

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ENGENHARIA
CAMPUS DE ILHA SOLTEIRA

JOÃO ANGELO FERRES BROGIN

GENERALIZAÇÃO DA TÉCNICA DE CORRELAÇÃO CANÔNICA PARA
APLICAÇÕES EM INTERFACE CÉREBRO-MÁQUINA

ILHA SOLTEIRA

2018

JOÃO ANGELO FERRES BROGIN

**GENERALIZAÇÃO DA TÉCNICA DE CORRELAÇÃO CANÔNICA PARA
APLICAÇÕES EM INTERFACE CÉREBRO-MÁQUINA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Mecânica, Faculdade de Engenharia de Ilha Solteira - UNESP, como parte dos requisitos necessários para obtenção do título de Mestre em Engenharia Mecânica.

Orientador: Prof. Dr. Douglas Domingues Bueno

Coorientador: Prof. Dr. Samuel da Silva

Coorientador: Prof. Dr. Jean Faber Ferreira de Abreu

ILHA SOLTEIRA

2018

FICHA CATALOGRÁFICA
Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

B866g Brogin, João Angelo Ferres.
Generalização da técnica de correlação canônica para aplicações em interface cérebro-máquina / João Angelo Ferres Brogin. -- Ilha Solteira: [s.n.], 2018
151 f. : il.

Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Ciências Térmicas, 2018

Orientador: Douglas Domingues Bueno
Coorientadores: Samuel da Silva e Jean Faber Ferreira de Abreu
Inclui bibliografia

1. Análise de sinais. 2. Interface cérebro-máquina. 3. Eletroencefalograma. 4. Potenciais visuais Evocados. 5. Análise de correlação canônica.

Raiane da Silva Santos
Raiane da Silva Santos

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Generalização da Técnica de Correlação Canônica para Aplicações em Interface Cérebro-Máquina

AUTOR: JOÃO ANGELO FERRES BROGIN
ORIENTADOR: DOUGLAS DOMINGUES BUENO
COORIENTADOR: SAMUEL DA SILVA
COORIENTADOR: JEAN FABER FERREIRA DE ABREU

Aprovado como parte das exigências para obtenção do Título de Mestre em ENGENHARIA MECÂNICA, área: Ciências Térmicas pela Comissão Examinadora:



Prof. Dr. DOUGLAS DOMINGUES BUENO
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira - UNESP



Profa. Dra. KARINA RABELLO CASALI
Departamento de Engenharia Biomédica / Universidade Federal de São Paulo



Prof. Dr. DIOGO COUTINHO SORIANO
Departamento de Engenharia Biomédica / Universidade Federal do ABC

Ilha Solteira, 17 de dezembro de 2018

Agradecimentos

Em primeiro lugar, agradeço a meus pais, Maria e João, por sempre batalharem em busca de uma educação de qualidade para seus filhos, a despeito de todas as dificuldades. Essa conquista, antes de tudo, pertence a vocês.

Agradeço, também, a minhas irmãs Fabiana e Juliana, por todo o carinho e paciência que sempre tiveram, contribuindo imensamente para meu crescimento pessoal.

Ao meu orientador, professor Douglas Domingues Bueno, por todo o conhecimento e confiança que sempre depositou em mim, contribuindo para meu desenvolvimento acadêmico e profissional, bem como construção de autonomia e senso crítico.

Aos meus coorientadores, professor Samuel da Silva e professor Jean Faber Ferreira de Abreu, por todo o suporte e atenção ao trabalho.

A todos os funcionários e professores da Faculdade de Engenharia de Ilha Solteira da Universidade Estadual Paulista “Júlio de Mesquita Filho”, bem como à Universidade Federal de São Paulo, pela colaboração neste trabalho.

Ao Programa de Pós-Graduação em Engenharia Mecânica e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico por todo o apoio necessário.

Por fim, obrigado a todos os colegas e amigos por sempre estarem ao meu lado, pelas horas de estudo e aprendizado, pelas horas de lazer, e por participarem ativamente no Programa de Pós-Graduação em Engenharia Mecânica.

Au milieu de l'hiver, j'apprenais enfin qu'il y avait en moi un été invincible

Albert Camus

Resumo

A busca por uma melhor compreensão das regiões do cérebro e suas funções nas ações humanas tem sido uma tarefa árdua, porém muito útil, principalmente para aplicações da engenharia de interface cérebro-máquina (ICM), bem como para o auxílio a diagnósticos médicos a partir de sinais obtidos dos pacientes em avaliação. No contexto do presente trabalho, destacam-se os trabalhos de interface cérebro-máquina (ICM) pela abrangência no envolvimento de técnicas, métodos e ferramentas comumente estudadas nos cursos de engenharia. Em particular, análises envolvendo técnicas de processamento de sinais de eletroencefalograma (EEG) têm se mostrado de significativa importância para o desenvolvimento dessa área. Uma abordagem amplamente utilizada nesse contexto é a ICM usando Potenciais Visuais Evocados de Estados Estacionários (SSVEP, do inglês *Steady-State Visual Evoked Potentials*), que, de forma geral, são sinais caracterizados pela resposta evocada do cérebro a estímulos visuais modulados em uma frequência específica. Assim, este trabalho tem o objetivo de propor uma generalização do coeficiente de correlação, conceito-base da análise de correlação canônica (CCA), técnica que tem se mostrado robusta e eficiente no reconhecimento de padrões, especialmente no caso dos SSVEP, e detalhar seu comportamento em função dos parâmetros relevantes para se estabelecer melhores práticas de uso em aplicações de ICM, incluindo fatores fisiológicos, técnicos e operacionais.

Palavras-chave: Análise de sinais. Interface cérebro-máquina. Eletroencefalograma. Potenciais visuais evocados. Análise de correlação canônica.

Abstract

The search for a better understanding of the brain's anatomy and its functions on human actions has been a harsh yet very useful task, especially for brain-computer interface engineering applications, as well as for medical diagnosis using signals from patients. In the context of this work, brain-computer interface (BCI) applications are highlighted due to their comprehensiveness related to techniques, methods and tools commonly studied in engineering. In particular, analyses involving electroencephalogram (EEG) signals processing have proven to be of great significance for developing this field of study. A widely used approach is Steady State Visual Evoked Potentials (SSVEP) based BCI, which, in general, are signals characterized by the brain's evoked response to visual stimuli modulated at a certain frequency. This work aims thus to propose a generalization of the correlation coefficient, which entails Canonical Correlation Analysis (CCA), a technique that has presented robustness and efficiency for pattern recognition, especially in SSVEP-based BCIs, and describe its behavior under relevant varying parameters to establish better use practices in BCI applications, comprising physiological, technical and operational factors.

Keywords: Signal analysis. Brain-computer interface. Electroencephalogram. Visual evoked potentials. Canonical correlation analysis.

Lista de Símbolos

f_i	- Primeira frequência de um trio de frequências-alvo
f_j	- Segunda frequência de um trio de frequências-alvo
f_k	- Terceira frequência de um trio de frequências-alvo
f_{alvo}	- Frequência alvo para um trio de frequências
R	- Taxa de atualização da imagem na tela do computador
$P(*)$	- Probabilidade
x	- Nível de amplitude qualquer
x_0	- Amplitude do sinal, em μV
t	- Tempo, em s
d	- Incremento de variável qualquer
T	- Período, em s
$p(*)$	- Função Densidade de Probabilidade
$F(*)$	- Função de Distribuição Acumulada
$f(*)$	- Função Densidade de Probabilidade
a	- Constante real qualquer
b	- Constante real qualquer
$erf(*)$	- Função erro
$g(*)$	- Momento de uma variável aleatória
$E[*]$	- Operador expectativa
D	- Variável aleatória qualquer
w	- Variável aleatória qualquer
m_r	- Coeficiente angular da reta
$\tilde{\mathbf{x}}$	- Variável canônica de x
$\tilde{\mathbf{y}}$	- Variável canônica de y
\mathbf{w}_x	- Autovetores do banco de sinais real
\mathbf{w}_y	- Autovetores do banco de sinais artificial
\mathbf{X}	- Conjunto de sinais reais do sistema BCI
\mathbf{Y}_m	- Conjunto de sinais artificial
E	- Número de canais/eletrodos
P	- Número de pontos

M	-	Constante real qualquer
f_m	-	Frequência de estímulo, em Hz
F_s	-	Frequência de amostragem do banco de sinais artificial, em Hz
H	-	Número de harmônicas
$\hat{\rho}$	-	Autovalores da CCA
\mathbf{S}	-	Matriz de covariância
\mathbf{A}	-	Matriz do problema de autovalores (CCA)
$\hat{\mathbf{x}}$	-	Matriz de autovetores qualquer
Λ	-	Matriz de autovalores qualquer
\hat{f}	-	Frequência reconhecida, em Hz
x_n	-	Amplitude das componentes da senoide do banco real
x_m	-	Amplitude das componentes da senoide do banco artificial
k	-	Constante qualquer $\in \mathbb{Z}$
F_i	-	Frequência de amostragem do conjunto de sinais reais, em Hz
F	-	Razão de frequência de amostragem
t_{max}	-	Tempo máximo de janela temporal, em s
n_P	-	Número de períodos
BW	-	Largura de banda em torno da frequência de estímulo, em Hz
FA	-	Fator de acurácia
$n_{P,ot}$	-	Número ótimo de períodos
$t_{max,ot}$	-	Tempo ótimo de janela temporal, em s
$f_{m,ot}$	-	Frequência ótima de estímulo, em Hz
f_r	-	Frequência reconhecida, em Hz
f	-	Razão entre frequência reconhecida e de estímulo
\mathbf{w}	-	Vetor direção, autovetor ou vetor perpendicular ao hiperplano para SVM
\mathbf{x}_p	-	Série de dados qualquer
\mathbf{x}_P	-	Projeção de dados no subespaço de componentes principais
\bar{x}	-	Valor médio para série discreta de dados
$L(\ast)$	-	Lagrangiana
d_e	-	Distância Euclidiana
\mathbf{p}_{CP}	-	Vetor de projeção dos dados qualquer
\mathbf{q}_t	-	Vetor teste qualquer
d_m	-	Distância de Mahalanobis
$\hat{\mathbf{S}}$	-	Matriz de covariância após projeção (LDA ou QDA)

\mathbf{S}_b	- Matriz de dispersão interclasse
$\tilde{\mathbf{S}}_b$	- Matriz de dispersão interclasse após projeção (LDA ou QDA)
\mathbf{S}_w	- Matriz de dispersão intraclasse
$\tilde{\mathbf{S}}_w$	- Matriz de dispersão intraclasse após projeção (LDA ou QDA)
$J(\mathbf{w})$	- discriminante linear de Fisher
w_0	- Constante real qualquer
\mathbf{b}^T	- Vetor de constantes para a equação de fronteira
\mathbf{c}^T	- Vetor de constantes para a equação de fronteira
\mathbf{u}	- Vetor teste de entrada (SVM)
\mathbf{x}_+	- Vetor de suporte de amostra positiva
\mathbf{x}_-	- Vetor de suporte de amostra negativa
d	- Constante real qualquer
Δx	- Faixa de valores da amplitude de um sinal, em mV
SW	- Teste de Shapiro-Wilk
\mathbf{m}'	- Vetor de valores esperados na estatística de ordem padrão normal
\mathbf{V}	- Matriz de covariância das estatísticas de ordem
\mathbf{Y}'	- Vetor ordenado de observações amostrais
\mathbf{a}'	- Vetor carregamento
H_0	- Hipótese nula
E_{total}	- Erro ou porcentagem de acerto total para validação cruzada
H_i	- Número de harmônicas no sinal de entrada
H_s	- Número de harmônicas no banco artificial
p_r	- Porcentagem de ruído
$r(t)$	- Ruído adicionado ao sinal

Letras Gregas

α	-	Constante qualquer $\in \mathbb{R}$
π	-	Constante irracional
σ	-	Desvio padrão
μ	-	Valor médio
μ_1	-	Média
μ_2	-	Média quadrática
σ^2	-	Variância
σ^3	-	Assimetria
σ^4	-	Curtose
β	-	Constante real qualquer
ρ_{xy}	-	Coefficiente de correlação (Pearson)
ρ_{max}	-	Coefficiente máximo de correlação
$\tilde{\rho}$	-	Coefficiente de correlação para ajuste de curvas
ω_n	-	Frequência das senoides do banco real
ω_m	-	Frequência das senoides do banco artificial
ϕ	-	Fase das senoides do banco real
φ	-	Fase das senoides do banco artificial
λ	-	Multiplicadores de Lagrange
$\tilde{\mu}$	-	Valor médio após projeção (LDA ou QDA)
Σ	-	Matriz de covariância (LDA ou QDA)
π_k	-	Probabilidade a priori de uma variável k qualquer
δ	-	Equação para limite de fronteira entre amostras
$\phi(*)$	-	Função <i>kernel</i>
γ	-	Constante qualquer $\in \mathbb{R}$
ζ	-	Coefficientes obtidos para as curvas LDA e QDA

Acrônimos

ICM	-	Interface Cérebro-Máquina
ELA	-	Esclerose Lateral Amiotrófica
AVC	-	Acidente Vascular Cerebral
BCI	-	<i>Brain-Computer Interface</i>
EEG	-	Eletroencefalograma
PVEEE	-	Potenciais Visuais Evocados de Estado Estacionário
SSVP	-	<i>Steady-State Visual Evoked Potentials</i>
CCA	-	<i>Canonical Correlation Analysis</i>
PSD	-	<i>Power Spectral Density</i>
FFT	-	<i>Fast Fourier Transform</i>
PCA	-	<i>Principal Component Analysis</i>
KKT	-	Condições de Karush-Kuhn-Tucker
LDA	-	<i>Linear Discriminant Analysis</i>
QDA	-	<i>Quadratic Discriminant Analysis</i>
SVM	-	<i>Support Vector Machine</i>
1VR	-	Abordagem um versus o restante
1V1	-	Abordagem um versus um
RMS	-	<i>Root Mean Square</i>

Lista de Figuras

1	Áreas corticais das superfícies laterais e médias dos hemisférios cerebrais. . .	28
2	Macrorregiões cerebrais.	28
3	Forma esquemática da anatomia de um neurônio.	30
4	Interconexão de neurônios, gravação de pulso e sinapse.	31
5	Ilustração do limiar de disparo.	31
6	Gradiente de cátions e ânions em um axônio para despolarização da membrana.	32
7	Frequência de potenciais de ação e visualização de seu efeito no EEG. . . .	33
8	Faixas de frequência características da dinâmica neural.	35
9	Aquisição de sinal em interface cérebro-máquina.	36
10	Posicionamento de eletrodos com o sistema 10-20.	43
11	Posicionamento de eletrodos de acordo com as macrorregiões do córtex cerebral e o sistema 10-10.	43
12	Protocolo SSVEP.	44
13	Exemplo de estímulos para o protocolo SSVEP.	45
14	Paciente com olhos abertos sem estímulo visual específico.	46
15	Paciente sob estímulo visual modulado em 5,75 Hz.	46
16	Paciente sob estímulo visual modulado em 7,75 Hz.	47
17	Paciente sob estímulo visual modulado em 8,75 Hz.	47
18	Paciente sob estímulo visual modulado em 9,75 Hz.	47
19	Exemplo de sinais determinístico e não determinístico.	54
20	Senoide com alta amplitude e sua respectiva PDF.	56
21	Senoide com baixa amplitude e sua respectiva PDF.	56

22	Sinal determinístico para obtenção da PDF discreta.	57
23	FDP e FDC uniformes.	58
24	FDP e FDC exponenciais.	59
25	FDP e FDC gaussianas.	60
26	Influência da variância, assimetria e curtose em uma PDF.	61
27	Distribuições marginais e conjunta.	64
28	Representação de probabilidade condicional.	65
29	Ilustração de correlação, anticorrelação, e ausência de correlação linear. . .	68
30	Exemplos de valores de correlação.	70
31	Reconhecimento de SSVEP baseado em CCA.	73
32	Dinâmica neural real e assunção adotada.	74
33	Representação das componentes principais.	80
34	Redução de dimensionalidade e projeção de dados SSVEP em 3 componen- tes principais.	81
35	Redução de dimensionalidade e projeção de dados a partir do discriminante linear de Fisher.	83
36	Classificação a partir do discriminante linear de Fisher.	86
37	Comparação entre LDA e QDA para 9,75 Hz e 7,75 Hz.	88
38	Ilustração do conceito de SVM.	89
39	Introdução de amostra cuja classe é desconhecida.	89
40	Definição da região entre margens para SVM.	90
41	Transformação para novo espaço aplicando uma função <i>kernel</i> em SVM. .	92
42	Teste visual de aderência à normalidade para a primeira série de dados. . .	94
43	Teste visual de aderência à normalidade para a segunda série de dados. . .	95
44	Ilustração do processo de validação cruzada por <i>k-fold</i>	97
45	Exemplo de uma matriz de confusão.	98
46	Ilustração de uma curva ROC para valores discretos.	99

47	Exemplo de distribuições de dados facilmente separáveis.	100
48	Exemplo de distribuições de dados mais próximas das reais.	100
49	Exemplo de distribuições de dados mais próximas das reais.	101
50	Estímulos visuais e posicionamento utilizados no experimento.	103
51	Ilustração geométrica do banco de dados utilizado para os testes com EEG.	103
52	Teste de aderência à normalidade para sinal de frequência de estímulo igual a 9,75 Hz.	104
53	Teste de aderência à normalidade para sinal de frequência de estímulo igual a 8,75 Hz.	104
54	Teste de aderência à normalidade para sinal de frequência de estímulo igual a 7,75 Hz.	105
55	Teste de aderência à normalidade para sinal de frequência de estímulo igual a 5,75 Hz.	105
56	Contribuição dos termos da equação generalizada CCA ($H_s = 3$).	106
57	Contribuição dos termos da equação generalizada CCA ($H_i = 3$).	107
58	Efeito de aumento do número de períodos no valor de correlação.	108
59	Efeito de adição de harmônicas.	110
60	Sub-harmônicas de cada harmônica adicionada no sinal de entrada.	110
61	Efeito de alterações em F	111
62	Relação entre F e f	112
63	Efeito da adição de ruído.	113
64	Comparação entre variável canônica (sinal real) e resultado analítico para $f_m = 9,75$ Hz.	114
65	Comparação de amplitudes em frequência entre sinal de um eletrodo e variável canônica para $f_m = 9,75$ Hz.	115
66	Pesos de cada eletrodo na classificação da frequência-alvo de 9,75 Hz.	115
67	Comparação entre variável canônica (sinal real) e resultado analítico para $f_m = 9,75$ Hz.	116

68	Componentes em frequência e correlação para 8,75 Hz, 7,75 Hz e 5,75Hz.	117
69	Comparação entre curvas analíticas e de sinal real para 9,75 Hz.	118
70	Fluxograma de similaridades e diferenças entre a abordagem clássica da CCA e a proposta de generalização.	118
71	Matriz de confusão utilizando CCA.	120
72	Influência indevida de frequências distintas na frequência de estímulo. . . .	121
73	Valores de correlação por grupos de frequência.	121
74	Curvas ROC para as frequências consideradas.	122
75	Comparação entre curvas analíticas e de sinal real para 9,75 Hz.	124
76	Comparação entre curvas analíticas e de sinal real para 9,75 Hz.	125
77	Dados utilizados de uma frequência de estímulo para validação cruzada em PCA.	126
78	Estimativas de função densidade de probabilidade para as componentes principais consideradas.	128
79	Teste visual e de Shapiro-Wilk para os dados de 9,75 Hz nas componentes principais.	128
80	Teste visual e de Shapiro-Wilk para os dados de 8,75 Hz nas componentes principais.	129
81	Teste visual e de Shapiro-Wilk para os dados de 7,75 Hz nas componentes principais.	129
82	Teste visual e de Shapiro-Wilk para os dados de 5,75 Hz nas componentes principais.	129
83	Estimativas de função densidade de probabilidade para as componentes principais após aplicação da LDA.	130
84	Subespaço de duas dimensões com as componentes principais após LDA. . .	131
85	Aplicação de LDA e QDA para 9,75 Hz e 8,75 Hz.	132
86	Aplicação de LDA e QDA para 9,75 Hz e 7,75 Hz.	133
87	Aplicação de LDA e QDA para 9,75 Hz e 5,75 Hz.	133
88	Aplicação de LDA e QDA para 8,75 Hz e 7,75 Hz.	133

89	Aplicação de LDA e QDA para 8,75 Hz e 5,75 Hz.	134
90	Aplicação de LDA e QDA para 7,75 Hz e 5,75 Hz.	134

Sumário

1	INTRODUÇÃO	22
1.1	OBJETIVOS DO TRABALHO	25
1.2	ORGANIZAÇÃO DO TRABALHO	26
2	ASPECTOS FUNDAMENTAIS DA ANATOMIA E DINÂMICA CEREBRAL	27
2.1	ANATOMIA CEREBRAL	27
2.2	DINÂMICA NEURAL	30
2.3	SISTEMAS DE INTERFACE CÉREBRO-MÁQUINA	35
2.3.1	Técnicas não Invasivas	36
2.3.2	Técnicas Invasivas	40
2.4	SISTEMA 10-20, 10-10 E 5-10	42
2.5	POTENCIAIS VISUAIS EVOCADOS EM ESTADO ESTACIONÁRIO . . .	44
2.5.1	Adequabilidade do Protocolo SSVEP	48
2.6	ESTADO DA ARTE	49
3	CONCEITOS BÁSICOS DE PROBABILIDADE E ESTATÍSTICA	53
3.1	ESPAÇO AMOSTRAL E AXIOMAS DE PROBABILIDADE	54
3.2	FUNÇÕES DE DENSIDADE E DISTRIBUIÇÃO DE PROBABILIDADE .	55
3.3	MOMENTOS DE VARIÁVEIS ALEATÓRIAS	60
3.4	PROBABILIDADE CONJUNTA	62
3.5	PROBABILIDADE CONDICIONAL	65
4	CLASSIFICAÇÃO ESTATÍSTICA	67

4.1	ANÁLISE DE CORRELAÇÃO CANÔNICA (CCA)	67
4.1.1	Análise de Correlação Canônica Clássica	71
4.1.2	Generalização Proposta para a Técnica CCA	73
4.2	OUTRAS METODOLOGIAS DE CLASSIFICAÇÃO	78
4.2.1	Análise de Componentes Principais (PCA)	78
4.2.2	Análise de Discriminante Linear/Quadrático (LDA/QDA)	83
4.2.3	Máquina de Vetores de Suporte (SVM)	88
4.3	CRITÉRIOS DE ADERÊNCIA À NORMALIDADE	93
4.3.1	Crítério Visual	94
4.3.2	Crítério Formal	95
4.4	VALIDAÇÃO CRUZADA	96
4.5	CURVA ROC	97
5	RESULTADOS E DISCUSSÃO	102
5.1	TESTE DE ADERÊNCIA À NORMALIDADE	103
5.2	ANÁLISE DE CORRELAÇÃO CANÔNICA (CCA)	105
5.2.1	Soluções Analíticas	106
5.2.2	Sinais de EEG	113
5.2.3	Análise de Porcentagem de Acerto	119
5.3	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	123
5.3.1	Visualização das Componentes Principais	123
5.3.2	Validação Cruzada e Porcentagem de Acerto	125
5.4	ANÁLISE DE DISCRIMINANTE LINEAR/QUADRÁTICO (LDA/QDA) .	127
5.4.1	Visualização Espacial do Efeito da LDA	127
5.4.2	Discriminante Linear e Quadrático	131
5.4.3	Validação Cruzada e Porcentagem de Acerto	134
5.5	MÁQUINA DE VETORES DE SUPORTE (SVM)	135

5.6	COMPARAÇÃO ENTRE CLASSIFICADORES	136
6	CONSIDERAÇÕES FINAIS	141
6.1	CONTRIBUIÇÕES ALCANÇADAS	142
6.2	LIMITAÇÕES DO ESTUDO	142
6.3	SUGESTÕES PARA TRABALHOS FUTUROS	143
	REFERÊNCIAS	145

1 INTRODUÇÃO

A busca por uma melhor compreensão das regiões do cérebro e suas funções nas ações humanas tem sido uma tarefa árdua, porém muito útil principalmente para aplicações da engenharia de interface cérebro-máquina, bem como para o auxílio a diagnósticos médicos a partir de sinais obtidos dos pacientes em avaliação. Trata-se de uma área em profunda expansão, cujas fronteiras do imaginável estão sendo ultrapassadas a cada nova publicação de resultados de pesquisas em neurociência. No contexto do presente trabalho, destacam-se os trabalhos de *interface cérebro-máquina* (ICM) pela abrangência no envolvimento de técnicas, métodos e ferramentas comumente estudadas nos cursos de engenharia.

Os sistemas de ICM têm sido desenvolvidos principalmente com o intuito de promover capacidades de comunicação e controle para pessoas com incapacidades motoras (LAFLEUR et al., 2013). O progresso desse tipo de sistema depende, entre outras coisas, de avaliações e comparações sistemáticas de sinais cerebrais, métodos de aquisição de dados, algoritmos de processamento das informações e protocolos de operação, conforme mostrado nas referências Goldberger et al. (2000), Schalk et al. (2004), entre outras.

