAAMAAJ>.
- [Ekman and Friesen, 1978a] Ekman and Friesen (1978a). FACS - Facial Action Coding System. Available from: <http://www.cs.cmu.edu/~face/facs.htm>.
- [Ekman, 1973] Ekman, P. (1973). *Darwin and Facial Expression: A Century of Research in Review*. Academic Press. Available from: <http://books.google.pt/books?id=2JtpAAAAMAAJ>.
- [Ekman et al., 1999] Ekman, P., Dalgleish, T., and Power, M. J. (1999). *Handbook of Cognition and Emotion*. John Wiley & Sons, Ltd, Chichester, UK. Available from: <http://doi.wiley.com/10.1002/0470013494>.
- [Ekman et al., 1978] Ekman, P., Friesen, W., and Hager, J. (1978). *Facial Action Coding System Investigator's Guide*. Consulting Psychologist Press, Palo Alto, CA. <http://face-and-emotion.com/dataface/facs/guide/FACSIVTi.html>.
- [Ekman and Friesen, 1978b] Ekman, P. and Friesen, W. V. (1978b). Facial action coding system: a technique for the measurement of facial movement. *Consulting Psychologists Press*.
- [Fridlund and Cacioppo, 1986] Fridlund, A. and Cacioppo, J. (1986). Guidelines for human electromyographic research. *Psychophysiology*, 23:567–589. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1986.tb00676.x/full>.
- [Gibert et al., 2009] Gibert, G., Pruzinec, M., and Stevens, C. (2009). Enhancement of Human Computer Interaction with facial Electromyographic sensors Locked Bag 1797. pages 2–5.
- [g.tec medical engineering GmbH, 2013] g.tec medical engineering GmbH (2013). Product Catalogue 2013/14.
- [Haslwanter, 2012] Haslwanter, T. (2012). Rotation Matrices. Available from: <http://work.thaslwanter.at/Kinematics/html/kinRotMats.html>.

- [Iáñez et al., 2013] Iáñez, E., Azorin, J. M., and Perez-Vidal, C. (2013). Using eye movement to control a computer: a design for a lightweight electro-oculogram electrode array and computer interface. *PloS one*, 8(7):e67099. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3700965&tool=pmcentrez&rendertype=abstract>.
- [Johns et al., 2007] Johns, M. W., Tucker, a., Chapman, R., Crowley, K., and Michael, N. (2007). Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie - Schlafforschung und Schlafmedizin*, 11(4):234–242. Available from: <http://link.springer.com/10.1007/s11818-007-0311-y>.
- [Karson et al., 1981] Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., and Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry research*, 5(3):243–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6948307>.
- [Kotsia et al., 2008] Kotsia, I., Buciu, I., and Pitas, I. (2008). An analysis of facial expression recognition under partial facial image occlusion. *Image and Vision Computing*, 26(7):1052–1067. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0262885607002107>.
- [Kotsia and Pitas, 2007] Kotsia, I. and Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, 16(1):172–187. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4032815.
- [Liu et al., 2010] Liu, Y., Sourina, O., and Nguyen, M. K. (2010). Real-Time EEG-Based Human Emotion Recognition and Visualization. *2010 International Conference on Cyberworlds*, pages 262–269. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5656346>.
- [Masahiro et al., 2012] Masahiro, B., Translated, M., and Kageki, N. (2012). The Uncanny Valley. pages 98–100.
- [MathWorks, 2014a] MathWorks (2014a). Discriminant Analysis. Available from: <http://www.mathworks.com/help/stats/discriminant-analysis.html#bs31msq>.
- [MathWorks, 2014b] MathWorks (2014b). Savitzky-Golay filter design. Available from: <http://www.mathworks.com/help/signal/ref/sgolay.html>.
- [Ohno and Mukawa, 2003] Ohno, T. and Mukawa, N. (2003). A Free-head , Simple Calibration , Gaze Tracking System That Enables Gaze-Based Interaction.
- [Partala et al., 2006] Partala, T., Surakka, V., and Vanhala, T. (2006). Real-time estimation of emotional experiences from facial expressions. *Interacting with Computers*, 18:208–226. Available from: <http://iwc.oxfordjournals.org/cgi/doi/10.1016/j.intcom.2005.05.002>.
- [Pettersson et al., 2013] Pettersson, K., Jagadeesan, S., Lukander, K., Henelius, A., Haeggström, E., and Müller, K. (2013). Algorithm for automatic analysis of electro-oculographic data. *Biomedical engineering online*, 12:110. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3830504&tool=pmcentrez&rendertype=abstract>.

- [Phinyomark et al., 2012] Phinyomark, A., Phukpattaranont, P., and Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8):7420–7431. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0957417412001200>.
- [Psychologist World, 2014] Psychologist World (2014). Eye Reading (Body Language). Available from: <http://www.psychologistworld.com/bodylanguage/eyes.php>.
- [Robinson, 1963] Robinson, D. A. (1963). A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field. pages 137–145.
- [Sarver, 2001] Sarver, J. J. (2001). EMG Analysis. Available from: http://www.biomed.drexel.edu/labs/biomechanics/emg_analy.htm.
- [Schels et al., 2013] Schels, M., Kächele, M., Glodek, M., Hrabal, D., Walter, S., and Schwenker, F. (2013). Using unlabeled data to improve classification of emotional states in human computer interaction. *Journal on Multimodal User Interfaces*, 8:5–16. Available from: <http://link.springer.com/10.1007/s12193-013-0133-0>.
- [Schiffman, 2000] Schiffman, H. R. (2000). *Sensation and Perception: An Integrated Approach*. Wiley, John & Sons, Incorporated, 5 edition.
- [SMI,] SMI. SensoMotoric Instruments homepage. Available from: <http://www.smivision.com/en.html>.
- [Stangor, 2014] Stangor, C. (2014). Introduction to Psychology, v. 1.0. Available from: http://catalog.flatworldknowledge.com/bookhub/127?e=stangor-ch10_s01.
- [Tian et al., 2001] Tian, Y.-I., Kanade, T., and Cohn, J. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=908962>.
- [Tobii, 2013] Tobii (2013). Eye tracking products. Available from: <http://www.tobii.com/en/eye-tracking-research/global/products/>.
- [Van Boxtel, 2001] Van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, 38(1):22–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11321618>.
- [van der Geest and Frens, 2002] van der Geest, J. N. and Frens, M. a. (2002). Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods. *Journal of neuroscience methods*, 114(2):185–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11856570>.
- [Zeng et al., 2009] Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19029545>.

[Zweig, 2011] Zweig, A. (2011). *Controlling a crane arm with EMG sensors*. Bachelor-thesis, Swiss Federal Institute of Technology Zurich.