De acordo com Schalk et al. (2004), muitas pessoas com restrições motoras severas carecem de tecnologia para aumento de comunicação com o corpo. O acidente vascular cerebral (AVC), por exemplo, é uma das maiores causas de geração de incapacidades motoras em adultos no mundo (HATEM et al., 2016), ocupando uma posição especial entre os diferentes tipos de danos cerebrais. Para aquelas pessoas que sobrevivem a um AVC, mais de 80% apresentam hemiparesis do membro superior contralateral de forma aguda, e mais de 40% de forma crônica (Cramer et al. (1997), apud Hatem et al. (2016)). A esclerose lateral amiotrófica (ELA), por sua vez, é uma doença neural motora progressiva que leva à completa destruição do sistema nervoso periférico, afetando de forma relativamente menor, porém, as funções sensoriais e cognitivas Chou e Norris (1993), apud Chaudhary, Birbaumer e Ramos-Murguialday (2016)). Não há cura ainda, e os últimos vestígios de controle muscular normalmente são os movimentos dos olhos. Quando há a paralisia completa, exceto pelo movimento vertical e piscar dos olhos, com preservação de consciência e cognição, o estado do paciente é denominado *locked-in*; em casos ainda mais extremos, quando não há movimento qualquer dos olhos, atribui-se o nome de estado *completely locked-in* (Bauer, Gerstenbrand e Rumpl (1979), apud Chaudhary, Birbaumer

e Ramos-Murguialday (2016)).

Devido à heterogeneidade dos sintomas vistos após os AVCs, por exemplo, um grande arsenal de técnicas dedicadas à reabilitação da saúde das pessoas tem sido desenvolvido. A maioria delas são baseadas em teorias de aprendizado motor que, em geral, assumem que: (a) o reaprendizado motor é comparável ao aprendizado motor e (b) os pacientes podem, de fato, aprender (HUANG; KRAKAUER, 2009). A lei fundamental de aprendizado motor é o “treino”, aquisição abrangente de capacidades, adaptação motora e tomada de decisão. Fatores essenciais para a prática bem sucedida são o número de repetições, alta intensidade, priorização sensorial, prática variável e, por último, mas não menos importante, provisão de *feedback*. O *feedback* facilita a avaliação detalhada do desempenho: ele enfatiza o aspecto sensorial no *loop* senso-motor. Ao salientar essas características ao paciente, aumenta-se seu comprometimento ativo, bem como sua motivação. Esta última é especialmente importante, de modo que nenhuma terapia é efetiva quando há falta de motivação para a prática (LEVIN; SVEISTRUP; SUBRAMANIAN, 2010).

Segundo Dokkum, Ward e Laffont (2015), a maioria das ferramentas de reabilitação comum requer um nível residual de controle motor para realizar as tarefas terapêuticas requeridas a fim de ter algo pelo qual oferecer *feedback*. Para pacientes com severas incapacidades motoras (pouco ou nenhum controle de movimento), as ferramentas existentes podem não ser suficientes. Para as pessoas consideradas *locked-in*, os casos mais delicados, não se pode usar tecnologias convencionais de reabilitação, pois geralmente requerem algum tipo de medida de controle muscular. Devido a isso, ao longo das duas últimas décadas, diversos estudos avaliaram a possibilidade de que sinais cerebrais gravados a partir de eletrodos ou escalpos de eletrodos¹ (técnica não invasiva), ou de dentro do próprio cérebro (técnica invasiva), poderiam prover uma nova tecnologia de aumento de comunicação que não necessite de controle muscular.

Neste ponto, as interfaces de cérebro-máquina (*Brain-Computer Interface* - BCI) prometem preencher esta lacuna. Os sistemas BCI gravam e decodificam atividade cerebral enquanto realizam ou tentam realizar tarefas motoras e/ou cognitivas. Esses sistemas podem, portanto, ser configurados de tal forma que mapeiem os sinais cerebrais decodificados em *feedback* útil na tarefa realizada para ambos paciente e terapeuta. Esse *feedback* pode tomar várias formas, inclusive de natureza visual, auditiva ou tátil. O sinal decodificado pode ser usado mesmo para controle de dispositivos externos que executem os movimentos pretendidos, provendo *feedback* proprioceptivo. Consequentemente, pode-se

¹Neste caso, escalpos são dispositivos fixados à pele do crânio com o intuito de se realizar o registro eletroencefalográfico a partir de eletrodos distribuídos ao longo das regiões cerebrais.

considerar os sistemas BCI como uma forma de tecnologia de reabilitação que aumenta as possibilidades terapêuticas a todos os pacientes, independentemente da severidade da disfunção motora. Quando a função motora normal é perdida, os sistemas BCI podem promover o recrutamento de áreas cerebrais envolvidas em uma tarefa particular, induzindo a plasticidade neural — isto é, habilidade do sistema nervoso de reorganizar suas estruturas, funções e conexões em resposta a estímulos intrínsecos ou extrínsecos (RUGE; LIOU; HOAD, 2012) — requerida para recuperação da função.

Em particular, análises envolvendo técnicas de processamento de sinais de eletroencefalograma (EEG) têm se mostrado de significativa importância para o desenvolvimento dessa área. A partir delas, observa-se as principais características peculiares a esses sinais: há, em maior ou menor grau (dependendo da atividade na qual o cérebro está engajado, bem como da pessoa que realiza determinada tarefa), tanto um comportamento parcialmente não determinístico quanto componentes de frequência, ou faixas de frequência, bem estabelecidas, denotadas por oscilações periódicas de pulsos ao longo do tempo, representativos das descargas elétricas oriundas das sinapses dos neurônios (BROGIN, 2016). A correta aquisição e classificação dos EEG, principalmente no que concerne às frequências desses disparos elétricos, a uma melhor compreensão da dinâmica cerebral e, também, à possibilidade de processamento da informação contida nos sinais para eventual integração de pacientes com capacidades motoras comprometidas com atuadores que os auxiliem em sua reabilitação.

Uma abordagem amplamente utilizada para se efetuar a integração é a interface cérebro-máquina com auxílio de Potenciais Visuais Evocados (*Visual Evoked Potentials*), consistindo em potenciais com pequena latência que representam a resposta do cérebro a estímulos visuais exógenos rápidos. A partir dessa abordagem, é possível fazer uso de Potenciais Visuais Evocados de Estado Estacionário (PVEEE) (ou do inglês *Steady-State Visual Evoked Potentials*, SSVEP), que, de forma geral, são sinais caracterizados pela resposta evocada do cérebro a estímulos visuais em frequências específicas (BEVERINA et al., 2003).

Quando a retina é excitada por um estímulo visual abrangendo a faixa de 3,5 a 75 Hz, o cérebro gera uma atividade cerebral na mesma frequência desse estímulo, ou em seus harmônicos, isto é, frequências múltiplas da fundamental (BEVERINA et al., 2003). Assim, empregando-se um dispositivo que permita ao usuário do sistema de interface cérebro-máquina fixar seu olhar em um sinal visual emitido em determinadas frequências, é possível detectá-las com o auxílio de técnicas físico-matemáticas de aquisição e processamento de sinais.

Neste contexto apresentado, este trabalho tem o intuito de investigar a aplicação da técnica de correlação canônica (CCA, do inglês *Canonical Correlation Analysis*) para reconhecimento de frequência apresentada pelos EEG de indivíduos sob o protocolo SS-VEP, com o objetivo de possibilitar o desenvolvimento de um sistema de interface cérebro-máquina que permita reconhecer uma intenção de movimento de um usuário com restrição motora. Em particular, considera-se sinais de EEG reais de indivíduos humanos, principalmente da região occipital, obtidos a partir de um banco de dados de domínio público. Os dados são relativos a um experimento utilizando 4 frequências de estímulo (5,75 Hz, 7,75 Hz, 8,75 Hz e 9,75 Hz) apresentadas visualmente e de forma simultânea na tela de um computador (ZHANG et al., 2014). No total, 20 rodadas por frequência foram feitas, a aquisição dos sinais se baseou em uma janela temporal de 4 segundos, a partir de 30 eletrodos (dos quais apenas 8, representativos da região occipital, estão disponíveis no banco de dados). Mais detalhes do experimento constam na seção de resultados.

Também, o trabalho compreende uma formulação estendida do método CCA que generaliza sua aplicação, eliminando atuais limitações mencionadas na literatura.

1.1 OBJETIVOS DO TRABALHO

O objetivo principal deste trabalho é avaliar técnicas de correlação de sinais para a análise de eletroencefalogramas e reconhecimento da intenção do usuário integrado ao sistema de interface cérebro-máquina. Como objetivos secundários, tem-se:

- aplicar um algoritmo de CCA (Análise de Correlação Canônica) em sinais de eletroencefalogramas, para identificar frequências presentes nos sinais EEG e classificá-las corretamente de acordo com os potenciais visuais evocados;
- avaliar métricas de correlação baseadas em CCA buscando aperfeiçoar o processamento dos sinais EEG;
- aplicar técnicas alternativas de classificação de parâmetros, a fim de se obter comparações em termos de desempenho e aplicabilidade;
- avaliar, a partir do reconhecimento da intenção do usuário, estratégias que possibilitem a integração de atuadores mecânicos que o auxiliem na retomada de controle motor e autonomia.

1.2 ORGANIZAÇÃO DO TRABALHO

A seguir, apresenta-se, de forma breve, a estrutura de organização deste texto:

- **Capítulo 1:** apresenta uma introdução contendo o contexto a motivação deste trabalho, bem como seus objetivos, contribuições alcançadas e organização;
- **Capítulo 2:** introduz conceitos fundamentais para uma compreensão geral sobre a anatomia básica do cérebro humano, bem como a dinâmica na qual os neurônios operam;
- **Capítulo 3:** apresenta um estudo preliminar sobre os principais conceitos de probabilidade e estatística, que são base para as técnicas de reconhecimento de padrão aplicadas neste trabalho;
- **Capítulo 4:** compreende as técnicas de classificação estatística utilizadas no presente trabalho, com destaque para a Análise de Correlação Canônica, primeira seção, onde uma generalização de seu modelo é desenvolvida. Também, apresenta-se técnicas de tratamento e validação de dados;
- **Capítulo 5:** apresenta as contribuições obtidas até o momento no que diz respeito à maior compreensão e generalização da CCA, bem como resultados relativos a sua acurácia no reconhecimento de padrões e comparações com os desempenhos de outras técnicas de classificação;
- **Capítulo 6:** destaca as conclusões obtidas com o estudo, bem como sugestões para trabalhos futuros.

2 ASPECTOS FUNDAMENTAIS DA ANATOMIA E DINÂMICA CEREBRAL

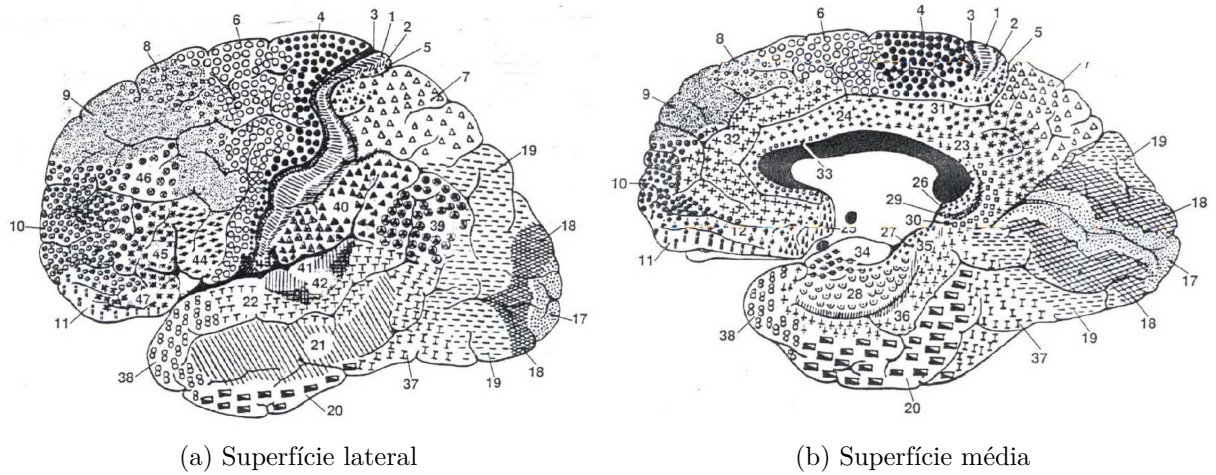
Este capítulo aborda conceitos gerais sobre a anatomia cerebral humana, bem como aspectos fundamentais sobre a dinâmica neural, considerados relevantes no contexto deste trabalho.

2.1 ANATOMIA CEREBRAL

Para os estudos envolvendo interface cérebro-máquina, é necessário, a princípio, ter-se um conhecimento básico da anatomia cerebral. A primeira e mais importante tarefa do mapeamento das regiões cerebrais é o de sua divisão de acordo com características anatômicas, isto é, o agrupamento sistemático de componentes neurais estruturalmente similares e a separação daqueles estruturalmente dissimilares, podendo, assim, identificar partes homólogas do córtex cerebral em diferentes animais, ou grupos de animais, baseados em sua estrutura (GAREY, 2006). Para tal, o recurso utilizado neste trabalho é o denominado mapa citoarquitetônico de Brodmann, estabelecido em 1909 pelo neurologista alemão Korbinian Brodmann, que define 52 regiões diferentes do córtex cerebral a partir de suas características histológicas, conforme mostrado na Figura 1.

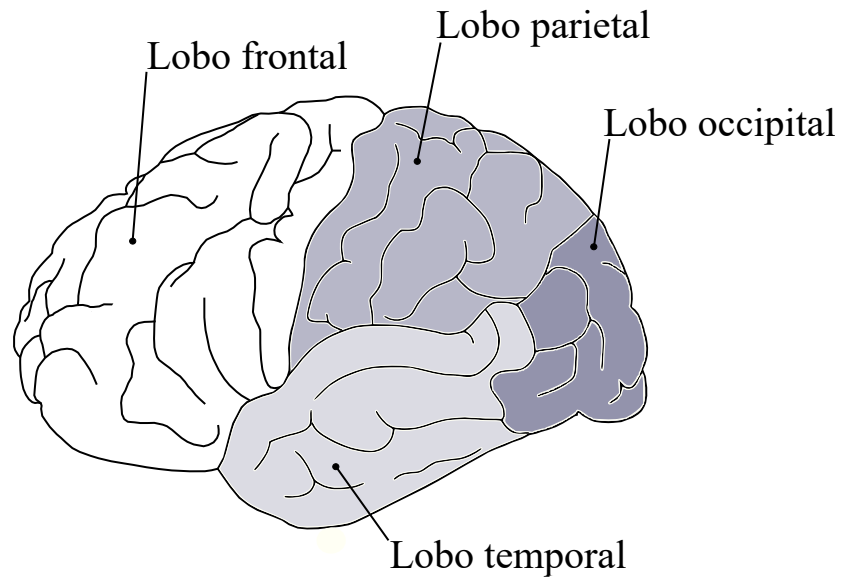
A Figura 2, por sua vez, apresenta as macrorregiões representativas dessas estruturas (MACHADO, 1993). Como exemplo de algumas das principais áreas funcionais e suas devidas localizações, a Tabela 1 apresenta os nomes de 27 das regiões mostrada no mapa de Brodmann, bem como suas respectivas funções ou responsabilidades nas ações humanas. Note em destaque as áreas 17, 18 e 19 do córtex visual, lobo occipital, que são as predominantemente consideradas neste trabalho.

Figura 1 – Áreas corticais das superfícies laterais e médias dos hemisférios cerebrais.



Fonte: Adaptado de Singi (1996).

Figura 2 – Macrorregiões cerebrais.



Fonte: Elaborado pelo autor.

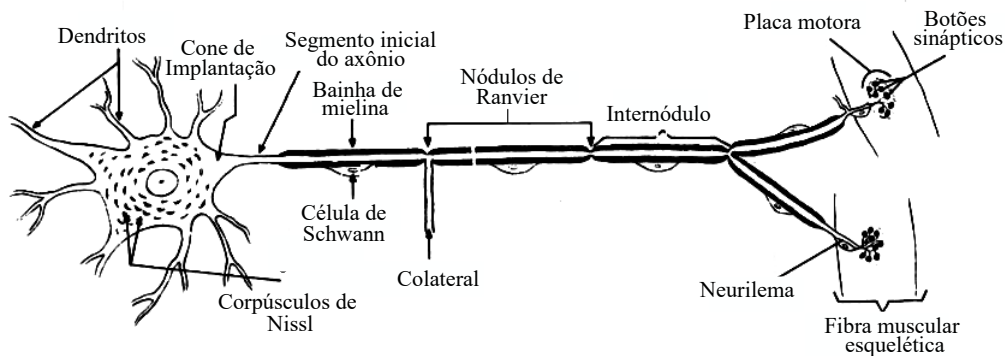
Tabela 1 – Áreas corticais das superfícies lateral e medial do hemisfério cerebral humano

Áreas	Áreas Funcionais	Localização	Função
1, 2, 3	Córtex sensitivo primário	Giro pós-central	Tato
4	Córtex motor primário	Giro pré-central	Controle do movimento voluntário
5	Córtex sensitivo somático terciário; área associativa parietal posterior	Lóbulo parietal superior	Estereognosia
6	Córtex motor suplementar; campo ocular suplementar; córtex pré-motor; campos oculares frontais	Giro pré-central e córtex adjacente rostral	Planejamento dos movimentos dos membros e oculares
7	Área associativa parietal posterior	Lóbulo parietal superior	Vísuo-motora e percepção
8	Campos oculares frontais	Giros frontal superior, médio, lobo frontal medial	Movimentos sacádicos oculares
9, 10, 11, 12	Córtex associativo pré-frontal; campos oculares frontais	Giros frontal superior, médio, lobo frontal medial	Pensamento, cognição, planejamento do movimento
17	Córtex visual primário	Margens da cisura calcarina	Visão
18	Córtex visual secundário	Giros occipitais medial e lateral	Visão, profundidade
19	Córtex visual terciário, área visual temporal média	Giros occipitais medial e lateral	Visão, cor, movimento e profundidade
20	Área temporal inferior visual	Giro temporal inferior	Forma visual, memória
21	Área temporal inferior visual	Giro temporal médio	Forma visual, memória
22	Córtex auditivo de ordem superior	Giro temporal superior	Audição, palavra, memória auditiva e interpretativa
23, 24, 25, 26, 27	Córtex associativo límbico	Giro cíngulo, área subcalosa, área retroesplenial e giro para-hipocampal	Emoções

2.2 DINÂMICA NEURAL

Segundo Izhikevich (2007), o funcionamento dos neurônios não deve se restringir a uma abordagem estritamente biológica ou teórica, mas, também, sob o ponto de vista de um sistema dinâmico não-linear, que trata dos sinais de entrada baseado em sua própria dinâmica intrínseca. Desse modo, pode-se compreender os neurônios como excitáveis, visto que trabalham próximo a uma transição, chamada bifurcação, desde seu estado inerte inicial até o estado contínuo de atividade de pulsação (*spiking*). A Figura 3 apresenta, de forma esquemática, a anatomia de um neurônio. Nota-se que ele pode ser caracterizado, essencialmente, por três partes principais: o corpo celular (de onde partem os dendritos), os dendritos (ramificações que se ligam a outros neurônios) e axônio (maior porção do neurônio, partindo da extremidade esquerda, que se liga ao corpo celular, até a extremidade direita, onde há a bifurcação para conexão com outros neurônios).

Figura 3 – Forma esquemática da anatomia de um neurônio.

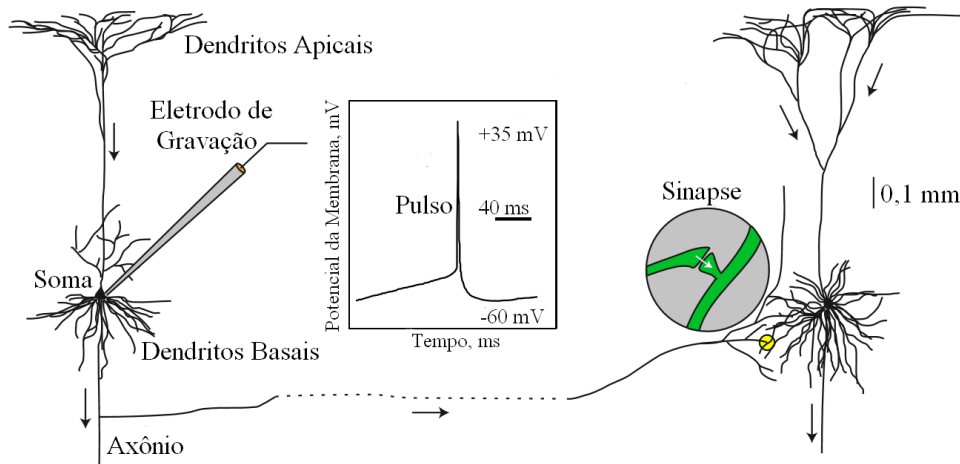


Fonte: Adaptado de Machado (1993).

Ainda, de acordo com Izhikevich (2007), um neurônio típico recebe entradas de mais de 1000 outros neurônios por meio do contato de seus dendritos, chamadas sinapses. As entradas produzem correntes elétricas transmembranas que mudam o potencial elétrico da membrana de um neurônio, chamados de potenciais pós-sinápticos (PSPs); quanto maior a corrente, mais significativo é o potencial. Um grande potencial pode ser amplificado pelos canais sensíveis à tensão elétrica integrados na membrana neural e gerar, de fato, um potencial, ou pulso, caracterizado pela descarga elétrica abrupta e transiente da tensão elétrica da membrana, que se propaga a outros neurônios pelo axônio. Assim, os pulsos caracterizados podem ser considerados como os meios de comunicação entre os neurônios,

bem como estes, por sua vez, não disparam de forma autônoma, eles o fazem devido ao acúmulo resultante de pulsos de outros neurônios. A Figura 4 ilustra dois neurônios interconectados, um pulso gravado *in vitro*, e a sinapse.

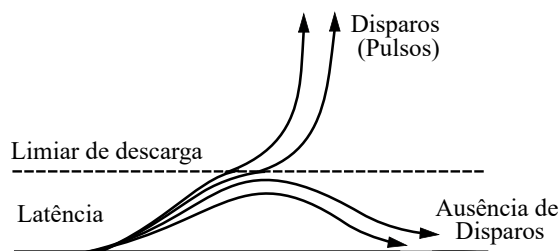
Figura 4 – Interconexão de neurônios, gravação de pulso e sinapse.



Fonte: Adaptado de Izhikevich (2007).

De forma geral, pode-se definir como limiar de descarga (*firing treshold*) o patamar de bifurcação do neurônio; isto é, o limiar a partir do qual seu estado muda e este, por fim, produz um pulso. Esses pulsos, por sua vez, comportam-se da forma “tudo ou nada”: há um disparo se um determinado valor de carga for atingido, não importando se este for igual ou maior que o limiar; caso contrário, os pulsos não ocorrem, como mostra a Figura 5.

Figura 5 – Ilustração do limiar de disparo.

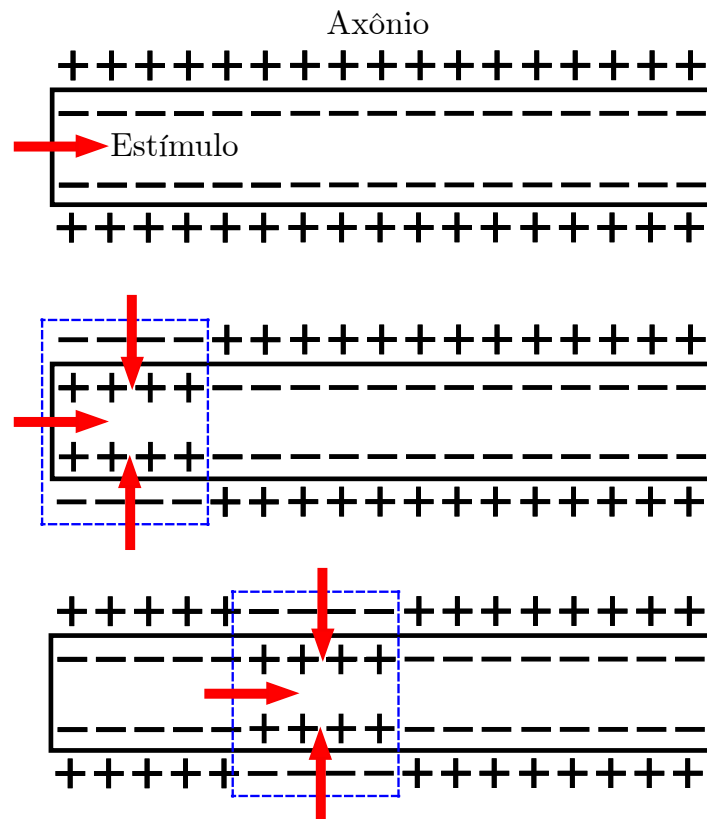


Fonte: Adaptado de Izhikevich (2007).

Esse processo de descarga é melhor ilustrado na Figura 6, e pode ser definido de acordo com Niedermeyer e Silva (2005): ao se gravar, por exemplo, a atividade de um neurô-

nio, sabe-se que este apresenta conexões sinápticas com outros neurônios; as estruturas pré-sinápticas são, também, exploradas pela gravação e, se um potencial de ação percorre ao longo das fibras dos neurônios (projeções, ou prolongamentos, axonais), um potencial sináptico ocorre no neurônio seguinte. Se mais de um potencial percorre essa mesma fibra em um pequeno intervalo de tempo, há uma soma do potencial sináptico, causando um potencial de ação no neurônio pós-sináptico após atingir o limiar de descarga, o que caracteriza uma despolarização da membrana neural. A permeabilidade dessa membrana a um dado íon é controlada pelas condições elétricas e químicas do ambiente no qual esta célula está inserida. Com a despolarização, um gradiente de potencial se desenvolve ao longo da membrana, tanto na porção interna quanto externa do neurônio. Devido a esse gradiente, cátions se movem através da porção extracelular em direção à região subsináptica, e um fluxo dirigido inversamente ocorre na porção intracelular. Esse gradiente que ocorre através e ao longo da membrana causa um efeito de “estrangulamento”, fazendo com que a descarga percorra toda a projeção axonal.

Figura 6 – Gradiente de cátions e ânions em um axônio para despolarização da membrana.

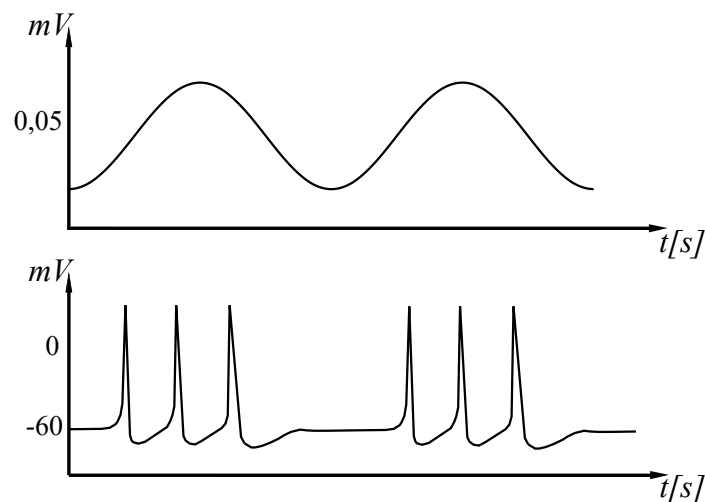


Fonte: Elaborado pelo autor.

Dependendo da atividade neural na qual o neurônio está engajado, sua frequência de pulsação varia (potencial de ação), como apresentado na Figura 7. A porção de corrente

elétrica que flui através do espaço de fora da célula neural, isto é, espaço extracelular, é responsável, por sua vez, pelo que se conhece como potenciais de campo. Quanto mais de um potencial percorre o mesmo neurônio, esses potenciais são somados; ou seja, de acordo com a sincronização e frequência de disparos, esses potenciais podem ser amplificados. O efeito que se visualiza no eletroencefalograma inclui o conjunto do efeito de potenciais de ação e de campo resultantes de vários neurônios em uma região cerebral de interesse, sendo, portanto, um efeito global de todo o somatório de potenciais individuais, apresentando menor amplitude que a dos potenciais de ação individuais, e forma próxima a de funções senoidais (Fig. 7).

Figura 7 – Frequência de potenciais de ação e visualização de seu efeito no EEG.



Fonte: Elaborado pelo autor.

Em relação aos pulsos dos neurônios, segundo Niedermeyer e Silva (2005) apud Siuly (2012), a frequência com que ocorrem é um dos parâmetros mais importantes para se avaliar anormalidades em eletroencefalogramas clínicos, bem como para entender comportamentos funcionais em pesquisa cognitiva. Contando com bilhões de comunidades oscilantes de neurônios como fonte, os potenciais desses EEG são manifestados como oscilações aperiódicas imprevisíveis com disparos de oscilações que são tipicamente categorizados em bandas, ou faixas, sucintamente descritas a seguir.

- Faixa δ : as ondas do tipo δ estão na faixa entre 0,5 a 4 Hz, tem predominância no lobo frontal e sua forma é caracterizada como a de maior amplitude e de menor frequência, sendo predominante em crianças, característica de estados de indivíduos durante o sono (BRONZINO, 1999), ou mesmo em concentração interna, por exemplo, envolvendo tarefas como meditação, cálculo mental ou processos cognitivos

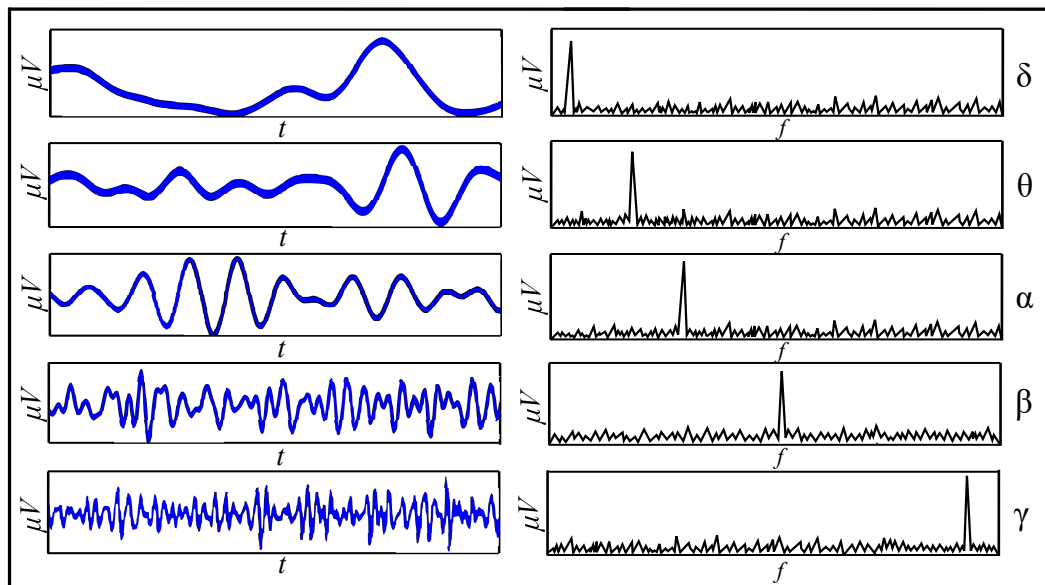
(HARMONY, 2013); quando associada a enfermidades, pode indicar estados de desordem neurológica (WORTZEL, 2008; FAUZAN; AMRAN, 2015), como desordens do espectro do autista;

- Faixa θ : as ondas do tipo θ estão na faixa de 4 a 8 Hz, com predominância no córtex pré-frontal, sensorial e mesmo hipocampo (UHLHAAS et al., 2008), estão associadas com estresse emocional, frustração ou desapontamento, e por vezes ao conteúdo inconsciente, inspiração criativa e meditação; também estão associadas a estados de alerta (BRONZINO, 1999), codificação ou recuperação da memória (HARMONY, 2013), ou plasticidade sináptica (UHLHAAS et al., 2008). Alterações ou excesso de atividade nessa faixa podem indicar desordens cerebrais (LARRAIN-VALENZUELA et al., 2017; FAUZAN; AMRAN, 2015), como desordens do espectro do autista;
- Faixa α : as ondas do tipo α estão na faixa entre 8 a 13 Hz, aparecendo principalmente nas regiões posteriores do cérebro (lobo occipital), quando o sujeito está com os olhos fechados (BRONZINO, 1999), ou acordado, em estado de relaxamento. Também, as ondas α são as predominantes no cérebro humano e há evidência de que tenham uma função inibitória e de seleção/supressão de atenção (GRAIMANN; ALLISON; PFURTSCHELLER, 2010; KLIMESCH, 2012). Alterações na faixa α , ou mesmo excesso em determinadas regiões do cérebro, podem indicar desordens neurológicas, como desordens do espectro do autista, problemas de atenção e ansiedade (FAUZAN; AMRAN, 2015);
- Faixa β : as ondas do tipo β estão na faixa entre 13 a 30 Hz. De amplitudes menores em relação às outras faixas, as ondas do tipo β ocorrem em todas as áreas corticais e várias estruturas subcorticais. Geralmente, são associadas ao estado cerebral desperto, estimulado, e ativamente engajado em atividades mentais, indicando uma variedade de tarefas cognitivas, envolvendo atenção e percepção; também, essa faixa está envolvida na coordenação em larga escala da atividade neural distribuída (UHLHAAS et al., 2008). Alterações ou atividade excessiva nessa faixa podem indicar desordens neurológicas, como transtornos alimentares, por exemplo (JÁUREGUI-LOBERA, 2012);
- Faixa γ : as ondas do tipo γ , presentes em todas as estruturas cerebrais (UHLHAAS et al., 2008), encontram-se acima de 30 Hz, podendo atingir um máximo em torno de 80 e 100 Hz, sendo associadas, normalmente, com várias funções motoras e cognitivas, como percepção, atenção, consciência, plasticidade sináptica e controle

motor (UHLHAAS et al., 2008). Alterações nessa faixa também indicam desordens neurológicas: o aumento pode estar associado a sintomas de alucinações ou epilepsia, e a diminuição, a sintomas de esquizofrenia e doença de Alzheimer (HERRMANN; DEMIRALP, 2005).

Na Figura 8, constam ilustrações qualitativas destes tipos de ondas, com destaque para se verificar, de forma relativa, suas faixas de frequência verificadas nos EEG.

Figura 8 – Faixas de frequência características da dinâmica neural.



Fonte: Adaptado de Siuly (2012).

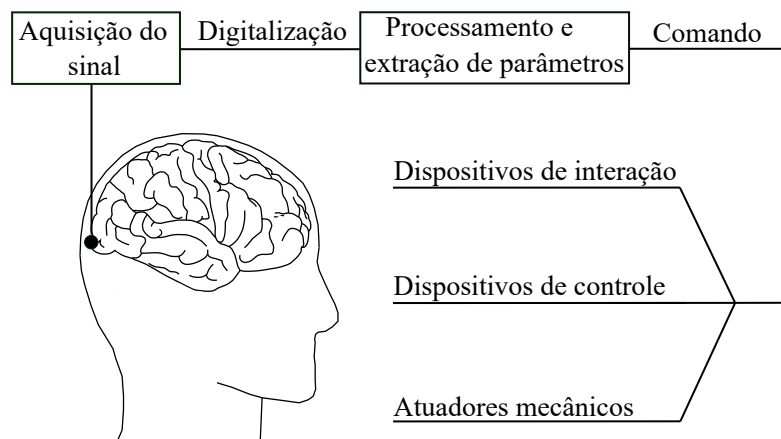
Uma vez abordados aspectos gerais sobre anatomia e dinâmica cerebral, é possível se introduzir o conceito de interface cérebro-máquina, conforme descreve a seção 2.3.

2.3 SISTEMAS DE INTERFACE CÉREBRO-MÁQUINA

Segundo Graimann, Allison e Pfurtscheller (2010), as formas naturais de comunicação ou controle começam a partir da intenção do indivíduo e requerem a atuação de nervos periféricos ou músculos: essa intenção inicia um processo complexo de ativação de determinadas áreas cerebrais (correspondentes à intenção envolvida) e transmissão de sinais aos músculos por meio do sistema nervoso periférico, que, por sua vez, são convertidos em movimento e controle, tarefas chamadas de resultado motor, ou resultado eferente. O processo contrário, ou aferente, descreve a comunicação a partir dos receptores sensoriais até o sistema nervoso central. Um sistema ICM é, essencialmente, um sistema artificial que contorna o processo natural de comunicação eferente.

Desta forma, sistemas de interface cérebro-máquina podem permitir que seres humanos operem dispositivos (como computadores, cadeiras de roda e próteses, por exemplo) sem o auxílio de músculos ou atividade neural periférica; isto é, eles decodificam a intenção do usuário apenas a partir de sinais neurais (KWAK; MÜLLER; LEE, 2017). Esses sistemas medem parâmetros específicos da atividade cerebral e os traduzem para dispositivos de controle de sinais, fato que demonstra sua adequabilidade ao propósito deste trabalho. A Figura 9 apresenta, de forma conceitual, como a atividade cerebral é medida em uma ICM: é feita a aquisição dos sinais cerebrais por eletrodos localizados no escalpo, na superfície cortical, ou de dentro do cérebro; estes são pós-processados para se extrair os parâmetros desejados (amplitudes a partir de córtex senso-motor ou frequência de pulsos de neurônios, por exemplo) que permitam identificar a intenção do usuário; os parâmetros são, então, traduzidos para comandos que operam um dispositivo (como exemplos: programas de simples processamento de palavras, cadeira de rodas, ou próteses neurais) (SCHALK et al., 2004).

Figura 9 – Aquisição de sinal em interface cérebro-máquina.



Fonte: Elaborado pelo autor.

Há, essencialmente, dois grupos de técnicas utilizadas nessas plataformas: invasivas e não invasivas, bem como tipos específicos de sinais que podem ser obtidos a partir dessas técnicas, apresentados sucintamente a seguir.

2.3.1 Técnicas não Invasivas

Interfaces cérebro-máquina não invasivas não requerem implantação cirúrgica de eletrodos e permitem gravações de sinais cerebrais conhecidos como eletroencefalogramas (EEG) (GRAIMANN; ALLISON; PFURTSCHELLER, 2010), como visualizado na Figura 7, a partir de eletrodos ou uma superfície externa de um escalpo de eletrodos. Natu-

ralmente, há vantagens e desvantagens no que tange à utilização dessa técnica.

Vantagens de Técnicas não Invasivas

Em relação às vantagens, pode-se citar (GRAIMANN; ALLISON; PFURTSCHELLER, 2010):

- é um método bem estabelecido, utilizado em aplicações clínicas ou no contexto de pesquisa há décadas;
- o equipamento usado para a aquisição do sinal é relativamente mais barato e fácil de aplicar que aquele usado em técnicas invasivas;
- a resolução temporal (ou capacidade de detectar mudanças dentro de um intervalo de tempo) é boa;
- o fato de não envolver cirurgia (e, conseqüentemente, questões éticas e financeiras mais críticas), o que torna sua aplicação mais simples e segura.

Desvantagens de Técnicas não Invasivas

As desvantagens, por sua vez, são (GRAIMANN; ALLISON; PFURTSCHELLER, 2010):

- a resolução espacial (referente ao posicionamento dos eletrodos) e faixa de frequências para aplicação são limitados;
- o EEG é suscetível aos chamados artefatos, isto é, contaminações no sinal causadas por outras atividades elétricas oriundas do próprio indivíduo (como movimento ou piscar de olhos e atividade muscular, por exemplo);
- fontes eletromagnéticas externas, como as linhas de corrente elétrica, também podem contaminar o EEG;
- embora o EEG seja de relativa fácil aplicação, o procedimento de uso do equipamento e configuração pode ser incômodo: os eletrodos a serem aplicados devem ser preparados cuidadosamente com gel condutor, que pode, inclusive, secar durante o processo e prejudicar a aquisição de sinais.

Devido às peculiaridades do EEG, potenciais, ou tipos de sinais, específicos podem ser obtidos a partir dessa técnica, o que determina, inclusive, o tipo de protocolo a ser empregado no sistema ICM. Nesse sentido, pode-se definir dois grupos: potenciais espontâneos (também categorizados como de imagética motora, *Motor Imagery*) e potenciais evocados (ou também chamados de potenciais relacionados a eventos, *Event-Related Potentials* – ERPs):

- Potenciais espontâneos: essas respostas cerebrais não dependem de estímulo externo (WOLPAW et al., 2002), mas da intenção ou processo imaginativo do usuário do sistema de interface cérebro-máquina, o que pode ser considerado uma vantagem; uma desvantagem, porém, é que a imagética motora é uma habilidade que deve ser aprendida e necessita, portanto, de treinamento por parte do usuário (GRAIMANN; ALLISON; PFURTSCHELLER, 2010). Exemplo desse tipo de resposta são os ritmos sensomotores, baseados na modulação voluntária de potenciais espectrais (LOPEZ et al., 2009);
- Potenciais evocados: são respostas resultantes de algum tipo de estímulo sensorial promovido pelo sistema de interface cérebro-máquina (WOLPAW et al., 2002), o que restringe o tipo de resultado obtível a respostas binárias, como abordado nas próximas seções; entretanto, sua vantagem reside no fato de que necessita pouco ou nenhum treinamento prévio comparado à imagética motora (GRAIMANN; ALLISON; PFURTSCHELLER, 2010). São distinguíveis por suas localizações no escalpo de eletrodos e latências, ou tempo relacionado ao evento visualizado no EEG (GRAIMANN; ALLISON; PFURTSCHELLER, 2010). Exemplos desse tipo de resposta são os potenciais visuais evocados e potenciais P300 (LEBEDEV; NICOLELIS, 2006).

A seguir, apresenta-se uma sucinta descrição para cada um dos tipos de potenciais mencionados.

Ritmos Sensomotores

Segundo Graimann, Allison e Pfurtscheller (2010), esses potenciais podem ser obtidos a partir de regiões específicas do córtex cerebral: se gravações de EEG são feitas sobre áreas corticais sensomotoras primárias (correspondentes às áreas 1 a 6 no Mapa de Brodmann), observa-se as frequências conhecidas como ritmos μ , de 8 a 12 Hz (análogas à faixa α , porém típicas dessas regiões), e ritmos β , de 18 a 26 Hz, estando associadas, portanto,

a áreas corticais conectadas a respostas neuromusculares. O movimento, preparação para o movimento, ou mesmo imaginação de movimento, é normalmente procedido por uma diminuição desses ritmos sensomotores (dessincronização relacionada a evento, ou *Event-Related Desynchronization* – ERD); após o movimento, por sua vez, ocorre a chamada sincronização relacionada a evento (*Event-Related Synchronization* – ERS). Por meio da ERD e imaginação de movimento, é possível aplicar esses potenciais para tarefas cuja resposta é discreta ou proporcional (dentro de uma faixa possível de valores), incluindo movimentação de membros do corpo, contração de músculos e controle de um cursor em duas dimensões (GRAIMANN; ALLISON; PFURTSCHELLER, 2010).

Potenciais Visuais Evocados

Potenciais visuais evocados (*Visual-Evoked Potentials*, VEP) dependem principalmente das propriedades de um estímulo visual apresentado ao indivíduo, e não têm demonstrado variar com a intenção do usuário durante sua aplicação (GRAIMANN; ALLISON; PFURTSCHELLER, 2010). O exemplo mais comum dessa categoria é chamado de Potencial Visual Evocado em Estado Estacionário (*Steady-State Visual-Evoked Potential* – SSVEP), resposta cerebral obtida a partir da fixação do olhar em um estímulo visual modulado em frequências específicas, e será abordado em detalhes neste texto devido à sua relevância no contexto deste trabalho. Para esse tipo de potencial, cada estímulo está associado a um comando específico, ou tarefa possível de ser executada, de modo que ele poderia ser útil em aplicações que envolvam atenção seletiva ou resposta binária (como sim ou não) (GRAIMANN; ALLISON; PFURTSCHELLER, 2010): seleção de opções ou letras na tela do computador e navegação bidimensional de um cursor ou cadeira de rodas são alguns exemplos.

P300

Os potenciais P300 são obtidos a partir de um protocolo conhecido como *oddball*: quando um estímulo visual, auditivo ou tátil que é infrequente e desejado é intercalado com outros estímulos frequentes, é possível se visualizar um pico de voltagem evocado aproximadamente 300 ms após o início do estímulo (GRAIMANN; ALLISON; PFURTSCHELLER, 2010); ou seja, a latência intrínseca a esse tipo de potencial é bem definida. Esse potencial também implica em tarefas de atenção seletiva, e pode ser utilizado em aplicações similares às dos potenciais visuais evocados.

Potenciais Corticais Lentos

De acordo com Graimann, Allison e Pfurtscheller (2010), esses potenciais, por sua vez, ocorrem entre 0,5 e 10 segundos após a apresentação de um estímulo e podem ser visualizados no EEG, caracterizando-se pelo seguinte comportamento: os potenciais corticais lentos negativos caracterizam atividades de preparação mental, enquanto os positivos, inibição mental, refletindo, respectivamente, um aumento e diminuição na excitação cortical dos neurônios. A partir de treinamento apropriado, é possível que usuários do sistema de ICM baseado nesses potenciais controlem sua amplitude. Aplicações para esse tipo de potenciais, assim como VEP e P300, incluem: processamento de palavras, atenção seletiva ou tarefas de controle simples.

2.3.2 Técnicas Invasivas

O uso de interfaces cérebro-máquina invasivas envolve implantação cirúrgica de eletrodos – ou redes de múltiplos eletrodos, por de meio da abertura do crânio em um processo chamado de craniotomia (GRAIMANN; ALLISON; PFURTSCHELLER, 2010) – capazes de mensurar padrões de atividade dos neurônios, que codificam informações comportamentais relevantes. Quando os eletrodos são dispostos na superfície do córtex, o sinal gravado é chamado de eletrocorticograma (ECoG), procedimento que não causa danos aos neurônios, pois os eletrodos não penetram o cérebro; quando o sinal é gravado a partir de eletrodos que penetram, de fato, o tecido cerebral, tem-se os chamados sinais intracorticais (GRAIMANN; ALLISON; PFURTSCHELLER, 2010).

Vantagens de Técnicas Invasivas

As vantagens para esse tipo de técnica são as seguintes (GRAIMANN; ALLISON; PFURTSCHELLER, 2010):

- os sinais apresentam qualidade de sinal excelente, bem como melhor resolução espacial e maior faixa de frequência aplicável (em comparação ao EEG);
- é menos suscetível a artefatos;
- o processo incômodo de utilização e reutilização de eletrodos visto na técnica não invasiva é desnecessário.

Desvantagens de Técnicas Invasivas

As desvantagens, por sua vez, são (GRAIMANN; ALLISON; PFURTSCHELLER, 2010):

- o fato de requerer implantação cirúrgica e, portanto, ser um procedimento mais complexo e de maior risco;
- envolve, desta forma, questões éticas e financeiras que não são um problema para aplicações não invasivas;
- eletrodos implantados cirurgicamente por muito tempo podem causar reações do tecido cerebral, o que pode deteriorá-los;
- técnicas não invasivas podem já atender às demandas do indivíduo, dependendo de seu propósito.

Como visto, dois são os tipos de sinais obtíveis a partir de técnicas invasivas: electrocortigrama (ECoG) e sinais intracorticais. A seguir, apresenta-se os potenciais característicos de cada um desses sinais.

Atividade de ECoG

Devido à sua maior amplitude e resolução, bem como ao fato de estar menos suscetível a artefatos, algumas características são evidenciadas nos electrocortigramas, ambos no domínio do tempo e frequência, e não em sinais EEG, como atividade de maior definição na faixa γ , por exemplo, de 30 a 200 Hz (GRAIMANN; ALLISON; PFURTSCHELLER, 2010). Apenas a atividade em baixa frequência γ é evidenciada no EEG, e considerando que essa frequência está associada tanto ao movimento quanto atenção e imaginação motora, pode-se aplicar sistemas ICM com uma resolução maior, provida pelo ECoG, e reconhecer a atividade em frequências maiores. Possíveis aplicações para esse potencial são: controle de cursor, direção de movimento e mesmo fala sintetizada (GRAIMANN; ALLISON; PFURTSCHELLER, 2010).

Potenciais Locais de Campo

Como abordado neste texto, sabe-se que existem potenciais de ação, relacionados à pulsação individual de neurônios, e potenciais de campo, originados da porção de corrente

elétrica que flui através do espaço de fora da célula neural. Segundo Graimann, Allison e Pfurtscheller (2010), esses potenciais são resultado da medição da atividade de neurônios envolvidos sob a área do eletrodo empregado, sendo estudados tanto no domínio do tempo quanto da frequência, e um exemplo conhecido é o chamado Potencial Evocado de Movimento (*Movement-Evoked Potential*, MEP), obtido a partir da gravação em áreas motoras do córtex, antes do início do movimento. Há registros de aplicação desses potenciais em pacientes com restrições motoras severas (como *locked-in*) para mover um cursor em uma dimensão por meio da modulação de amplitude desses sinais (GRAIMANN; ALLISON; PFURTSCHELLER, 2010).

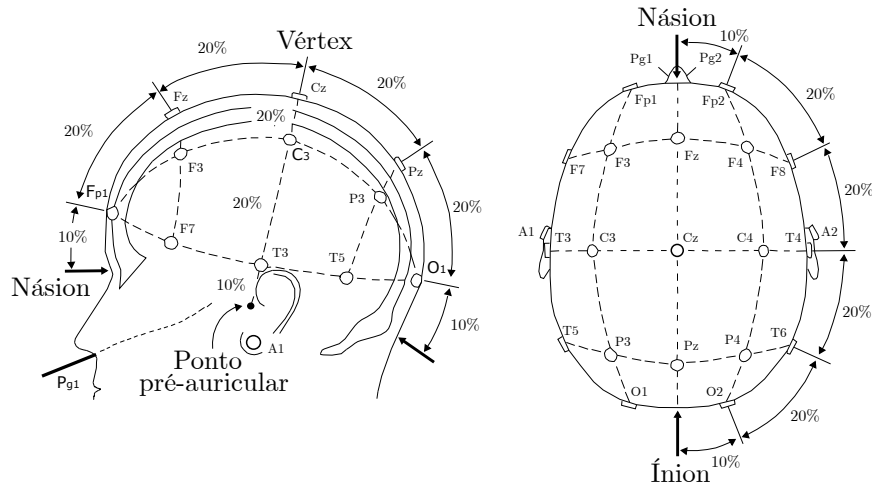
Potenciais de Ação (ou Atividade de um Único Neurônio)

Esses sinais, por fim, são obtidos a partir da gravação da atividade de um único neurônio utilizando microeletrodos. Nesse caso, busca-se registrar as taxas de pulsos (disparos) de neurônios individuais (GRAIMANN; ALLISON; PFURTSCHELLER, 2010), como visto na Figura 7. Neste caso, conhece-se aplicações desse tipo de potencial para tarefas como: controle de posição de cursor, ou movimento em uma dimensão.

2.4 SISTEMA 10-20, 10-10 E 5-10

Independente da técnica e protocolo adotados para a aquisição dos sinais, a disposição dos eletrodos é fundamental para se obter potenciais, de fato, representativos da região cerebral de interesse, e com bons níveis de amplitude. Para atender esses requisitos, posiciona-se os eletrodos de acordo com as macrorregiões do córtex cerebral e o sistema 10-20 (resultado de estudos realizados inicialmente por H. H. Jasper (JASPER, 1958) e fruto da necessidade de padronização do posicionamento de eletrodos em eletroencefalografia), determinado a partir de marcações com distância de 10% e 20% de intervalo entre si (WEBSTER, 2006), como mostra a Figura 10. Nota-se que as iniciais marcadas para cada eletrodo correspondem à respectiva região cerebral; por exemplo: T3 corresponde ao eletrodo 3, no lobo temporal; P4, por sua vez, corresponde ao eletrodo 4, no lobo parietal; O1, por fim, corresponde ao eletrodo 1, no lobo occipital.

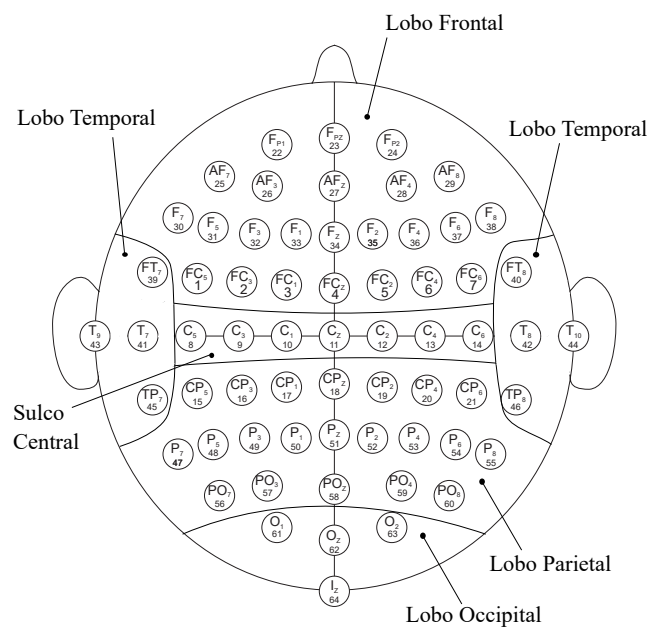
Figura 10 – Posicionamento de eletrodos com o sistema 10-20.



Fonte: Adaptado de Sepulveda et al. (2015)

Atualmente, há variações desse sistema, como o sistema 10-10 e o sistema 10-5 (OOSTENVELD; PRAAMSTRA, 2001). Para clareza de entendimento e visualização dos eletrodos de acordo com as macrorregiões cerebrais, a Figura 11 apresenta 64 eletrodos organizados de acordo com o sistema 10-10, mostrando, entre outras, a transição de lobos parietal e occipital, representados pela parte posterior do córtex, cujos eletrodos têm as identificações PO ou O, que são as de maior relevância no contexto deste trabalho.

Figura 11 – Posicionamento de eletrodos de acordo com as macrorregiões do córtex cerebral e o sistema 10-10.

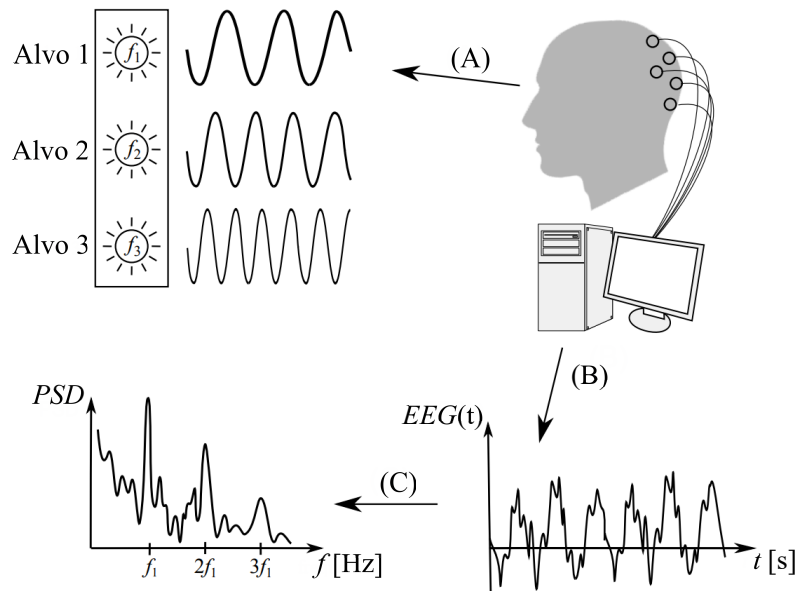


Fonte: Adaptado de Schalk et al. (2004).

2.5 POTENCIAIS VISUAIS EVOCADOS EM ESTADO ESTACIONÁRIO

De acordo com Wang et al. (2006), Potenciais Evocados Visuais (*Visual Evoked Potentials*) são potenciais derivados da resposta do cérebro a estímulos visuais, refletindo, dessa forma, seu mecanismo de processamento de informação visual. Potenciais Evocados Visuais em Estado Estacionário — cujo primeiro sistema BCI proposto foi apresentado em 1996 (Calhoun e McMillan (1996), apud Zhu et al. (2010)) — são as respostas a estímulos visuais modulados aproximadamente a uma frequência maior que 6 Hz, e podem ser gravadas a partir de escalpos de eletrodos posicionados sobre o córtex visual, com máxima amplitude na região occipital. Isso se deve ao fato de que, de acordo com o Mapa Citoarquitetônico de Brodmann (GAREY, 2006), essa região representa o córtex visual, cujas funções incluem visão, cor, movimento e profundidade. A Figura 12 representa a sequência completa do protocolo SSVEP: após o estímulo aos olhos do usuário (A), a sincronização da frequência de pulsos dos neurônios em relação àquela modulada pelo estímulo ocorre; a aquisição (B) e consequente transformação do sinal para o domínio da frequência (C) — via Densidade Espectral de Potência (do inglês, *Power Spectral Density* - PSD), por exemplo — revela qual componente de frequência escolhida para o estímulo.

Figura 12 – Protocolo SSVEP.

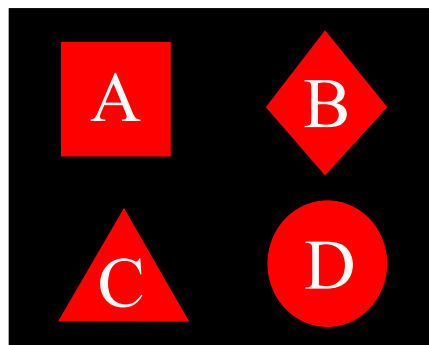


Fonte: Adaptado de Chumerin et al. (2012).

A princípio, os alvos indicados acima podem ser qualquer fonte de luz capaz de ser modulada em uma frequência específica, como um LED ou estímulo oscilante na tela do

computador; ademais, esses alvos não necessariamente devem ser aplicados simultaneamente, mas sim dependendo do número de graus de liberdade que se deseja e seja possível. Desta forma, a intenção do usuário é conhecida a partir do momento em que este fixa seu olhar sobre uma das frequências de estímulo. Por exemplo: ao olhar para um alvo modulado em 10 Hz, após o reconhecimento do padrão, realiza-se a ação A a partir de um atuador; ao olhar para um alvo modulado em 11 Hz, realiza-se a ação B; ao se fixar o olhar no alvo modulado em 12 Hz, realiza-se a ação C; ao olhar para o alvo de 13 Hz, por fim, realiza-se a ação D. A Figura 13 ilustra 4 estímulos para o protocolo SSVEP, podendo ser aplicáveis na tela de um computador, cada qual representado por um formato distinto, indicando as ações possíveis de serem executadas. Nota-se que, neste protocolo, o que se entende por “intenção do usuário” não é um reflexo espontâneo de seu pensamento ou processo imaginativo (isto é, não está na categoria de imagética motora), mas sim a possibilidade de escolha de uma ação, por parte de pessoas com limitações motoras severas (como tetraplegia ou ELA, por exemplo), dentre as ações possíveis de serem realizadas pelo sistema BCI.

Figura 13 – Exemplo de estímulos para o protocolo SSVEP.



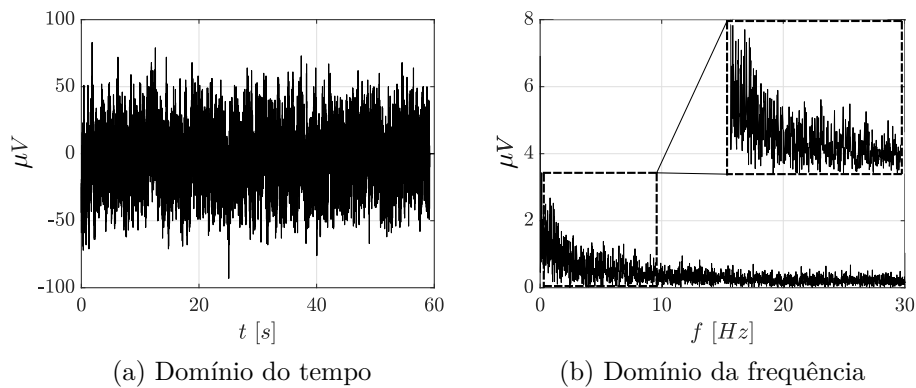
Fonte: Elaborado pelo autor.

Para fins de exemplo e maior clareza de entendimento quanto ao protocolo em questão, as Figuras 14 a 18 são apresentadas; estas ilustram a diferença, no domínio do tempo e da frequência, dos sinais adquiridos de um paciente sob a simples atividade de estar com os olhos abertos (isto é, sem estímulo visual específico), obtido a partir de um banco de dados *online* de domínio público (SCHALK et al., 2004), e outro sob estímulos visuais distintos, modulados em 5,75 Hz, 7,75 Hz, 8,75 Hz, e, em seguida, 9,75 Hz, respectivamente, obtido de outro banco de dados de domínio público (ZHANG et al., 2014) — neste caso, pode-se representar cada uma dessas frequências correspondentemente às ações A, B, C e D, o que totaliza 4 graus de liberdade.

O fato de não se utilizar valores nominais de 6 Hz, 8 Hz, 9 Hz e 10 Hz se deve, essencial-

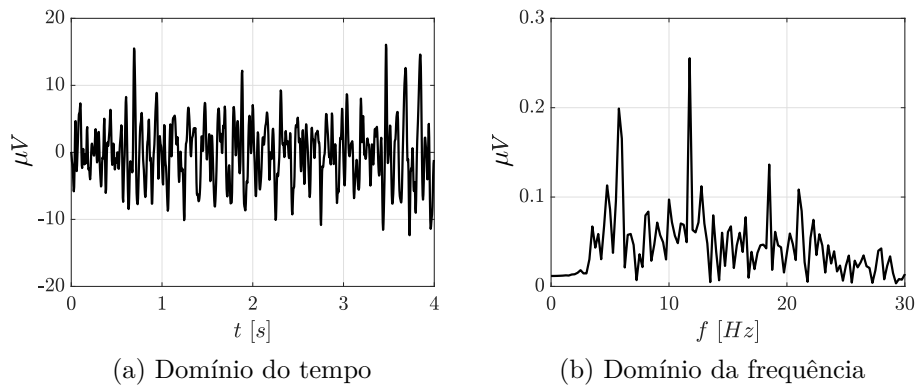
mente, dada a natureza não-linear da dinâmica cerebral, ao desejo de evitar superposição de harmônicos causada pelos estímulos, o que certamente aumenta a probabilidade de identificação incorreta. Também, devido a restrições de cunho fisiológico, comentadas ao longo deste texto, é necessário manter o maior número possível de opções para seleção de frequência; por isso a opção dos autores em utilizar frequências próximas uma à outra. Todos os sinais são referentes à região occipital, como visto na Figura 11. Percebe-se que os sinais obtidos com protocolo SSVEP apresentam uma componente bem definida (e, embora não seja possível se visualizar em todos os casos, há componentes harmônicas também, evidenciando o comportamento não-linear da dinâmica cerebral), o que demonstra que é possível, de fato, conseguir um sinal modulado a uma frequência específica de forma induzida em um curto período de tempo. Esse fato é importante, especialmente quando se considera a possibilidade de integração com atuadores mecânicos, de modo que quanto mais precisa e rápida é a resposta, maior autonomia do usuário.

Figura 14 – Paciente com olhos abertos sem estímulo visual específico.



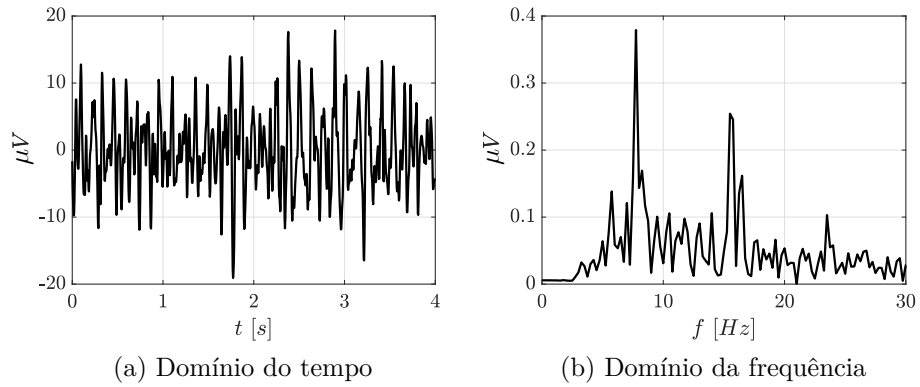
Fonte: Elaborado pelo autor.

Figura 15 – Paciente sob estímulo visual modulado em 5,75 Hz.



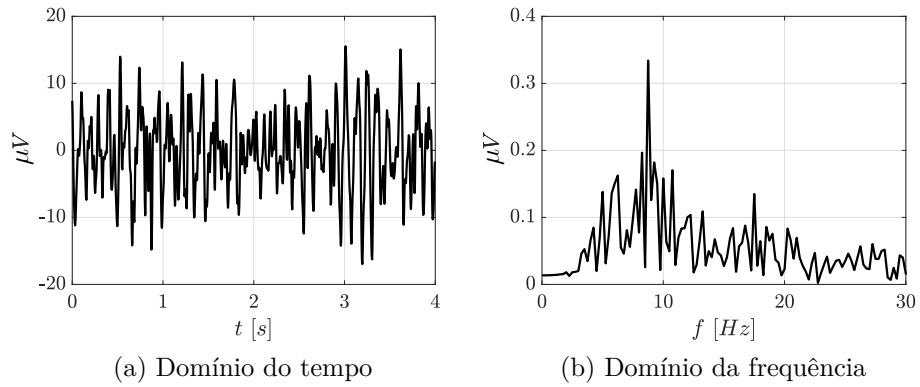
Fonte: Elaborado pelo autor.

Figura 16 – Paciente sob estímulo visual modulado em 7,75 Hz.



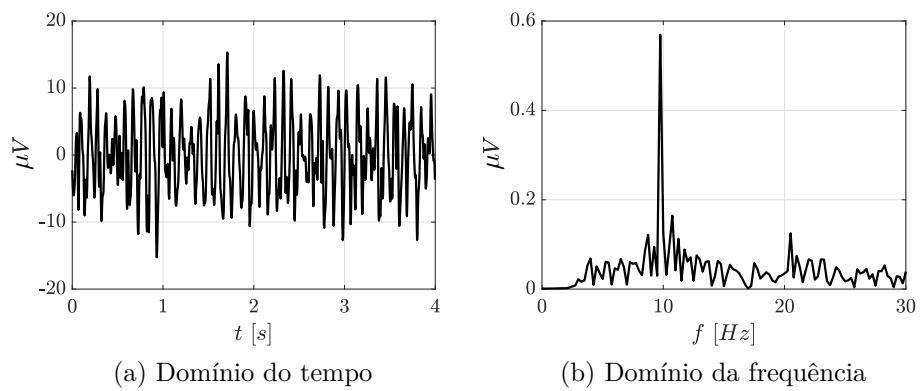
Fonte: Elaborado pelo autor.

Figura 17 – Paciente sob estímulo visual modulado em 8,75 Hz.



Fonte: Elaborado pelo autor.

Figura 18 – Paciente sob estímulo visual modulado em 9,75 Hz.



Fonte: Elaborado pelo autor.

2.5.1 Adequabilidade do Protocolo SSVEP

Dentre outros possíveis protocolos, como já abordado, há razões pelas quais se optar pelo SSVEP no contexto deste trabalho, explicadas brevemente a seguir:

1. é um protocolo que utiliza uma técnica não invasiva e, portanto, pode ser aplicado de forma mais segura (não envolvendo questões financeiras e éticas mais profundas), relativamente simples e rápida;
2. os sistemas BCI baseados em SSVEP abrangem uma faixa em frequência limitada, que corresponde a do estímulo efetuado. É possível, portanto, estabelecer uma relação bem definida entre sinal de entrada e sinal esperado de saída, o que facilita o reconhecimento;
3. como abordado anteriormente, em comparação aos potenciais espontâneos, esse protocolo carece de pouco ou nenhum treinamento;
4. o reconhecimento dos potenciais SSVEP tem apresentado altas porcentagens de acerto usando diferentes técnicas (LIN et al., 2006; BIN et al., 2009; KALUNGA et al., 2013; ZHANG et al., 2014; NAKANISHI et al., 2015; SAGAHÓN-AZÚA; TOVAR-CORONA; ZÚNIGA-VALLADARES, 2017; ZHANG et al., 2018);
5. tem demonstrado desempenho confiável em termos de exatidão e resposta no tempo (mesmo com um número reduzido de eletrodos) a uma taxa relativamente alta de transferência de informação (KWAK; MÜLLER; LEE, 2017);
6. apesar de estar mais sujeito a artefatos devido à utilização de EEG, o reconhecimento desses potenciais tem mostrado bom desempenho mesmo durante a execução do protocolo sob determinadas perturbações, como falar, pensar e ouvir (İŞCAN; NIKULIN, 2018);
7. dado que o resultado obtido a partir do reconhecimento de padrões neste protocolo é binário, ele é interessante para se testar a acurácia e robustez da técnica empregada em termos de acertos e classificações incorretas, fato especialmente útil no contexto deste trabalho, onde diferentes técnicas são aplicadas para fins comparativos;
8. também no referente a tarefas de resposta binária, as possibilidades de aplicação deste protocolo, embora mais simples, são maiores que aquelas envolvendo ritmos sensomotores ou potenciais que necessitem de técnicas invasivas, por exemplo.

Uma vez compreendido o conceito do protocolo SSVEP, apresenta-se, a seguir, uma seção contendo o estado da arte com algumas das principais informações no que tange à seleção de frequências neste protocolo.

2.6 ESTADO DA ARTE

Conforme apresentado na seção anterior, apesar dos sistemas BCI baseados em SSVEP serem relativamente recentes, já produziram diversos trabalhos e contribuições ao conhecimento. A fim de se ter uma melhor visão sobre o assunto, bem como as principais características e limitações desse protocolo, esta seção inclui uma breve revisão da literatura. Como o principal objetivo deste trabalho relaciona-se à correta identificação de frequências de estímulo, o foco se dá na seleção e modulação de frequências, levando em conta aspectos fisiológicos, técnicos e operacionais.

De acordo com Volosyak, Cecotti e Gräser (2009), a resposta SSVEP a estímulos visuais oscilantes de frequência f_0 é bem modelada como a superposição de n senoides de frequências $f_0, 2f_0, \dots, nf_0$, isto é, a frequência fundamental e suas harmônicas, bem como recomenda-se considerar $n = 3$ devido à obtenção de uma acurácia de classificação significativamente maior quando comparada a $n = 1$ ou $n = 2$, baseado em um estudo sobre o impacto de se incorporar oscilações harmônicas evocadas visualmente na classificação (MÜLLER-PUTZ et al., 2005). Também, é possível se obter uma resposta na faixa de 1 – 100 Hz, com picos de ressonância em 10, 20, 40 e 80 Hz (Herrmann (2001), apud Volosyak, Cecotti e Gräser (2009)).

Zhu et al. (2010), por sua vez, afirmam que as frequências escolhidas não devem ser sub-harmônicas ou harmônicas uma da outra, bem como estas podem ser divididas em três faixas nas aplicações em SSVEP: baixa (1 – 12 Hz), média (12 – 30 Hz) e alta (30 – 60 Hz), observando-se maiores amplitudes próximas a 10 Hz, seguida por 16 – 18 Hz, e a menor amplitude visualizada na faixa de alta. Ainda, recomendam atenção para efeitos em outras faixas específicas: entre 8 – 13 Hz, faixa alpha, pode ocorrer considerável número de falsos positivos; entre 5 – 25 Hz, avaliações subjetivas constataram maior perturbação e fadiga visual; o efeito dos estímulos pode acarretar crises epiléticas, especialmente entre 15 – 25 Hz (Fisher et al. (2005), apud Zhu et al. (2010)). Dessa forma, embora as maiores amplitudes estejam em baixas frequências, pode-se tentar contornar as restrições citadas aplicando-se maiores frequências de estímulo; aliás, mesmo apresentando menor amplitude, nessas faixas há, também, menor atividade cerebral em comparação às outras (Wang et al. (2006), apud Zhu et al. (2010)). O mesmo trabalho relata que as cores, o

tamanho e intensidade do estímulo, bem como a manutenção do foco ou distração do usuário, influenciam na intensidade de resposta do SSVEP. Por fim, com o intuito de aumentar as opções de escolha de frequências, pode-se modular um único estímulo em duas frequências (frequências duais).

Segundo Pastor et al. (2003), apud Stawicki, Gemblar e Volosyak (2015), as melhores respostas SSVEP são obtidas na faixa de 5 – 20 Hz, com máxima amplitude em 15 Hz. Também, a máxima diferença em frequência confiável detectada é de 0,1 Hz (Chen et al. (2014), apud Stawicki, Gemblar e Volosyak (2015)). Ainda, a fim de se obter uma resposta mais precisa, deve-se evitar frequências mútuas de estímulos, e cada par ou trio destas deve respeitar às seguintes restrições (STAWICKI; GEMBLER; VOLOSYAK, 2015):

$$f_i \neq \frac{[f_j + f_k]}{2} \quad f_i \neq 2f_j - f_k \quad f_i \neq 2f_k - f_j \quad (1)$$

onde i , j e k correspondem aos índices de uma sequência de frequências a serem empregadas. Finalmente, a seleção da frequência deve ser tal que (Sugiarto, Allison e Gräser (2009), apud Xie et al. (2016)):

$$f_{alvo} < \frac{R}{2} \quad (2)$$

onde R é a taxa de atualização da imagem na tela do computador, caso se utilize um para a realização de testes. Essa restrição se deve ao fato de se respeitar o critério de Nyquist-Shannon (SHANNON, 1949) e, desta forma, evitar o efeito de subamostragem (*aliasing*). No que tange à CCA, principal técnica de estudo neste trabalho, justifica-se sua utilização devido a uma série de fatores que, juntos, colaboram para uma identificação simples e precisa; a saber:

1. A CCA tem apresentado robustez e eficiência no reconhecimento de frequências-alvo em protocolo SSVEP, principalmente em contraste com técnicas de identificação tradicionais, como a Análise de Densidade Espectral de Potência (PSDA) (NAKANISHI et al., 2015; LIN et al., 2006);
2. Essa técnica processa o EEG diretamente no domínio do tempo (SAGAHÓN-AZÚA; TOVAR-CORONA; ZÚNIGA-VALLADARES, 2017), reduzindo a quantidade de etapas de comparação de sinais e, conseqüentemente, promovendo uma identificação rápida;
3. Pode melhorar significativamente a relação ruído-sinal (SNR) de sinais SSVEP (Bin et al. (2009), apud Nakanishi et al. (2015)), de modo a obter bom desempenho na

presença de ruído, especialmente com o aumento de número de eletrodos empregados;

4. Considerando janelas temporais curtas ($1 - 4s$), obteve um reconhecimento razoável ($\sim 50 - 80\%$ de acerto) (ZHANG et al., 2014);
5. Tem apresentado desempenho superior em relação a outras técnicas, como Análise de Discriminante Linear, Análise de Componentes Principais e Máquina de Vetores de Suporte, por exemplo (SAGAHÓN-AZÚA; TOVAR-CORONA; ZÚNIGA-VALLADARES, 2017; KALUNGA et al., 2013);
6. Ao utilizar-se a forma clássica da CCA, não é necessário realizar qualquer tipo de treinamento pelo usuário;
7. Seu conceito é simples e de fácil implementação computacional.

Não obstante, há aparentes restrições ou uma necessidade de preencher algumas lacunas: a abordagem clássica da CCA, mostrada em detalhes mais adiante, sempre considera o mesmo valor para ambas as taxas de amostragem do conjunto artificial de sinais de comparação e os sinais de entrada, como visto em Lin et al. (2006), Zhang et al. (2014), Nakanishi et al. (2015), Kwak, Müller e Lee (2017). Também, diversos estudos — Lin et al. (2006), Bin et al. (2009), Li et al. (2011), Kalunga et al. (2013), Sagahón-Azúa, Tovar-Corona e Zúñiga-Valladares (2017) — focam na obtenção de porcentagens de acerto cada vez melhores, ou na comparação entre CCA e outras técnicas, não analisando individualmente como cada parâmetro do sistema influencia no resultado final de correlação. Finalmente, alguns estudos abordam o impacto da adição de harmônicos ao conjunto artificial de sinais na porcentagem de acerto para a CCA e suas variações (BIN et al., 2009; ZHANG et al., 2014), ou o impacto de se incorporar oscilações harmônicas evocadas visualmente (MÜLLER-PUTZ et al., 2005), como visto anteriormente, porém não mencionam detalhes de outros parâmetros.

Como evidenciado nas referências citadas, não obstante pequenas divergências quanto às frequências que apresentam maior amplitude, há um certo consenso sobre suas faixas de ocorrência. Todas as recomendações e restrições, por sua vez, devem-se, em geral, a fatores fisiológicos — uma vez que o sistema em questão é biológico, não apenas físico — e, portanto, não são facilmente contornáveis. Embora significativo conhecimento esteja disponível para o emprego dessa técnica, constata-se na literatura algumas lacunas de entedimento, especialmente para se definir uma forma mais geral de seu uso, o que implica em maior flexibilidade para aplicação prática, bem como correta identificação da

frequência-alvo que, em particular para a formulação atualmente conhecida, pode não ser adequadamente calculada em algumas condições discutidas neste texto.

Quanto à análise de correlação canônica, por fim, pretende-se, a partir de um estudo detalhado e variando-se diversos fatores de influência no reconhecimento (como frequência do sinal de entrada e dos sinais do banco artificial, número de harmônicos considerados, número de períodos, entre outros), generalizar seu comportamento, adquirindo, assim, uma base sólida de conhecimento capaz de auxiliar, também, na otimização do processo envolvendo protocolo SSVEP, bem como demonstrar que o banco de sinais artificiais não necessariamente precisa ser construído com a mesma frequência de amostragem dos sinais de entrada. Assim, entende-se que esse estudo é, também, útil para uma seleção de frequências mais criteriosa para uma melhor compreensão e flexibilização da CCA para fins práticos, inclusive para aplicá-la a outros problemas de engenharia envolvendo reconhecimento de padrões, principalmente aqueles cujos equipamentos não conseguem atingir frequências de amostragem suficientemente altas.

Para a correta aplicação dessa técnica (bem como comparação com outras), é necessário, previamente, um estudo sobre os conceitos básicos de probabilidade e estatística, apresentado no capítulo a seguir.

3 CONCEITOS BÁSICOS DE PROBABILIDADE E ESTATÍSTICA

Embora as respostas dos SSVEP apresentem espectros relativamente estáveis ao longo do tempo, como observado nas Figuras 15 a 18, elas tendem a ser contaminadas por outras atividades cerebrais, ou mesmo ruídos do ambiente. Um algoritmo eficaz que reconheça as frequências dos SSVEP com uma alta precisão em um curto período de tempo é, pois, de extrema importância para o desenvolvimento de um sistema BCI baseado em SSVEP e de alto desempenho (ZHANG et al., 2014).

Segundo Lin et al. (2006), o método mais amplamente empregado para esse fim é a análise por Densidade Espectral de Frequência (PSD), que pode ser estimada a partir do sinal EEG do usuário para uma janela de tempo e, conseqüentemente, ter seu pico de amplitude detectado. A frequência correspondente a esse pico é tida como a frequência de estímulo visual. O periodograma é um método de estimativa de potência espectral não paramétrico que pode ser calculado diretamente pela Transformada Discreta de Fourier (MARPLE, 1987). Portanto, a Transformada Rápida de Fourier (do inglês *Fast Fourier Transform*) pode ser utilizada nesse cálculo, tendo um custo computacional relativamente baixo.

Comparado aos métodos de estimativa de potência espectral, o periodograma não necessita de uma etapa anterior de seleção de ordem e pode ser facilmente implementado. Além disso, quando os dados estão contaminados por ruído, a abordagem do periodograma tem se mostrado mais robusta do que os métodos de estimativa paramétricos (MANOLAKIS; INGLE; KOGON, 2000). Embora a análise baseada em Densidade Espectral de Potência utilizando FFT possa gerar bons resultados, ela tem suas desvantagens, como a possibilidade de ser sensível ao ruído se o sinal a ser analisado for de um único canal (ou de uma montagem bipolar). O processamento de sinais em conjunto, como visto para técnicas como a CCA, usando a informação de covariância dos canais, pode melhorar a relação entre sinal e ruído, o que justifica a melhora nos resultados obtidos em relação a uma abordagem empregando-se apenas FFT (LIN et al., 2006).

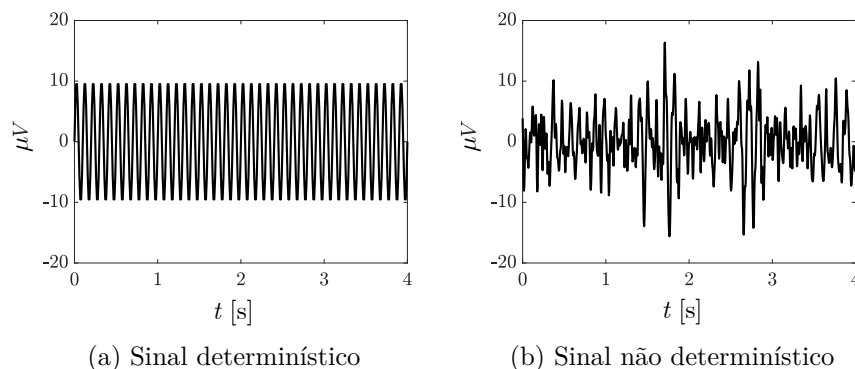
Desta forma, a análise de sinais no domínio da frequência utilizando a FFT ou PSD pode ser empregada como estratégia alternativa. Assim, as técnicas principais, por sua

vez, de forma geral, são fundamentadas a partir de conceitos de probabilidade e estatística. Neste contexto, este capítulo aborda os conceitos básicos que devem ser compreendidos previamente à introdução dessas técnicas.

3.1 ESPAÇO AMOSTRAL E AXIOMAS DE PROBABILIDADE

Espaço amostral de eventos se refere, no caso de um experimento, ao conjunto de todos os seus resultados possíveis; esses eventos são mapeados numericamente a valores reais, chamados de variável aleatória (SHIAVI, 2006). O lançamento de um dado, por exemplo, tem como espaço amostral os números inteiros de 1 a 6. Na engenharia, por sua vez, um sistema que vibra, ou ao qual tem uma frequência de eventos relacionada, resulta em um sinal que pode ser tratado como uma variável aleatória, seja ele determinístico — isto é, pode ser descrito matematicamente, como senos e cossenos, por exemplo — ou não determinístico — não pode ser modelado da mesma forma e deve, portanto, ser tratados a partir de suas propriedades estatísticas. Os sinais das Figuras 14 a 18 são exemplos de sinais não determinísticos: embora haja componentes de frequência bem definidas para os sinais das Figuras 15 a 18, o valor preciso do resultado em amplitude (mV) escolhido em um instante de tempo qualquer não pode ser predito. Uma comparação desses comportamentos é apresentada na Figura 19, onde o sinal determinístico é dado pela seguinte equação: $x(t) = 10\text{sen}(2\pi ft)$, sendo $f = 10$ Hz, definida arbitrariamente para exemplificar.

Figura 19 – Exemplo de sinais determinístico e não determinístico.



Fonte: Elaborado pelo autor.

De acordo com (SHIAVI, 2006), todos os valores em um espaço amostral (ou amostras contidas em um sinal) estão associados a probabilidades de ocorrência: em um espaço amostral S , a cada valor de qualquer conjunto A , por exemplo, seja ele contínuo ou

discreto, está associada uma probabilidade $P[A]$. Esses valores, por sua vez, devem obedecer a axiomas de probabilidade; a saber:

1. $0 \leq P[A] \leq 1$;
2. $P[S] = 1$;
3. $P[A \cup B] = P[A] + P[B]$, se os conjuntos A e B são mutuamente exclusivos;
4. o conjunto nulo contém nenhum dos valores no espaço amostral, e tem probabilidade zero.

Em relação à frequência de ocorrência desses valores no espaço amostral, pode-se descrevê-la a partir de duas funções complementares: função densidade de probabilidade (FDP) e função de distribuição cumulativa (FDC) (ou simplesmente de distribuição de probabilidade). Esses conceitos são abordados com detalhes na próxima seção.

3.2 FUNÇÕES DE DENSIDADE E DISTRIBUIÇÃO DE PROBABILIDADE

Funções densidade de probabilidade (FDP) são obtidas, essencialmente, a partir da quantidade de amostras, em um espaço amostral, contidas em determinada faixa de valores. A função de distribuição cumulativa (FDC) (integral da PDF), por sua vez, pode ser interpretada como a atribuição, de fato, de valores de probabilidade para as faixas em questão. Tomando novamente a função seno (sinal determinístico) como exemplo, se desejar-se calcular, genericamente, a proporção de tempo dt , em um intervalo T , que o sinal passa entre dois níveis de amplitude x e $x + dx$, obtém-se a seguinte equação (NEWLAND, 1993):

$$P[x \leq x(t_0) \leq x + dx] = \frac{2dt}{T} = \frac{dx}{\pi\sqrt{x_0^2 - x^2}} \quad (3)$$

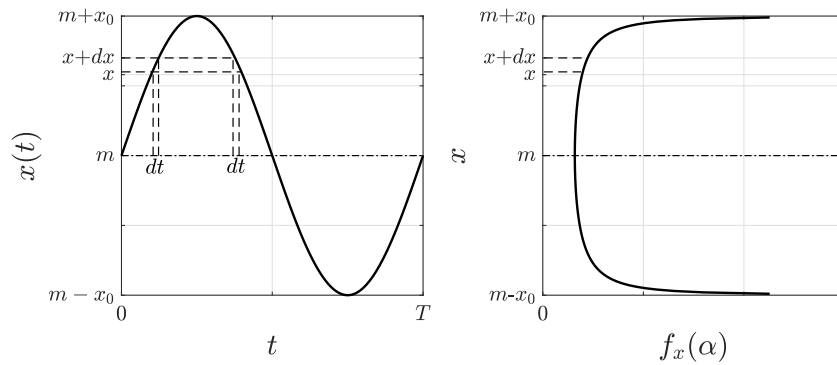
onde x_0 é a amplitude do seno, e x os demais valores. A função densidade de probabilidade de primeira ordem² $p(x)$ é, portanto, definida como:

$$p(x) = \frac{1}{\pi\sqrt{x_0^2 - x^2}}, \quad -x_0 \leq x \leq x_0 \quad (4)$$

²O termo primeiro ordem implica em apenas uma variável aleatória; funções de probabilidade de segunda ordem são, também, abordadas neste texto.

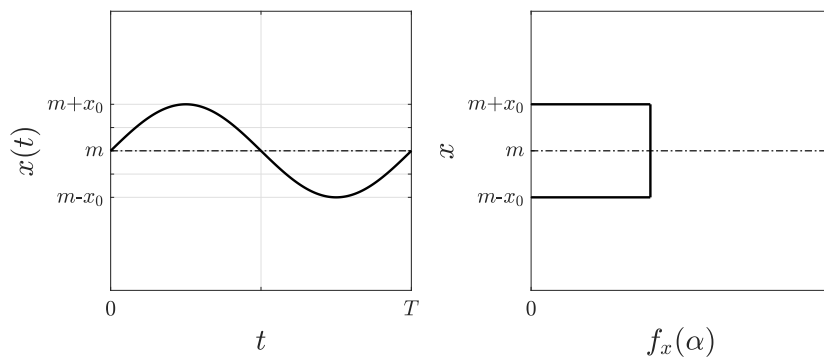
Graficamente, essa relação é mostrada na Figura 20. Nota-se que, devido ao comportamento oscilatório (de amplitude considerável), espera-se que exista maior chance de se escolher uma amostra nos picos ou vales em comparação ao centro da senoide, representada pela média m . Entretanto, ao diminuir-se progressivamente a amplitude do seno, todos os valores têm a mesma probabilidade de serem escolhidos (Figura 21) — a oscilação em torno do eixo das abscissas é menor e, portanto, concentra maior quantidade de amostras.

Figura 20 – Senoide com alta amplitude e sua respectiva PDF.



Fonte: Adaptado de Newland (1993).

Figura 21 – Senoide com baixa amplitude e sua respectiva PDF.

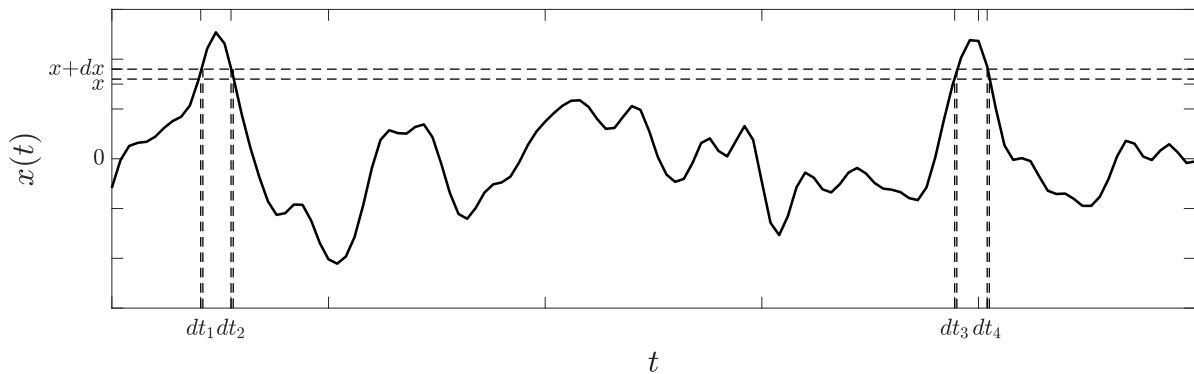


Fonte: Elaborado pelo autor.

Tomando como exemplo um sinal não determinístico, não se pode obter uma expressão analítica, mas, sim, uma aproximação para a fração de tempo que $x(t)$ passa pelos níveis x e $x + dx$, apresentada pela equação 5 (NEWLAND, 1993). A Figura 22, por sua vez, mostra um sinal não determinístico cuja relação para a PDF deve ser, necessariamente, a equação 5.

$$p(x)dx = \frac{(dt_1 + dt_2 + \dots + dt_n)}{T} = \frac{\sum dt}{T} \quad (5)$$

Figura 22 – Sinal determinístico para obtenção da PDF discreta.



Fonte: Adaptado de Newland (1993).

Há diversos tipos de funções de densidade e distribuição de probabilidade, porém sempre as propriedades de ambas também satisfazem aos axiomas de probabilidade; desta forma, para a FDC de uma variável aleatória x , seja ela discreta ou contínua, tem-se que (SHIAVI, 2006):

1. $P[x \leq \alpha] = F_x(\alpha)$ é a probabilidade de x ser menor ou igual a α ;
2. $0 \leq F_x(\alpha) \leq 1, -\infty \leq \alpha \leq \infty$;
3. $F_x(-\infty) = 0, F_x(\infty) = 1$;
4. $F_x(\alpha)$ é não decrescente com α ;
5. $P[\alpha_1 \leq x \leq \alpha_2] = F_x(\alpha_1) - F_x(\alpha_2)$

A FDP f_x da mesma variável aleatória x , contínua, é a derivada da FDC, e suas propriedades são:

1. $f_x(\alpha) \geq 0, -\infty \leq \alpha \leq \infty$;
2. $\int_{-\infty}^{\infty} f_x(u) du = 1$;
3. $F_x(\alpha) = \int_{-\infty}^{\alpha} f_x(u) du$;
4. $P[\alpha_1 \leq x \leq \alpha_2] = \int_{\alpha_1}^{\alpha_2} f_x(u) du$.

Para variáveis aleatórias discretas, a FDC é descontínua: $f_x(\alpha) = P[x = \alpha]$. A FDP para esse caso não é abordada em detalhes nesta seção; entretanto, para casos reais de sinais discretos, utiliza-se a relação 5. A seguir, exemplos conhecidos desses tipos de funções são apresentados.

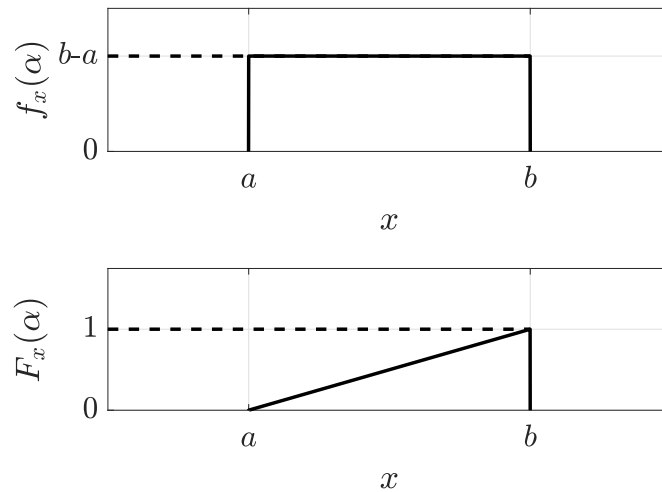
Probabilidade Uniforme

Essa função descreve valores cuja probabilidade de ocorrência dentro de uma faixa finita é igual, como visto para o seno da Figura 21. As expressões representativas de seu comportamento são dadas pela equações 6 e 7. As curvas para suas FDP (acima) e FDC (abaixo), por sua vez, são mostradas na Figura 23

$$f_x(\alpha) = \begin{cases} \frac{1}{b-a}, & a \leq \alpha \leq b \\ 0, & \text{caso contrário} \end{cases} \quad (6)$$

$$F_x(\alpha) = \frac{x}{b-a} \quad (7)$$

Figura 23 – FDP e FDC uniformes.



Fonte: Adaptado de Shiavi (2006).

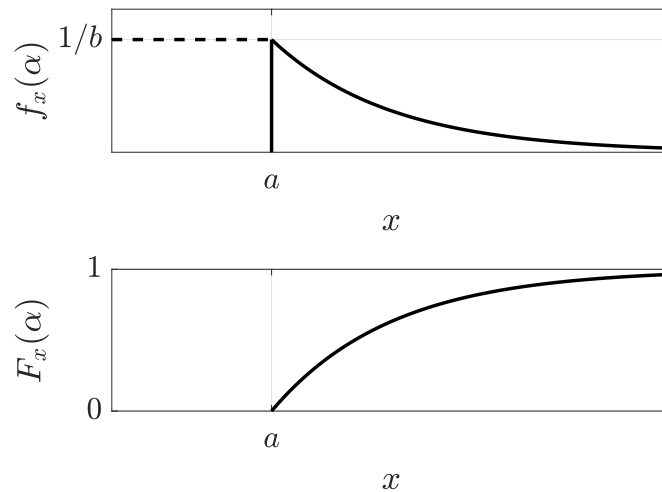
Probabilidade Exponencial

Essa função tem um faixa semi-infinita de valores com a função densidade de probabilidade, e é comumente utilizada para descrever falhas em equipamentos, por exemplo (SHIAVI, 2006). As expressões que a representam são dadas pelas equações 8 e 9. As curvas para sua FDP (acima) e FDC (abaixo) são mostradas na Figura 24.

$$f_x(\alpha) = \begin{cases} \frac{1}{b} e^{-(\alpha-a)/b}, & a \leq \alpha \leq \infty \\ 0, & \alpha < a \end{cases} \quad (8)$$

$$F_x(\alpha) = 1 - e^{-(x-a)/b} \quad (9)$$

Figura 24 – FDP e FDC exponenciais.



Fonte: Adaptado de Shiavi (2006).

Probabilidade Normal (Gaussiana)

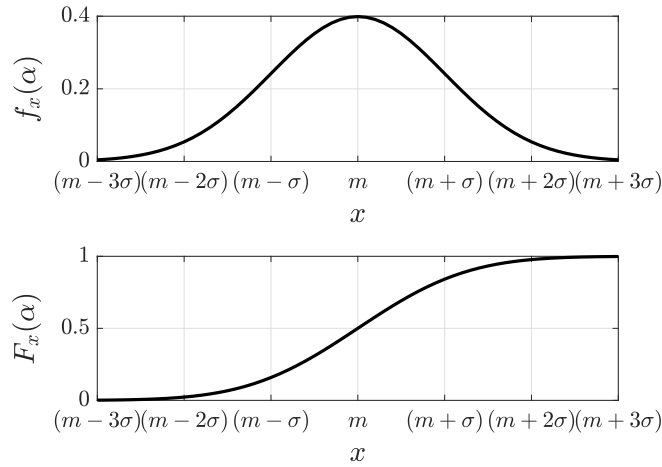
Essa função, por sua vez, amplamente utilizada, é representativa de muitas vibrações aleatórias naturais (NEWLAND, 1993), tais como ruído aleatório ou variações biológicas (SHIAVI, 2006), tendo o conhecido formato de sino. As equações que descrevem seu comportamento são:

$$f_x(\alpha) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty \leq \alpha \leq \infty \quad (10)$$

$$F_x(\alpha) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x-\mu}{\sigma\sqrt{2}} \right) \right] \quad (11)$$

onde os parâmetros m e σ são momentos estatísticos conhecidos como média e desvio padrão, e serão abordados na próxima seção. O operador erf , por sua vez, é chamado de função erro. As curvas para a FDP (acima) e FDC (abaixo) de probabilidade gaussiana são mostradas na Figura 25.

Figura 25 – FDP e FDC gaussianas.



Fonte: Adaptado de Shiavi (2006).

3.3 MOMENTOS DE VARIÁVEIS ALEATÓRIAS

Os momentos não apenas são propriedades estatísticas úteis, mas também, em muitas situações, são as únicas informações disponíveis; eles podem assumir diferentes formas, porém usualmente trata-se de um polinômio (SHIAVI, 2006). O operador que define esse tipo de cálculo é denominado expectativa, e é dado em função do momento $g(x)$ de uma variável aleatória x qualquer, da seguinte forma:

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx \quad (12)$$

Neste caso, considerando $f(x)$ como a FDP $p(x)$ e lembrando que esta pode ser dada em função do tempo (equação 5), tem-se o operador expectativa em termos de t :

$$E[x] = \frac{1}{T} \int_0^T x(t)dt \quad (13)$$

Baseado nesse contexto, pode-se definir alguns conceitos aumentando-se a ordem do polinômio, como o de média, dada da seguinte forma:

$$E[x] = \mu_1 = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{T} \int_0^T x(t)dt \quad (14)$$

ou média quadrática,

$$E[x^2] = \mu_2 = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{T} \int_0^T x(t)^2 dt \quad (15)$$

Essas expressões estão diretamente relacionadas a qual valor se espera obter, na média,

para determinada variável aleatória em um período T , por exemplo. Da mesma forma, é possível definir momentos centrais, isto é, momentos sobre a média, da seguinte forma: $g(x) = (x - \mu)^k$ (SHIAVI, 2006). Os mais conhecidos são aqueles para $k = 2$, $k = 3$ e $k = 4$, denominados variância, assimetria e curtose, cujas equações são, respectivamente:

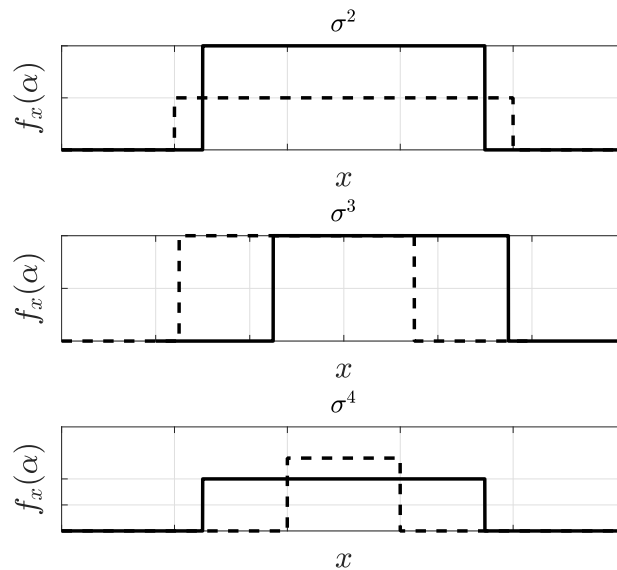
$$E[(x - \mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \frac{1}{T} \int_0^T (x(t) - \mu)^2 dt \quad (16)$$

$$E[(x - \mu)^3] = \sigma^3 = \int_{-\infty}^{\infty} (x - \mu)^3 f(x) dx = \frac{1}{T} \int_0^T (x(t) - \mu)^3 dt \quad (17)$$

$$E[(x - \mu)^4] = \sigma^4 = \int_{-\infty}^{\infty} (x - \mu)^4 f(x) dx = \frac{1}{T} \int_0^T (x(t) - \mu)^4 dt \quad (18)$$

A interpretação gráfica dessas equações pode ser vista na Figura 26. Nota-se que a variância representa o quão dispersos os valores estão em relação à média, isto é, o quanto eles se espalham, deixando de se concentrar em torno do valor central. Conhecido como desvio padrão, por sua vez, $\sigma = \sqrt{\sigma^2}$ mapeia numericamente essa dispersão em faixas de valores, como visto na Figura 25. A assimetria, por sua vez, representa a distorção da curva de densidade de probabilidade em relação a uma posição original conhecida. A curtose, por fim, relaciona-se com o quão afunilado é o topo de uma função de densidade de probabilidade em relação ao valor central (KOTZ; SEIER, 2007).

Figura 26 – Influência da variância, assimetria e curtose em uma PDF.



Fonte: Adaptado de Shiavi (2006).

Todos os momentos estatísticos podem ser normalizados, de modo a representarem a porcentagem do efeito que causam em suas devidas FDPs (KIM; WHITE, 2004):

$$E\left[\left(\frac{x - \mu}{\sigma}\right)^n\right] = \frac{\mu^n}{\sigma^n} \quad (19)$$

É possível, ainda, extrair outra relação importante que auxilia na compreensão do conceito de variância, de alta relevância neste trabalho. O operador esperança é linear; desta forma, expandindo a equação 16, tem-se:

$$\begin{aligned} E[(x - \mu)^2] &= E[(x - \mu)(x - \mu)] = E[x^2 - 2x\mu + \mu^2] = E[x^2] - 2E[x]E[x] + E[x]^2 \\ \sigma^2 &= E[x^2] - E[x]^2 = \mu_2 - \mu_1^2 \end{aligned} \quad (20)$$

ou seja, a variância é, em última instância, a dispersão de magnitude de uma variável aleatória em relação à média central. Uma vez compreendidos os conceitos inerentes a variáveis aleatórias de primeira ordem, pode-se passar às variáveis aleatórias de segunda ordem, conceito fundamental para a análise, comparação e, conseqüentemente, extração de informações sobre duas séries de dados, ou sinais.

3.4 PROBABILIDADE CONJUNTA

A fim de se descrever as relações de dependência entre duas variáveis aleatórias (ou relações de probabilidade de segunda ordem), valores de duas variáveis de um mesmo (ou diferente) conjunto são comparados, obtendo-se, assim, um espaço amostral bidimensional (SHIAVI, 2006). Por exemplo: ao considerar um processo $x(t)$, ou dois processos distintos $x(t)$ e $y(t)$, nos instantes de tempo t_1 e t_2 , as probabilidades são dadas como:

$$P[x(t_1) \leq \alpha_1, x(t_2) \leq \alpha_2] \quad \& \quad P[x(t_1) \leq \alpha_1, y(t_2) \leq \beta_2] \quad (21)$$

No que concerne às funções de distribuição e densidade de probabilidade (neste caso, chamada de conjuntas, ou bivariadas), tem-se, respectivamente:

$$F_{xy}(\alpha, \beta) = P[x \leq \alpha, y \leq \beta] \quad (22)$$

$$f_{xy}(\alpha, \beta) = \int_{-\infty}^{\beta} \int_{-\infty}^{\alpha} f_{xy}(u, v) du dv \quad (23)$$

Logicamente, ambas equações respeitam os axiomas da probabilidade, apresentando, portanto, as seguintes propriedades:

1. $F_{xy}(\alpha, \infty) = F_x(\alpha)$, que a distribuição de probabilidade marginal para a variável x ;
2. $F_{xy}(-\infty, \infty) = 1$;
3. $f_{xy}(\alpha, \beta) \geq 0$;
4. $f_x(\alpha) = \int_{-\infty}^{\infty} f_{xy}(\alpha, v)dv$, que é a função densidade de probabilidade marginal para x .

O mesmo vale para a variável y , fazendo as devidas mudanças de variáveis. Tomando duas séries de dados distintas $x(t)$ e $y(t)$, se desejar-se calcular a probabilidade de que uma amostra escolhida em um tempo t_0 esteja nas faixas de amplitude x e $x + dx$, bem como y e $y + dy$, tem-se (NEWLAND, 1993):

$$P[x \leq x(t_0) \leq x + dx \quad \& \quad y \leq y(t_0) \leq y + dy] = p(x, y)dx dy \quad (24)$$

de modo que, semelhante à equação 23, para a probabilidade em uma faixa de valores, obtém-se:

$$P[x_1 \leq x(t_0) \leq x_2 \quad \& \quad y_1 \leq y(t_0) \leq y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y)dx dy \quad (25)$$

onde $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y)dx dy = 1$. Nota-se que, necessariamente, x e y são indissociáveis. Entretanto, se analisar-se x independente de y , isto é, com y irrestrito, tem-se:

$$P[x \leq x(t_0) \leq x + dx \quad \& \quad -\infty \leq y(t_0) \leq \infty] = dx \int_{-\infty}^{\infty} p(x, y)dy \quad (26)$$

Lembrando a variável aleatória de primeira ordem, como apresentado no início deste capítulo, $P[x \leq x(t_0) \leq x + dx] = p(x)dx$. Ou seja:

$$P[x \leq x(t_0) \leq x + dx] = P[x \leq x(t_0) \leq x + dx \quad \& \quad -\infty \leq y(t_0) \leq \infty] \quad (27)$$

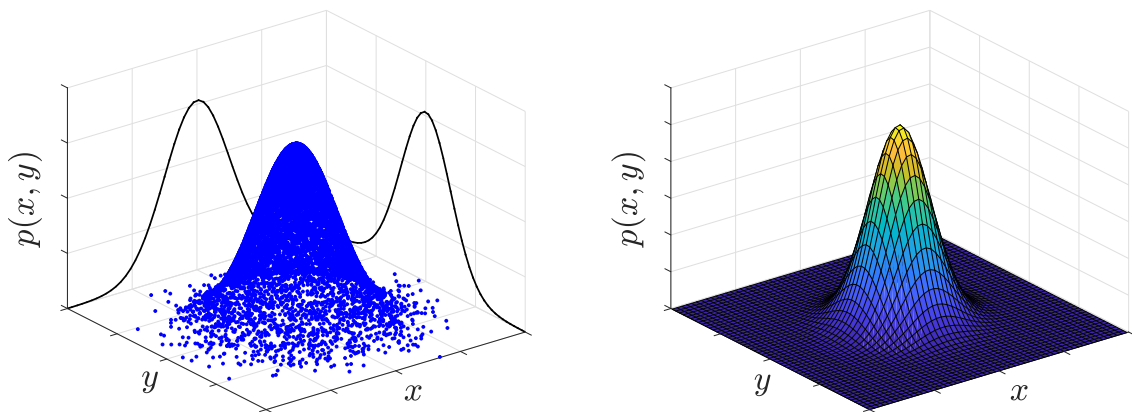
Consequentemente,

$$p(x) = \int_{-\infty}^{\infty} p(x, y)dy \quad (28)$$

$$p(y) = \int_{-\infty}^{\infty} p(x, y)dx \quad (29)$$

Em suma, observa-se que a interpretação para a variável aleatória de primeira ordem é a de um caso particular obtido a partir das funções de densidade de probabilidade de segunda ordem, e que estabelece, portanto, uma relação de independência entre duas séries de dados. A representação gráfica desse efeito é apresentada na Figura 27, onde constam as distribuições marginais (independentes), bem como o espaço amostral bidimensional gerado pela distribuição conjunta. A função adotada para a imagem é a distribuição Gaussiana.

Figura 27 – Distribuições marginais e conjunta.



Fonte: Elaborado pelo autor.

Em relação aos momentos de distribuições conjuntas, aplica-se o mesmo conceito visto na seção anterior. Não obstante, neste caso tratam-se de variáveis aleatórias de segunda ordem, a integral simples passa a ser uma integral dupla, da seguinte forma:

$$E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad (30)$$

Assim, ao invés de se considerar, por exemplo, a média quadrática x^2 , ou a variância σ^2 , considera-se o produto das médias $g(x, y) = xy$ e a denominada covariância $g(x, y) = (x - \mu_x)(y - \mu_y)$. Este último conceito é de extrema importância no contexto deste trabalho, sendo então discutido em detalhes nos capítulos seguintes. Aplicando o operador expectativa a ambos, tem-se:

$$E[xy] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy = \frac{1}{T} \int_0^T x(t)y(t) dt \quad (31)$$

$$E[(x - \mu_x)(y - \mu_y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy$$

$$E[(x - \mu_x)(y - \mu_y)] = \frac{1}{T} \int_0^T (x(t) - \mu_x)(y(t) - \mu_y) dt \quad (32)$$

Por fim, baseando-se no mesmo raciocínio da equação 20, obtém-se a relação entre esses dois momentos:

$$E[xy] = \sigma_{xy} + \mu_x \mu_y \quad (33)$$

Uma vez abordada e compreendida probabilidade conjunta, pode-se prosseguir para outro conceito importante na interpretação de variáveis aleatórias deste trabalho: probabilidade condicional, que será discutido na próxima seção.

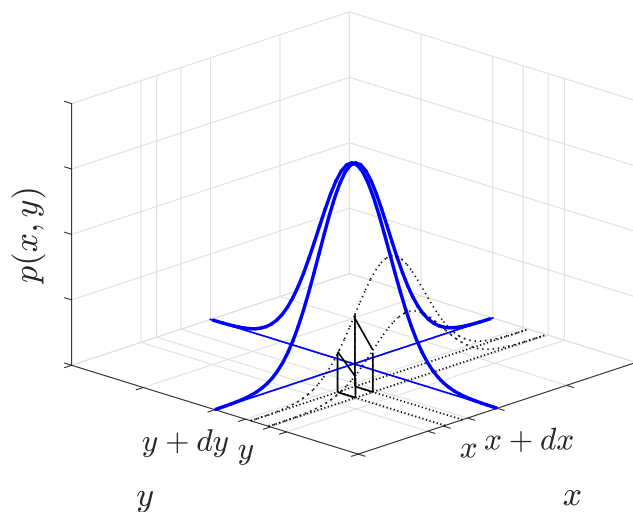
3.5 PROBABILIDADE CONDICIONAL

Como mostram as equações 28 e 29, pode-se determinar a probabilidade de x ou y estar nas faixas x a $x + dx$ ou y a $y + dy$, respectivamente, independentemente. A probabilidade condicional, por outro lado, estabelece a probabilidade de uma dessas variáveis quando a outra, ao invés de irrestrita, deve limitar-se a uma faixa de valores qualquer, reduzindo, por conseguinte, a quantidade de amostras (NEWLAND, 1993). Desta forma, a probabilidade condicional deve ser o volume obtido entre x e $x + dx$ e y e $y + dy$ em relação à probabilidade de y e $y + dy$ quando x é irrestrito (NEWLAND, 1993):

$$p(x|y)dx = \frac{p(x,y)dxdy}{dy \int_{-\infty}^{\infty} p(x,y)dx} = \frac{p(x,y)}{p(y)} \quad (34)$$

A representação gráfica desse conceito é mostrada na Figura 28.

Figura 28 – Representação de probabilidade condicional.



Fonte: Adaptado de Newland (1993).

Ou seja, caso as duas funções densidade de probabilidade sejam independentes, caracterizando sinais não correlacionados, tem-se a expressão dada na equação 35.

$$p(x|y) = \frac{p(x)p(y)}{p(y)} = p(x) \quad (35)$$

Esse resultado é importante para endossar o apresentado sobre probabilidade conjunta, principalmente no que concerne à dependência ou não — determinando, também, correlação ou não, respectivamente — de variáveis aleatórias. Além disso, em termos de análise de sinais, é uma aplicação direta do teorema de Bayes (BISHOP, 2006):

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \quad (36)$$

onde D e w (evidências e parâmetros) são dois conjuntos de dados (ou, no caso, sinais), $p(w)$ e $p(D)$ são as probabilidades individuais de ocorrência de cada uma dessas variáveis, $p(D|w)$ é a probabilidade de D ocorrer dada a existência de w , e $p(w|D)$, por fim, é a probabilidade condicional em questão. Por definição, essas variáveis recebem os nomes de a priori, caso não haja evidência prévia a ser conhecida para a análise (como $p(w)$ e $p(D)$), e a posteriori, caso alguma evidência seja importante para o cálculo, como $p(w|D)$. Desta forma, conceitualmente, a seguinte relação é válida (BISHOP, 2006):

Probabilidade a posteriori \propto Função Verossimilhança x Probabilidade a priori

Considerando os conceitos de probabilidade e estatística, tem-se a base necessária para se entender as técnicas de classificação estatística que são apresentadas no próximo capítulo.

4 CLASSIFICAÇÃO ESTATÍSTICA

Com os avanços da tecnologia, atualmente se pode obter, armazenar e processar grandes conjuntos de dados com facilidade. Sabe-se que, em diversas situações, deseja-se reconhecer padrões importantes entre tais dados (que, frequentemente, estão implícitos), de modo que, se não for possível explicar completamente o processo ou os princípios que regem o comportamento do fenômeno representativo desses dados, se possa ao menos obter uma aproximação boa o suficiente, e útil, para determinados fins (ALPAYDIN, 2009). Essencialmente, classificação estatística é, portanto, a capacidade de, dado um conjunto de dados representativo de um fenômeno qualquer, perceber similaridades/divergências entre as amostras a partir da aplicação de técnicas específicas e, por conseguinte, conseguir separá-las em categorias ou subgrupos, cada qual com suas peculiaridades ou características. A todo esse processo de manipulação estatística e de dados, bem como de extração de parâmetros comuns, dá-se nomes bem conhecidos hoje na literatura, como classificação estatística (*statistical classification*), mineração de dados (*data mining*), ou aprendizado de máquina (*machine learning*) (BISHOP, 2006; ALPAYDIN, 2009)

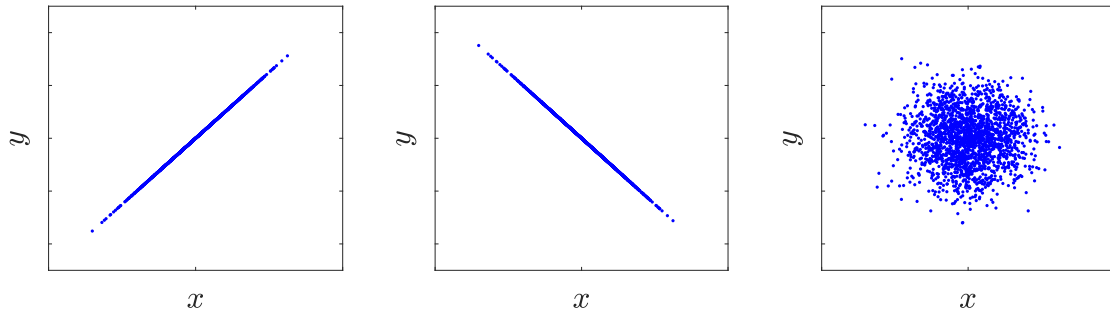
Existem diversas técnicas de classificação estatística. A de maior enfoque deste trabalho é a Análise de Correlação Canônica (CCA), introduzida na próxima seção. Não obstante, ainda neste capítulo, outras metodologias de classificação são estudadas, a fim de se realizar uma comparação entre resultados obtidos. Destacam-se entre elas a Análise de Componentes Principais (*Principal Component Analysis*), Análise de Discriminante Linear/Quadrático (*Linear/Quadratic Discriminant Analysis*) e Máquina de Vetores de Suporte *Support Vector Machine*. Essas técnicas são empregadas no contexto de reconhecimento de padrões eletroencefalográficos para sinais SSVEP.

4.1 ANÁLISE DE CORRELAÇÃO CANÔNICA (CCA)

Para maior compreensão da CCA, inicialmente nesta seção se discute de que modo ocorre a classificação. De forma geral, pode-se entendê-la a partir do conceito de correlação linear, isto é, o grau similaridade linear possivelmente existente entre duas séries de dados. Considere duas variáveis aleatórias x e y : se os valores de x crescem à medida que os valores de y também crescem, diz-se que há correlação; se, por outro lado, os valores de

x crescem conforme os valores de y decrescem, diz-se que há anticorrelação; se, por fim, nenhum desses efeitos corre, não se pode afirmar que existe qualquer grau de similaridade. A visualização desses efeitos, respectivamente, é mostrada na Figura 29.

Figura 29 – Ilustração de correlação, anticorrelação, e ausência de correlação linear.



Fonte: Elaborado pelo autor.

É possível expressar uma relação funcional entre x e y na forma de uma linha reta a partir da minimização por mínimos quadrados (NEWLAND, 1993). Considerando, por conveniência, que a reta desejada cruze o eixo das ordenadas e abscissas no ponto zero (isto é, $E[x] = E[y] = 0$), a equação representativa torna-se:

$$y = m_r x \quad (37)$$

De outra forma, pode-se dizer que o desvio de qualquer amostra em relação à reta é dado por: $\Delta = y - m_r x$. Se aplicar-se o operador expectativa a esse desvio, tem-se que:

$$E[\Delta^2] = E[(y - m_r x)^2] = E[y^2] - 2m_r E[xy] + m_r^2 E[x^2] \quad (38)$$

A fim de se minimizar essa função, aplica-se a derivada em relação à inclinação da reta m e iguala-se o resultado a zero:

$$2m_r E[x^2] - 2E[xy] = 0 \quad (39)$$

ou, alternativamente,

$$m_r = \frac{E[xy]}{E[x^2]} = \frac{y}{x} \quad (40)$$

Lembrando a equação 16 e o fato de que se considera, em particular, média zero, tem-se, neste caso, que a variância é igual à média quadrática (equação 15): $\sigma_x^2 = E[x^2]$

e $\sigma_y^2 = E[y^2]$. Assim, é válido que:

$$\frac{y}{\sigma_y^2} = \frac{x}{\sigma_y^2} \frac{E[xy]}{\sigma_x^2} \quad (41)$$

Cancelando o termo σ_y em ambos os lados da equação e rearranjando-a, tem-se:

$$\frac{y}{\sigma_y} = \left\{ \frac{E[xy]}{\sigma_x \sigma_y} \right\} \frac{x}{\sigma_x} \quad (42)$$

A partir dessa equação, é possível definir o coeficiente de correlação ρ_{xy} , também conhecido como coeficiente de Pearson (RODGERS; NICEWANDER, 1988), da seguinte forma:

$$\rho_{xy} = \frac{\frac{y}{\sigma_y}}{\frac{x}{\sigma_x}} = \left\{ \frac{E[xy]}{\sigma_x \sigma_y} \right\} \quad (43)$$

Por fim, retomando a possibilidade das médias serem diferente de zero, tem-se a equação generalizada para o coeficiente de correlação (NEWLAND, 1993):

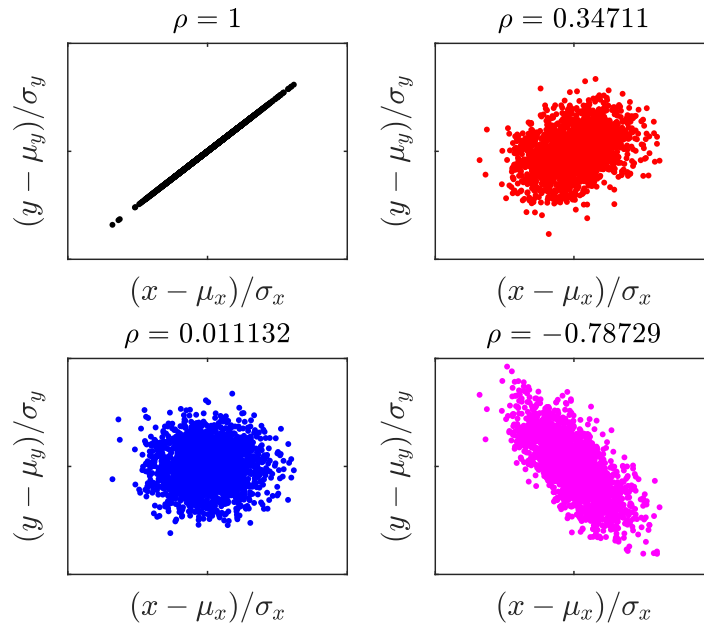
$$\rho_{xy} = \frac{\frac{y-\mu_y}{\sigma_y}}{\frac{x-\mu_x}{\sigma_x}} = \left\{ \frac{E[(x-\mu_x)(y-\mu_y)]}{\sigma_x \sigma_y} \right\} \quad (44)$$

É interessante manter a forma da equação anterior a fim de perceber a equivalência entre cada termo: como se pode notar, o coeficiente de correlação linear nada mais é do que a medida de covariância entre as séries de dados x e y ; quanto maior a covariância (entenda-se: quanto maior a dependência da série x em variar a partir da série y), mais correlacionadas elas estão. Para se tornar claro o quão expressivo é, de fato, esse valor, normaliza-se os dados dividindo-os por seus respectivos desvios-padrão. Em última instância, ρ_{xy} é, também, chamado de covariância normalizada (NEWLAND, 1993). Desta forma, o valor de correlação linear abrange a seguinte faixa de valores:

$$-1 \leq \rho_{xy} \leq 1 \quad (45)$$

Para clareza de entendimento, a Figura 30 apresenta exemplos de correlação de sinais e seus devidos valores para ρ_{xy} ; observa-se que, de fato, quanto mais correlacionados estão os dados, mais próximos se tornam de uma reta.

Figura 30 – Exemplos de valores de correlação.



Fonte: Elaborado pelo autor.

Em termos de classificação, ao comparar-se diversos sinais entre si, procura-se aqueles que melhor estabeleçam uma relação de dependência entre si, isto é, aqueles cujo ρ_{xy} seja mais próximo de 1 ou -1 . Esse valor, ou o máximo possível dentre os existentes, conclui a classificação. Para ilustrar esse efeito, considere como exemplo a equação da densidade de probabilidade gaussiana (equação 10); a expressão geral para sua função densidade de segunda ordem considerando duas variáveis aleatórias x e y é (NEWLAND, 1993):

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} e^{-\frac{1}{2(1-\rho_{xy}^2)} \left\{ \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right\}} \quad (46)$$

Caso $\rho_{xy} = 0$, essa equação se torna:

$$p(x, y) = \left\{ \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \right\} \left\{ \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \right\} = p(x)p(y) \quad (47)$$

ou seja, caso não haja correlação, as densidades de probabilidade são estatisticamente independentes. Baseado nesse contexto, introduziu-se o conceito de Análise de Correlação Canônica, que é abordado na seção seguinte em sua forma clássica. Em seguida, propõe-se uma generalização dessa técnica e apresenta-se uma análise de influência dos parâmetros da formulação por uma abordagem analítica.

4.1.1 Análise de Correlação Canônica Clássica

De acordo com Hotelling (1936), apud Lin et al. (2006), a Análise de Correlação Canônica (CCA) é uma extensão da correlação ordinária de dois conjuntos de variáveis. Essencialmente, pode ser definida como um método estatístico de múltiplas variáveis cujo intuito é buscar uma correlação inerente e, geralmente, implícita, entre duas séries de dados (ZHANG et al., 2014). Assim, dados dois conjuntos de variáveis aleatórias $\mathbf{X} \in R^{I_1 \times J}$ e $\mathbf{Y}_m \in R^{I_2 \times J}$, que são normalizados para ter média zero e variância unitária, a CCA busca um par de transformações lineares $\mathbf{w}_x \in R^{I_1}$ e $\mathbf{w}_y \in R^{I_2}$ de modo que a correlação entre as combinações lineares (também chamadas de variáveis canônicas) $\tilde{\mathbf{x}} = \mathbf{w}_x^T \mathbf{X}$ e $\tilde{\mathbf{y}} = \mathbf{w}_y^T \mathbf{Y}_m$ seja maximizada como (ZHANG et al., 2014):

$$\max_{\mathbf{w}_x, \mathbf{w}_y} (\rho) = \frac{E[\tilde{\mathbf{x}}\tilde{\mathbf{y}}^T]}{\sqrt{E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T]E[\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T]}} = \frac{\mathbf{w}_x^T \mathbf{X} \mathbf{Y}_m^T \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y}_m \mathbf{Y}_m^T \mathbf{w}_y}} \quad (48)$$

Nota-se a semelhança com a equação 43, onde o máximo do coeficiente de correlação ρ em relação a \mathbf{w}_x e \mathbf{w}_y representa a correlação canônica máxima. Se o objetivo for, por exemplo, reconhecer a frequência-alvo, isto é, a frequência do potencial visual evocado a partir de M frequências de estímulo em um sistema BCI baseado em SSVEP, tem-se que $\mathbf{X} \in R^{E \times P}$, onde E é o número de eletrodos, P é o número de pontos de cada sinal obtido no canal, e $\mathbf{Y}_m \in R^{2H \times P}$, normalmente, é o conjunto de sinais de referência pré-construídos na m -ésima frequência de estímulo f_m , sendo $m = 1, 2, \dots, M$, podendo ser representado da seguinte forma (LIN et al., 2006; LI et al., 2011; ZHANG et al., 2014; NAKANISHI et al., 2015):

$$\mathbf{Y}_m = \begin{Bmatrix} \text{sen}(2\pi f_m t) \\ \text{cos}(2\pi f_m t) \\ \vdots \\ \text{sen}(2\pi H_s f_m t) \\ \text{cos}(2\pi H_s f_m t) \end{Bmatrix}, \quad t = \frac{1}{F_s}, \frac{2}{F_s}, \dots, \frac{P}{F_s} \quad (49)$$

onde H é o número de harmônicos, F_s é a taxa de amostragem e sen e cos são as funções seno e cosseno, respectivamente. É importante ressaltar que \mathbf{Y}_m não necessariamente deve ser um banco de sinais artificial; Zhang et al. (2014) introduziram uma variação da técnica de CCA a partir da utilização de um banco formado a partir de sinais de treinamento previamente obtidos de cada indivíduo no protocolo SSVEP, e variações desta técnica podem ser vistas em detalhe em Nakanishi et al. (2015); ainda assim, a forma clássica (e mais usada) consta de um banco artificial. Alternativamente, a equação 48 pode ser

escrita como:

$$\hat{\rho} = \frac{\mathbf{w}_x^T \mathbf{S}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{S}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{S}_{yy} \mathbf{w}_y}} \quad (50)$$

ou, ainda (ANDERSON et al., 1958):

$$\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{w}_x = \hat{\rho}^2 \mathbf{w}_x \quad (51)$$

$$\mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{w}_y = \hat{\rho}^2 \mathbf{w}_y \quad (52)$$

onde \mathbf{S} representa a matriz de covariância respectiva de cada combinação de série de dados, e $\hat{\rho}^2$ é a correlação canônica quadrada. Na prática, este é um clássico problema de autovalores e autovetores, podendo ser escrito da seguinte forma:

$$\mathbf{A} \hat{\mathbf{x}} = \Lambda \hat{\mathbf{x}} \quad (53)$$

onde

$$\mathbf{A} = \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \Leftrightarrow \mathbf{A} = \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \quad (54)$$

$$\hat{\mathbf{x}} = \mathbf{w}_x \Leftrightarrow \hat{\mathbf{x}} = \mathbf{w}_y \quad (55)$$

bem como

$$\Lambda = \hat{\rho}^2 \quad (56)$$

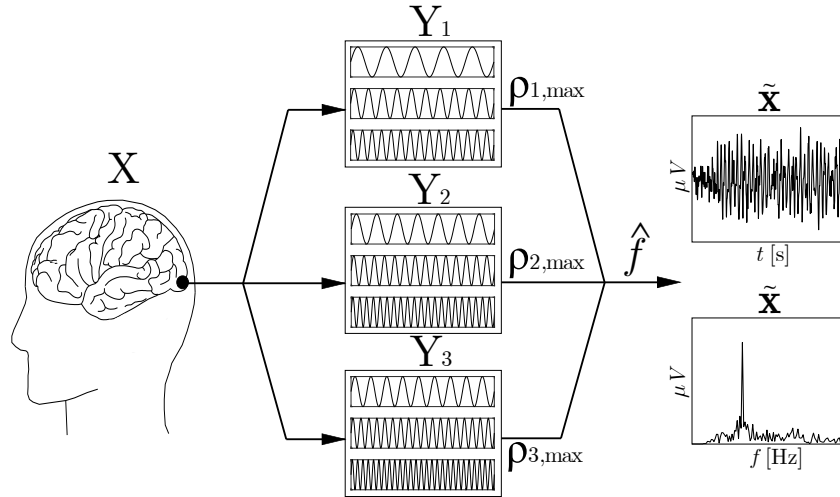
Supondo $m = 1, 2, \dots, M$ e encontrando o máximo autovalor ρ_m na matriz $\hat{\rho}_m$ para cada m , entre \mathbf{X} e \mathbf{Y}_m , a frequência SSVEP é dada, portanto, por (LIN et al., 2006; ZHANG et al., 2014):

$$\hat{f} = \underset{f_m}{\operatorname{argmax}}(\rho_m), \quad m = 1, 2, \dots, M \quad (57)$$

Como apresentado neste trabalho, pode-se demonstrar que essa notação para CCA é um caso particular e, também, que um parâmetro pode ser usado para reescrever essa equação, resultando em uma expressão mais geral, podendo ser aplicada mais convenientemente na prática. Uma ilustração do reconhecimento de SSVEP baseado na CCA é apresentada na Figura 31. Observa-se a aplicação da CCA para um sinal de entrada qualquer e um banco de sinais de referência pré-construídos (isto é, não obtidos a partir de indivíduos) e a obtenção da variável canônica transformada a partir desses dados de teste, representada no domínio da frequência em termos de espectro de potência. Para se efetuar a Análise de Correlação Canônica, embora não haja necessidade de se aplicar al-

goritmos tradicionais para a visualização do sinal no domínio da frequência, como a FFT ou PSD, por exemplo, pode-se utilizá-las de forma complementar, a fim de se comprovar a existência de uma componente de frequência esperada para determinada atividade cerebral.

Figura 31 – Reconhecimento de SSVEP baseado em CCA.



Fonte: Elaborado pelo autor.

Em termos de reconhecimento de padrões (neste caso, padrões encefalográficos, a partir das frequências de estímulo visual), é interessante obter-se uma generalização do comportamento da técnica em uso, a fim de não apenas melhor compreender o processo estatístico e computacional de classificação, mas também conseguir-se flexibilizar sua utilização e ampliar acurácia. Dessa forma, a seção seguinte apresenta uma proposta de generalização da CCA, incluindo informações sobre seu comportamento analítico e variáveis relacionadas a processamento de sinais.

4.1.2 Generalização Proposta para a Técnica CCA

A proposta de generalização da CCA é introduzida baseada no coeficiente de correlação (Pearson), como apresentado pela equação 44:

$$\rho_{xy} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y}$$

onde $x(t)$ e $y(t)$, neste caso, são duas séries de dados quaisquer, μ_x e μ_y suas médias, e σ_x e σ_y seus desvios padrão. A primeira hipótese dessa proposta é que ρ_{xy} corresponde ao máximo autovalor resultante do processo de maximização (Equação 48) — isto é,

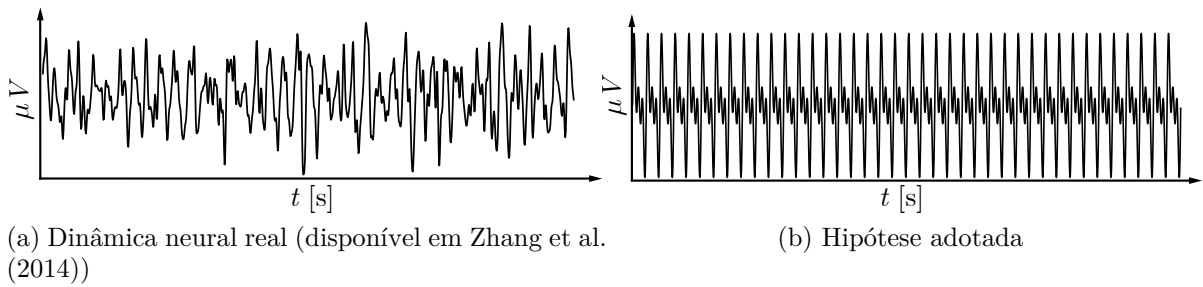
o coeficiente de correlação (autovalor) das variáveis canônicas (projeções das variáveis originais no maior autovetor): $\rho_{max} = \max(\rho_{xy})$. Também, considera-se que, apesar da dinâmica cerebral apresentar atividades mais complexas, como visto na Figura 32a, os sinais de entrada e o conjunto de sinais artificial são descritos como somatórios de senoides (VOLOSYAK; CECOTTI; GRÄSER, 2009) (Figura 32b):

$$x(t) = \sum_{n=1}^N x_n \text{sen}(n\omega_x t + \phi) \quad (58)$$

$$y(t) = \sum_{m=1}^M y_m \text{sen}(m\omega_y t + \varphi) \quad (59)$$

onde x_n e y_m são as amplitudes de cada seno, n e m os respectivos números de harmônicas de cada sinal, ω_x e ω_y as frequências, e ϕ e φ as fases, respectivamente.

Figura 32 – Dinâmica neural real e assunção adotada.



Fonte: Elaborado pelo autor.

Lembrando a equação 16 de momentos de variáveis aleatórias de segunda ordem para distribuição de probabilidade conjunta (covariância), pode-se escrever o coeficiente de Pearson da seguinte forma:

$$\rho_{max} = \rho_{xy} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x) f(y) dx dy}{\sigma_x \sigma_y} \quad (60)$$

ou, de forma intercambiável, no domínio do tempo (equação 16):

$$\rho_{max} = \frac{\frac{1}{T} \int_0^T (x(t) - \mu_x)(y(t) - \mu_y) dt}{\sigma_x \sigma_y} \quad (61)$$

Em particular, para os fins que se pretende à seguinte demonstração, os desvios-padrão são mantidos da forma original, visto que apenas normalizam o resultado final e, também, por se ter maior interesse, nesta demonstração, no numerador da equação, conforme apresenta-se abaixo. Assim, aplicando-se ambos sinal de entrada $x(t)$, cuja frequência se deseja identificar, e banco artificial $y(t)$ na integral, bem como considerando,

por simplicidade, a hipótese de médias iguais a zero ($\mu_x = \mu_y = 0$), obtém-se:

$$\rho_{max} = \frac{\frac{1}{T} \int_0^T \sum_{n=1}^N x_n \text{sen}(n\omega_x t + \phi) \sum_{m=1}^M y_m \text{sen}(m\omega_y t + \varphi) dt}{\sigma_x \sigma_y} \quad (62)$$

sendo M e N o número existente de harmônincos em cada sinal (note que tais sinais são escritos como a soma desses harmônicos). Baseando-se na seguinte identidade trigonométrica:

$$\text{sen}(\alpha)\text{sen}(\beta) = \frac{1}{2}[\cos(\alpha - \beta) - \cos(\alpha + \beta)] \quad (63)$$

pode-se escrever, ainda:

$$\rho_{max} = \frac{\frac{1}{2T} \int_0^T \sum_{n=1}^N \sum_{m=1}^M x_n y_m [\cos(n\omega_x t - m\omega_y t + \phi - \varphi) - \cos(n\omega_x t + m\omega_y t + \phi + \varphi)] dt}{\sigma_x \sigma_y} \quad (64)$$

Fazendo as seguintes substituições de variáveis $(n\omega_x t - m\omega_y t + \phi - \varphi) = u$ e $(n\omega_x t + m\omega_y t + \phi + \varphi) = v$, tem-se: $(n\omega_x - m\omega_y)dt = du$ e $(n\omega_x + m\omega_y)dt = dv$. Retornando à integral, por sua vez:

$$\rho_{max} = \frac{\sum_{n=1}^N \sum_{m=1}^M x_n y_m}{2T \sigma_x \sigma_y} \left[\int_0^T \frac{\cos(u) du}{n\omega_x - m\omega_y} - \int_0^T \frac{\cos(v) dv}{n\omega_x + m\omega_y} \right] \quad (65)$$

Resolvendo as integrais:

$$\int_0^T \frac{\cos(u) du}{n\omega_x - m\omega_y} = \frac{\text{sen}[(n\omega_x - m\omega_y + \phi - \varphi)T] - \text{sen}(0)}{n\omega_x - m\omega_y} \quad (66)$$

$$\int_0^T \frac{\cos(v) dv}{n\omega_x + m\omega_y} = \frac{\text{sen}[(n\omega_x + m\omega_y + \phi + \varphi)T] - \text{sen}(0)}{n\omega_x + m\omega_y} \quad (67)$$

Substituindo o resultado anterior na equação do coeficiente de Pearson, tem-se:

$$\rho_{max} = \frac{\sum_{n=1}^N \sum_{m=1}^M x_n y_m}{2\sigma_x \sigma_y} \left\{ \frac{\text{sin}[(n\omega_x - m\omega_y + \phi - \varphi)T]}{n\omega_x - m\omega_y} \right\} - \frac{\sum_{n=1}^N \sum_{m=1}^M x_n y_m}{2\sigma_x \sigma_y} \left\{ \frac{\text{sin}[(n\omega_x + m\omega_y + \phi + \varphi)T]}{n\omega_x + m\omega_y} \right\} \quad (68)$$

Pode-se observar que a equação acima comporta-se como uma função sinc, cujo máximo é dado na frequência (e harmônicas) de interesse; isto é, quando $n\omega_x = m\omega_y$. Para clareza de entendimento, sem perda de generalidade, assumindo $\phi = \varphi = 0$ e aplicando a regra de L'Hôpital, o primeiro termo dentro da separação por chaves da Equação 68 é máximo

quando:

$$\lim_{(n\omega_x - m\omega_y) \rightarrow 0} \frac{\sin(n\omega_x - m\omega_y)}{n\omega_x - m\omega_y} \Leftrightarrow \lim_{(n\omega_x - m\omega_y) \rightarrow 0} \cos(n\omega_x - m\omega_y) = 1 \quad (69)$$

ou seja, quando ocorre efeito semelhante à “ressonância”³, o que não é válido para o segundo termo ($n\omega_x = -m\omega_y$): a menos que $n\omega_x = m\omega_y = 0$, a curva não atinge um valor extremo, mas apenas adiciona sua contribuição ao resultado final da correlação. O comportamento dessa curva é abordado com mais detalhes na seção de resultados. Baseado nessa condição, ρ_{max} apresenta correlação máxima se $(n\omega_x - m\omega_y) = k2\pi$, para $k = 0, 1, 2, \dots$, e n e $m \in \mathbb{N}^*$, ou $(nf_x - mf_y) = k$. Assumindo f_x como a frequência do sinal a ser identificada pelo algoritmo CCA, e considerando aplicações práticas nas quais os sinais sejam valores discretos (na prática, ρ_{max} é computado para cada incremento df), o máximo valor de correlação atingido na frequência f_y é tal que:

$$f_y = \frac{nf_x - k}{m} \quad (70)$$

o que implica que $f_x = f_y$ se, e somente se, $k = 0$ e $m = n$. Do contrário, a Equação 57 não assegura adequada identificação da frequência-alvo. Note que, se existem harmônicas apenas no sinal de entrada, super-harmônicas são reconhecidas pela CCA; se existem harmônicas apenas no conjunto de sinais artificial, sub-harmônicas são reconhecidas; por fim, se existem harmônicas em ambos os conjuntos, a frequência principal, suas super-harmônicas, e sub-harmônicas de suas super-harmônicas são reconhecidas. Dessa forma, o número total de frequências reconhecidas é, portanto $n_{f_r} = nm$.

Considerando aspectos práticos de processamento de sinais, se estes são discretizados utilizando um intervalo de tempo dt , então $T = t_{max} = dt_i P$ ou $t_{max} = dt_s P$, onde dt_i e dt_s são os intervalos usados no sinal de entrada e conjunto artificial, respectivamente, e P é o número de amostras; alternativamente, $t_{max} = P/F_i$ ou $t_{max} = P/F_s$, onde F_i e F_s são suas respectivas taxas de amostragem — isto é, neste ponto assume-se que o sistema de aquisição, por exemplo, amostra com uma frequência insuficiente, enquanto que o conjunto artificial apresenta uma taxa suficientemente alta. Assim, para cada k -ésimo valor discreto:

$$x(k) = \sum_{n=1}^N x_n \sin(n\omega_x dt_i k + \phi) = \sum_{n=1}^N x_n \sin\left(\frac{n\omega_x k}{F_i} + \phi\right) \quad (71)$$

$$y(k) = \sum_{m=1}^M y_m \sin(m\omega_y dt_s k + \varphi) = \sum_{m=1}^M y_m \sin\left(\frac{m\omega_y k}{F_s} + \varphi\right) \quad (72)$$

³Neste contexto, entende-se por ressonância a correlação máxima apresentada quando a frequência a ser comparada é igual à frequência de entrada do sistema.

Seja $F = F_i/F_s$ a razão de taxa de amostragem, pode-se reescrever $y(k)$ como:

$$y(k) = \sum_{m=1}^M y_m \sin\left(\frac{m\omega_y k F}{F_i} + \varphi\right) \quad (73)$$

Tem-se, então, $x(k)$ e $y(k)$ na mesma variável F_i introduzindo a razão F proposta. Seja, por fim, $t_{max} = n_P/f_m$, onde n_P é o número de períodos. A Equação 68 pode ser expressa, por analogia, de forma mais geral após integrar-se novamente a Equação 61 com as modificações mencionadas acima:

$$\rho_{max} = \frac{f_m \sum_{n=1}^N \sum_{m=1}^M x_n y_m}{2n_P \sigma_x \sigma_y} \left\{ \frac{\sin\left[\left(n\omega_x - m\omega_y F + \phi - \varphi\right)\frac{n_P}{f_m}\right]}{n\omega_x - m\omega_y F} \right\} - \frac{f_m \sum_{n=1}^N \sum_{m=1}^M x_n y_m}{2n_P \sigma_x \sigma_y} \left\{ \frac{\sin\left[\left(n\omega_x + m\omega_y F + \phi + \varphi\right)\frac{n_P}{f_m}\right]}{n\omega_x + m\omega_y F} \right\} \quad (74)$$

A equação acima é considerada de extrema relevância para o presente trabalho, visto que explicita o coeficiente de correlação tanto em termos de seu comportamento analítico quanto aspectos de processamento de sinais (os resultados do capítulo 5 são assim calculados). Na seção de resultados, essa equação deve demonstrar que: a largura de banda $BW[\text{Hz}]$ em torno da frequência de estímulo diminui conforme n_P aumenta, garantindo uma melhor acurácia, e dois novos parâmetros podem ser introduzidos para melhor selecionar a frequência de estímulo e escolher a aplicação do sistema BCI:

$$t_{max} = n_P/f_m \quad (75)$$

onde t_{max} é o tempo máximo da janela temporal do sinal, e:

$$FA = \frac{n_P}{n_{P,ot}} \Leftrightarrow FA = \frac{t_{max,ot}}{t_{max}} = \frac{f_{m,ot}}{f_m} \quad (76)$$

chamado de fator de acurácia (FA), onde $n_{P,ot}$, $t_{max,ot}$ e $f_{m,ot}$ são os valores ótimos do número de períodos, tempo máximo de janela do sinal e frequência de estímulo; também, será demonstrado que a frequência reconhecida é melhor expressa pela seguinte equação:

$$\hat{f} = F[\text{argmax}(\rho_m)], \quad m = 1, 2, \dots, M \quad (77)$$

onde F é um “fator de correção” (razão de frequências de amostragem); é possível relacionar F a $f = f_r/f_m$, onde f_r é a frequência reconhecida pela CCA:

$$f = \frac{1}{F} \quad (78)$$

também, nota-se que o ruído acrescentado ao sinal de entrada influi de forma expressiva no resultado da correlação, sendo ainda mais severo quando o número de componentes harmônicas no sinal de entrada é diferente daquele contido no conjunto de sinais artificiais. Por fim, é possível utilizar a CCA como método de verificação de ajuste de curvas, dado por:

$$\tilde{\rho} = \frac{E[(\hat{x} - \mu_{\hat{x}})(\hat{y} - \mu_{\hat{y}})]}{\sigma_{\hat{x}}\sigma_{\hat{y}}} \quad (79)$$

onde \hat{x} e \hat{y} são as curvas obtidas a partir do diagrama f_r vs ρ , onde f_r é a frequência reconhecida e ρ é o valor absoluto de ρ_{max} . Uma vez detalhada essa técnica, pode-se prosseguir para as outras metodologias de classificação que são abordadas neste trabalho para finalidade principal de comparação.

4.2 OUTRAS METODOLOGIAS DE CLASSIFICAÇÃO

Esta seção aborda técnicas comumente empregadas no contexto de interfaces cérebro-máquina em protocolo SSVEP, e que, eventualmente, são aplicadas paralelamente à CCA no presente trabalho; a saber: Análise de Componentes Principais (*Principal Component Analysis*), Análise de Discriminante Linear/Quadrático (*Linear/Quadratic Discriminant Analysis*) e *Support Vector Machine*.

4.2.1 Análise de Componentes Principais (PCA)

Também conhecida como transformada de Karhunen-Loève, a Análise de Componentes Principais é uma técnica amplamente usada para visualização e redução de dimensionalidade de dados, bem como extração de parâmetros nele contidos (Jolliffe (2002), apud Bishop (2006)), podendo ser definida como a projeção ortogonal dos dados originais em um espaço cuja dimensionalidade é menor, de modo que a variância dos dados projetados seja maximizada (Hotelling (1933), apud Bishop (2006)).

Se deseja-se projetar um conjunto de amostras $\{x_i\}$ ($i = 1, \dots, P$) de dimensionalidade original D em um subespaço de dimensionalidade M , tal que $M < D$, por exemplo, pode-se definir o vetor direção \mathbf{w}_m desse novo espaço, de modo que a projeção é dada como (BISHOP, 2006):

$$\mathbf{x}_p = \mathbf{w}_m^T \mathbf{x}_i \quad (80)$$

Lembrando a equação 14 para média estatística, porém considerando o número finito dado de amostras P , intervalo de tempo dt entre uma amostra e outra, de modo que o período total seja $T = Pdt$, tem-se:

$$\mu = \frac{1}{P} \sum_{n=1}^P x_i \quad (81)$$

Baseado na equação 16, tem-se que a variância para esses dados é dada, portanto, como:

$$\frac{1}{P} \sum_{n=1}^P \{ \mathbf{w}_m^T x_i - \mathbf{w}_m^T \mu \} = \mathbf{w}_m^T \mathbf{S} \mathbf{w}_m \quad (82)$$

onde S é definida como a matriz de covariância, dada por:

$$\mathbf{S} = \frac{1}{P} \sum_{n=1}^P (x_i - \mu)(x_i - \mu)^T \quad (83)$$

Para maximizar a variância projetada com respeito a \mathbf{w}_m , evitando que ocorra, por exemplo, $\|\mathbf{w}_m\| \rightarrow \infty$, a restrição de normalização é imposta: $\mathbf{w}_m^T \mathbf{w}_m = 1$ (BISHOP, 2006). A maximização é efetuada considerando-se uma função-objetivo $f(x)$, uma restrição $g(x)$, e multiplicadores de Lagrange λ_m . Nesse caso, a primeira é dada pela projeção em si, e a última, pela condição de normalização. A forma clássica de proposição desse problema é sob a Lagrangiana L (DUDA; HART; STORK, 2012; BERTSEKAS, 2014):

$$L(x, \lambda_m) = f(x) + \lambda_m g(x) \quad (84)$$

onde busca-se por pontos em que ocorra:

$$\nabla L(x, \lambda_m) = 0 \quad (85)$$

ou seja, procura-se pontos onde o gradiente da função-objetivo seja igual ao gradiente da restrição; em outras palavras, a direção de crescimento da função-objetivo encontra um máximo quando é igual à direção de crescimento da restrição, multiplicada por qualquer fator. No caso em questão, adequa-se as variáveis de forma que: $f(\mathbf{w}_m) = \mathbf{w}_m^T \mathbf{S} \mathbf{w}_m$ e $g(\mathbf{w}_m) = 1 - \mathbf{w}_m^T \mathbf{w}_m$. Assim:

$$L(\mathbf{w}_m, \lambda_m) = \mathbf{w}_m^T \mathbf{S} \mathbf{w}_m + \lambda_m (1 - \mathbf{w}_m^T \mathbf{w}_m) \quad (86)$$

Aplicando-se o operador gradiente, pode-se estabelecer as conhecidas condições de Karush-Kuhn-Tucker (KKT), necessárias para a otimização em problemas diferenciáveis de programação não linear que devem satisfazer a certas restrições (BAZARAA; JARVIS;

SHERALI, 2011):

$$\frac{\partial L(\mathbf{w}_m, \lambda_m)}{\partial \mathbf{w}_m} = 2\mathbf{S}\mathbf{w}_m - 2\lambda_m\mathbf{w}_m = 0 \quad (87)$$

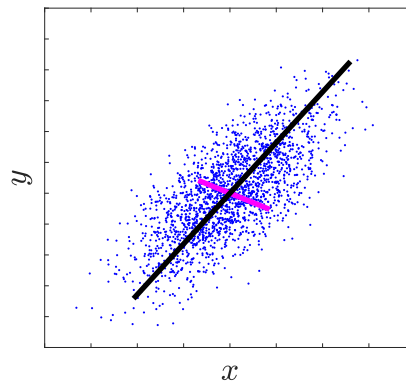
$$\frac{\partial L(\mathbf{w}_m, \lambda_m)}{\partial \lambda_m} = 1 - \mathbf{w}_m^T \mathbf{w}_m = 0 \quad (88)$$

A partir da equação 87, tem-se, portanto:

$$\mathbf{S}\mathbf{w}_m = \lambda_m \mathbf{u}\mathbf{w}_m \quad (89)$$

E da equação 88, retoma-se a condição de ortonormalidade. Percebe-se que as componentes principais são que os autovetores \mathbf{w}_m da matriz de covariância \mathbf{S} , cuja magnitude são os autovalores λ_m . A representação dessas componentes pode ser vista na Figura 33, onde $m = 2$. Nela, observa-se os autovetores indicando as direções onde ocorre maior dispersão de um conjunto dados; seus módulos dependem da magnitude de seus autovalores.

Figura 33 – Representação das componentes principais.

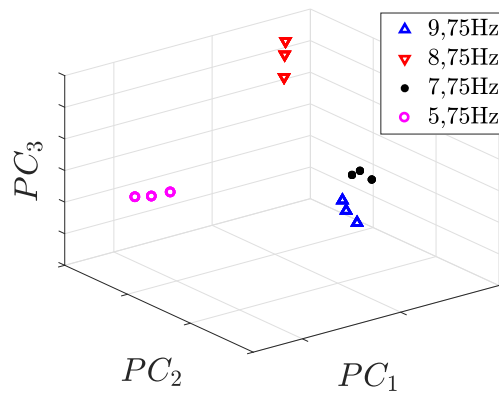


Fonte: Elaborado pelo autor.

Como visto anteriormente, ao se projetar os dados originais nas componentes principais, obtém-se um subespaço de dimensão reduzida. Esse fato é especialmente útil para aplicações práticas da PCA: se houver um número de conjuntos quaisquer de dados, cada um representativo de um estado ou condição específica de um fenômeno qualquer, por exemplo, a partir de sua matriz geral de covariância, é possível não só reduzir a quantidade de amostras, mas também separar-se espacialmente os conjuntos. A Figura 34 apresenta quatro conjuntos de dados de sinais SSVEP — com 8 eletrodos, representativos da região occipital, córtex visual, salientado na Figura 11, 1000 amostras por eletrodo,

por rodada, de um total de 20 rodadas — obtidos a partir de um banco de dados de domínio público (ZHANG et al., 2014), relativos a cada uma das frequências de estímulo do banco, projetados no subespaço constituído por 3 componentes principais; os 3 eletrodos mostrados a seguir apresentaram maior distância entre si dentre os 8 possíveis, e são provenientes de uma das rodadas. Percebe-se que cada ponto no subespaço representa um eletrodo, e que a distância entre os grupos é, de fato, considerável. Uma discussão mais detalhada sobre a disposição dos eletrodos nas componentes principais é apresentada na seção de resultados.

Figura 34 – Redução de dimensionalidade e projeção de dados SSVEP em 3 componentes principais.



Fonte: Elaborado pelo autor.

Não obstante, apesar de separar convenientemente os dados, a PCA em si não é um classificador; é necessário, portanto, introduzir-se uma métrica a fim de, ao se obter um novo sinal e projetá-lo no subespaço das componentes principais, realizar a classificação. Métricas popularmente aplicadas incluem: Euclidiana, *cityblock*, Minkowski, Chebyshev e Mahalanobis (OIKONOMOU et al., 2016). No presente trabalho, utiliza-se a distância Euclidiana e a de Mahalanobis, definidas, respectivamente, abaixo:

$$d_e = \sqrt{(\mathbf{p}_{cp} - \mathbf{q}_t)(\mathbf{p}_{cp} - \mathbf{q}_t)^T} \quad (90)$$

$$d_m = \sqrt{(\mathbf{p}_{cp} - \mathbf{q}_t)\mathbf{S}^{-1}(\mathbf{p}_{cp} - \mathbf{q}_t)^T} \quad (91)$$

onde os subíndices cp e t representam as coordenadas dos vetores de projeções dos dados originais e do sinal de teste, respectivamente, no subespaço de componentes principais, e \mathbf{S} é a matriz de covariância. Nota-se que a distância Euclidiana é, dessa forma, um caso particular da distância de Mahalanobis, onde \mathbf{S} é igual à identidade. Assim, d_m é mais geral, incluindo informações sobre a dispersão dos dados. Optou-se por essas métricas

devido ao fato de, além de ser possível compará-las nesse sentido, serem comumente utilizadas na literatura (EVA; LAZAR, 2015; KWAK; MÜLLER; LEE, 2017; CHEN et al., 2017; ISA et al., 2017). Por fim, o menor valor de d_e ou d_m corresponde ao grupo ao qual o sinal de teste pertence, concluindo, assim, a classificação. Também, além das métricas de distância, optou-se por aplicar um terceiro classificador baseado no conceito de probabilidade condicional, conhecido como Naïve-Bayes. Retomando a equação 36 e considerando que se deseja verificar a probabilidade de uma variável qualquer \mathbf{x} pertencer à classe C_k ($k = 1, 2, \dots, n$), tem-se:

$$p(C_k|\mathbf{x}) = \frac{p(C_k|\mathbf{x})p(C_k)}{p(\mathbf{x})} \quad (92)$$

Suponho que \mathbf{x} seja o vetor de características de interesse (no caso da PCA, essas características são as componentes principais), tem-se que $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$, onde m é a quantidade de componentes principais consideradas. Substituindo \mathbf{x} na equação 92, tem-se:

$$p(C_k|x_1, x_2, \dots, x_m) = \frac{p(C_k|x_1, x_2, \dots, x_m)p(C_k)}{p(x_1, x_2, \dots, x_m)} \quad (93)$$

Essa equação pode ser reescrita considerando-se a regra da cadeia, de modo que (DUDA; HART; STORK, 2012):

$$p(C_k|x_1, x_2, \dots, x_m) = \frac{p(x_1|x_2, \dots, x_m, C_k)p(x_2|x_3, \dots, x_m, C_k)\dots p(x_m|C_k)p(C_k)}{p(x_1, x_2, \dots, x_m)} \quad (94)$$

que é uma generalização da equação de probabilidade condicional. O classificador de Naïve-Bayes assume, desta forma, duas hipóteses: i) o denominador da equação 94 apenas normaliza o resultado, de modo que o numerador, por conter as classes, retém a informação mais importante; ii) as características são condicionalmente independentes (DUDA; HART; STORK, 2012). Assim, re-escreve-se a equação acima como:

$$p(C_k|x_1, x_2, \dots, x_m) \propto p(x_1|C_k)p(x_2|C_k)\dots p(x_m|C_k)p(C_k) = p(C_k) \prod_i^m p(x_i|C_k) \quad (95)$$

Desta forma, para se verificar a probabilidade das características pertencerem à classe C_k , tem-se o classificador:

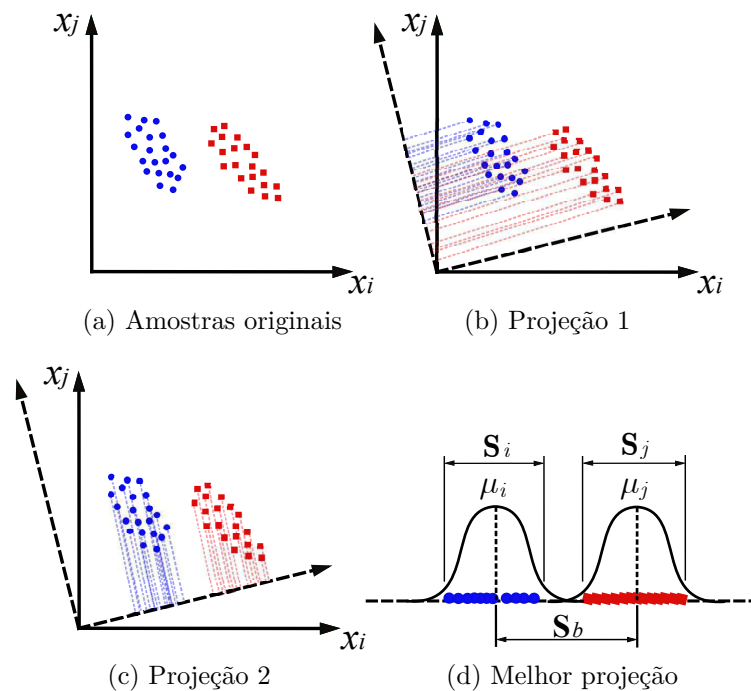
$$\hat{y} = \underset{k \in \{1, \dots, n\}}{\operatorname{argmax}} \left\{ p(C_k) \prod_i^m p(x_i|C_k) \right\} \quad (96)$$

indicando que o maior valor de \hat{y} indica a qual classe pertence o grupo de características. A seguir, é apresentada a próxima técnica de classificação estatística: Análise de Discriminante Linear/Quadrático (*Linear/Quadratic Discriminant Analysis*).

4.2.2 Análise de Discriminante Linear/Quadrático (LDA/QDA)

O conceito dessa técnica, introduzida por Fisher (1936), é similar a de componentes principais. Seu propósito é, da mesma forma, reduzir a dimensão dos dados. A principal diferença consiste em procurar as direções onde a projeção dos dados fornece a melhor separabilidade. Considerando que haja, por exemplo, duas séries $\mathbf{x}_i = \{x_1, x_2, \dots, x_P\}$ e $\mathbf{x}_j = \{x_1, x_2, \dots, x_P\}$ pertencentes a classes distintas i e j , procura-se uma projeção ótima $\mathbf{w}^T \mathbf{x}$ onde os dados apresentem máxima distância (\mathbf{S}_b) entre suas médias e mínima dispersão das amostras individualmente ($\mathbf{S}_i, \mathbf{S}_j$). Esse efeito é mostrado na Figura 35.

Figura 35 – Redução de dimensionalidade e projeção de dados a partir do discriminante linear de Fisher.



Fonte: Elaborado pelo autor.

Dessa forma, as médias inicialmente definidas como:

$$\mu_i = \frac{1}{P} \sum_{i=1}^p x_i, \quad x_i \in i \quad (97)$$

$$\mu_j = \frac{1}{P} \sum_{j=1}^p x_j, \quad x_j \in j \quad (98)$$

tornam-se:

$$\tilde{\mu}_i = \frac{1}{P} \mathbf{w}^T \sum_{i=1}^p x_i = \mathbf{w}^T \mu_i, \quad x_i \in i \quad (99)$$

$$\tilde{\mu}_j = \frac{1}{P} \mathbf{w}^T \sum_{j=1}^p x_j = \mathbf{w}^T \mu_j, \quad x_j \in j \quad (100)$$

e suas variâncias individuais originais:

$$\mathbf{S}_i^2 = \frac{1}{P} \sum_{i=1}^p (x_i - \mu_i)(x_i - \mu_i)^T, \quad x_i \in i \quad (101)$$

$$\mathbf{S}_j^2 = \frac{1}{P} \sum_{j=1}^p (x_j - \mu_j)(x_j - \mu_j)^T, \quad x_j \in j \quad (102)$$

tornam-se:

$$\tilde{\mathbf{S}}_i^2 = \frac{1}{P} \sum_{i=1}^p (\mathbf{w}^T x_i - \mathbf{w}^T \tilde{\mu}_i)(\mathbf{w}^T x_i - \mathbf{w}^T \tilde{\mu}_i)^T = \mathbf{w}^T \mathbf{S}_i \mathbf{w}, \quad x_i \in i \quad (103)$$

$$\tilde{\mathbf{S}}_j^2 = \frac{1}{P} \sum_{j=1}^p (\mathbf{w}^T x_j - \mathbf{w}^T \tilde{\mu}_j)(\mathbf{w}^T x_j - \mathbf{w}^T \tilde{\mu}_j)^T = \mathbf{w}^T \mathbf{S}_j \mathbf{w}, \quad x_j \in j \quad (104)$$

O discriminante linear de Fisher é definido, portanto, da seguinte forma (DUDA; HART; STORK, 2012):

$$J(\mathbf{w}) = \frac{|\tilde{\mu}_i - \tilde{\mu}_j|^2}{\tilde{\mathbf{S}}_i^2 + \tilde{\mathbf{S}}_j^2} \quad (105)$$

Alternativamente, o numerador da equação 105 pode ser escrito como:

$$(\tilde{\mu}_i - \tilde{\mu}_j)^2 = (\mathbf{w}^T \mu_i - \mathbf{w}^T \mu_j)^2 = \mathbf{w}^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \mathbf{w} \quad (106)$$

Define-se, assim, $\mathbf{S}_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$, conhecida como matriz de dispersão inter-classe (*between-class scatter matrix*), cuja projeção ótima é (YU; YANG, 2001):

$$\tilde{\mathbf{S}}_b = \mathbf{w}^T \mathbf{S}_b \mathbf{w} \quad (107)$$

O denominador, por sua vez, pode ser expandido como:

$$\tilde{\mathbf{S}}_i^2 + \tilde{\mathbf{S}}_j^2 = \mathbf{w}^T \mathbf{S}_i \mathbf{w} + \mathbf{w}^T \mathbf{S}_j \mathbf{w} = \mathbf{w}^T (\mathbf{S}_i + \mathbf{S}_j) \mathbf{w} \quad (108)$$

Define-se, assim, $\mathbf{S}_w = \mathbf{S}_i + \mathbf{S}_j$, conhecida como matriz de dispersão intra-classe (*within-*

class scatter matrix), cuja projeção ótima é (YU; YANG, 2001):

$$\tilde{\mathbf{S}}_w = \mathbf{w}^T \mathbf{S}_w \mathbf{w} \quad (109)$$

Por fim, o discriminante linear de Fisher pode ser escrito, também, da seguinte maneira:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \quad (110)$$

e a maximização desse discriminante ocorre quando:

$$J(\mathbf{w})\mathbf{w} = \mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w} \quad (111)$$

ou seja, semelhante ao coeficiente de Pearson, esse problema é caracterizado por autovetores \mathbf{w} e autovalores da matriz $\mathbf{S}_w^{-1} \mathbf{S}_b$. Portanto:

$$\hat{J}(\mathbf{w}) = \underset{w}{argmax}(\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w}) \quad (112)$$

Esse resultado também pode ser obtido diferenciando-se a equação 110 com respeito a w e igualando a zero:

$$\frac{J(\mathbf{w})}{d\mathbf{w}} = (\mathbf{w}^T \mathbf{S}_w \mathbf{w}) \frac{d(\mathbf{w}^T \mathbf{S}_b \mathbf{w})}{d\mathbf{w}} - (\mathbf{w}^T \mathbf{S}_b \mathbf{w}) \frac{d(\mathbf{w}^T \mathbf{S}_w \mathbf{w})}{d\mathbf{w}} = 0 \quad (113)$$

alternativamente,

$$\left(\frac{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \right) \mathbf{S}_b \mathbf{w} - \left(\frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \right) \mathbf{S}_w \mathbf{w} = 0 \quad (114)$$

Portanto:

$$J(\mathbf{w})\mathbf{S}_w \mathbf{w} = \mathbf{S}_b \mathbf{w} \quad (115)$$

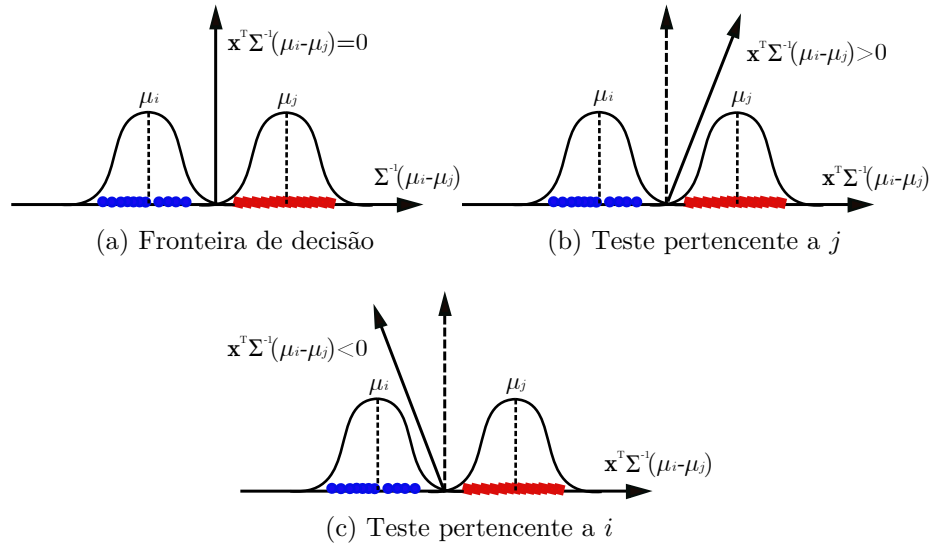
Ao isolar-se $J(\mathbf{w})$ realizando-se a inversa de \mathbf{S}_w , tem-se a mesma expressão que a equação 111. O maior autovetor obtido a partir de $\hat{J}(\mathbf{w})$ caracteriza-se pela projeção que melhor separa as distribuições dos dados. Quanto à classificação, por fim, pode-se estabelecer uma fronteira de decisão baseando-se na projeção $(\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w})$, da seguinte forma (DUDA; HART; STORK, 2012):

$$\tilde{\mathbf{w}}^T \mathbf{z} + w_0 = 0 \quad (116)$$

onde \mathbf{z} é o vetor de entrada, w_0 é uma constante a ser encontrada, que depende das propriedades estatísticas do conjunto (DUDA; HART; STORK, 2012), e $\tilde{\mathbf{w}} = \Sigma^{-1}(\mu_i - \mu_j)$,

com Σ representando a matriz de covariância (ou de dispersão intra-classe). A representação gráfica da classificação é mostrada na Figura 36. Observa-se que, considerando uma função de densidade gaussiana, por exemplo, para o caso particular de $w_0 = 0$, se o produto escalar entre o vetor de entrada e a projeção ótima for maior que zero, o vetor teste pertence à classe j ; caso o produto seja menor que zero, pertence à classe i .

Figura 36 – Classificação a partir do discriminante linear de Fisher.



Fonte: Elaborado pelo autor.

Para se determinar w_0 , é necessário estabelecer-se qual a densidade de probabilidade do conjunto de dados. Considerando a função gaussiana (equação 10), a mais frequentemente utilizada, pode-se defini-la, para as classes i e j , como (AGGARWAL, 2015):

$$f_i(\mathbf{x}) = \frac{1}{2\pi^{-p/2}|\Sigma_i|^{-1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i (\mathbf{x}-\mu_i)} \quad (117)$$

$$f_j(\mathbf{x}) = \frac{1}{2\pi^{-p/2}|\Sigma_j|^{-1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j (\mathbf{x}-\mu_j)} \quad (118)$$

onde p é a dimensão do conjunto de dados. A classificação é dada utilizando-se a função de densidade e a probabilidade condicional, como visto na equação 36, de forma que:

$$P(\mathbf{x} \in C|\mathbf{x}) = \frac{f(\mathbf{x}|\mathbf{x} \in C_i)P(\mathbf{x} \in C_i)}{f(\mathbf{x})} \propto f_i(\mathbf{x})\pi_i \quad (119)$$

onde $f_i(\mathbf{x})$ é a função verissimilhança, e π_i é a probabilidade a priori para o conjunto i ; na prática, essa probabilidade se dá pelo número de amostras de um conjunto dividido pelo total de amostras. Aplicando-se a propriedade do logaritmo nas equações 117 e 118,

e substituindo-as na equação 119, tem-se (FRIEDMAN, 1989):

$$\delta_i(\mathbf{x}) = -\ln(\pi_i) + \frac{1}{2} \log|\Sigma_i| + \frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma^{-1}(\mathbf{x} - \mu_i) \quad (120)$$

$$\delta_j(\mathbf{x}) = -\ln(\pi_j) + \frac{1}{2} \log|\Sigma_j| + \frac{1}{2}(\mathbf{x} - \mu_j)^T \Sigma^{-1}(\mathbf{x} - \mu_j) \quad (121)$$

Nota-se que o terceiro termo de cada equação corresponde à distância de Mahalanobis, podendo ser reescrita na forma de quadrado perfeito:

$$(\mathbf{x} - \mu_h)^T \Sigma_h^{-1}(\mathbf{x} - \mu_h) = \mu_h^T \Sigma_h^{-1} \mu_h - 2\mu_h^T \Sigma_h^{-1} \mathbf{x} + \mathbf{x}^T \Sigma_h^{-1} \mathbf{x} \quad (122)$$

onde h é uma classe qualquer, seja i ou j . A fronteira se dá quando $\delta_i(\mathbf{x}) = \delta_j(\mathbf{x})$, ou, alternativamente, $\delta_i(\mathbf{x}) - \delta_j(\mathbf{x}) = 0$. Percebe-se que, ao se juntar cada parcela da equação, o termo $\mathbf{x}^T \Sigma^{-1} \mathbf{x}$ faz com que a equação final de fronteira seja de segunda ordem:

$$(\mathbf{a}_i - \mathbf{a}_j) + (\mathbf{b}_i^T - \mathbf{b}_j^T)\mathbf{x} + \mathbf{x}^T(\mathbf{c}_i^T - \mathbf{c}_j^T)\mathbf{x} = 0 \quad (123)$$

e o termo w_0 , por fim, é: $w_0 = (\mathbf{a}_i - \mathbf{a}_j)$. Esse fato caracteriza a classificação como Análise de Discriminante Quadrático (*Quadratic Discriminant Analysis*). Um caso particular, comumente utilizado, dá-se para quando as matrizes de covariância para todas as classes são iguais (homocedasticidade), de forma que (FRIEDMAN, 1989):

$$\Sigma_i = \Sigma_j = \Sigma \quad (124)$$

Caso essa hipótese seja feita, a classificação passa a Análise de Discriminante Linear (*Linear Discriminant Analysis*), pois os termos quadrados se cancelam. A equação de fronteira será, por conseguinte:

$$\ln(\pi_i) - \frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i) = \ln(\pi_j) - \frac{1}{2}(\mathbf{x} - \mu_j)^T \Sigma_j^{-1}(\mathbf{x} - \mu_j) \quad (125)$$

e a classificação é dada com:

$$\mathbf{x}^T \Sigma^{-1}(\mu_i - \mu_j) = \ln\left(\frac{\pi_i}{\pi_j}\right) - \frac{1}{2}\left(\mu_i^T \Sigma^{-1} \mu_i - \mu_j^T \Sigma^{-1} \mu_j\right) \quad (126)$$

Dessa forma, uma nova amostra pertence ao conjunto i caso:

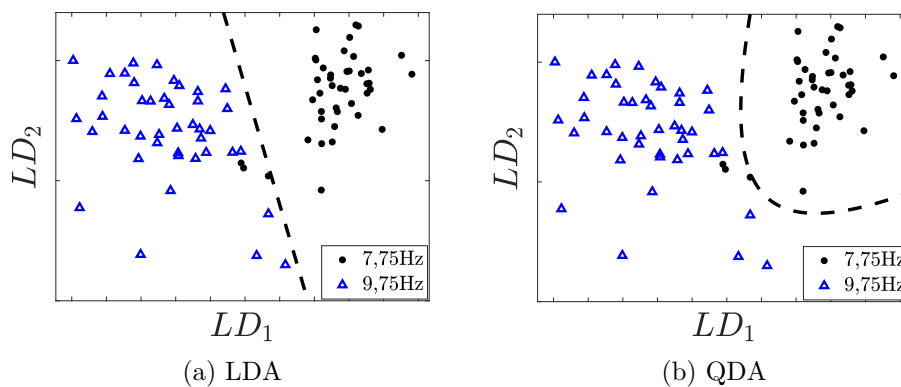
$$\mathbf{x}^T \Sigma^{-1}(\mu_i - \mu_j) < \ln\left(\frac{\pi_i}{\pi_j}\right) - \frac{1}{2}\left(\mu_i^T \Sigma^{-1} \mu_i - \mu_j^T \Sigma^{-1} \mu_j\right) \quad (127)$$

e ao conjunto j , caso:

$$\mathbf{x}^T \Sigma^{-1} (\mu_i - \mu_j) > \ln\left(\frac{\pi_i}{\pi_j}\right) - \frac{1}{2} \left(\mu_i^T \Sigma^{-1} \mu_i - \mu_j^T \Sigma^{-1} \mu_j \right) \quad (128)$$

Observa-se que essa expressão é a mesma obtida na equação 116, porém mais geral. Assim, quando se considera a hipótese de homocedasticidade, a QDA se torna o próprio coeficiente de Fisher. A Figura 37 ilustra a diferença na classificação entre a QDA e a LDA para duas frequências de estímulo distintas, feita a partir do mesmo banco de dados disponibilizado por Zhang et al. (2014) e aplicado para a PCA. Nota-se que a fronteira da QDA é mais flexível, fornecendo uma classificação mais refinada. Em todo caso, a classificação por LDA é, em muitos casos, suficiente; mesmo a distância Euclidiana ou de Mahalanobis podem ser utilizada para se realizar a análise.

Figura 37 – Comparação entre LDA e QDA para 9,75 Hz e 7,75 Hz.



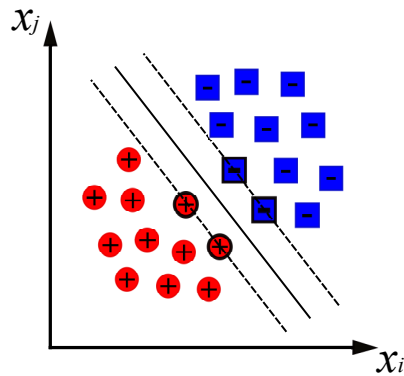
Fonte: Elaborado pelo autor.

A seguir, na próxima seção, apresenta-se a última técnica entre as metodologias alternativas de classificação apresentadas neste trabalho: *Support Vector Machine*.

4.2.3 Máquina de Vetores de Suporte (SVM)

O conceito de *Support Vector Machines* envolve a separação (e, conseqüentemente, classificação) de dois conjuntos de dados a partir da busca de um hiperplano ótimo cujas margens — isto é, distâncias entre as amostras de cada conjunto mais próximas entre si e a fronteira de decisão — sejam maximizadas (BISHOP, 2006). Considerando dois conjuntos de variáveis aleatórias, uma com amostras positivas, e outra com amostras negativas; a fronteira de decisão encontra-se equidistante das amostras positiva e negativa, respectivamente, mais próximas uma da outra, chamadas de vetores suporte, como mostra a Figura 38.

Figura 38 – Ilustração do conceito de SVM.



Fonte: Elaborado pelo autor.

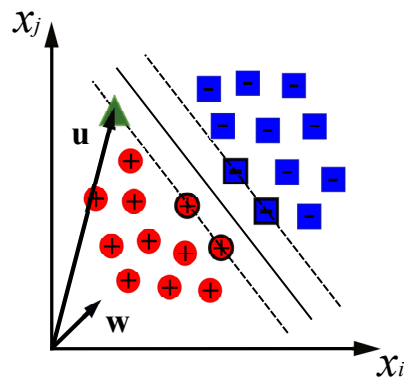
Com a introdução de uma nova amostra (classe previamente desconhecida) \mathbf{u} , por exemplo, a classificação se dá a partir de sua projeção no vetor \mathbf{w} (Figura 39), perpendicular ao hiperplano de separação, e verificação a partir da seguinte regra de decisão:

$$\mathbf{w}^T \mathbf{u} \geq c \quad (129)$$

onde c é uma constante, valor de referência que delimita se uma amostra é positiva ou negativa. Alternativamente, pode-se escrever essa equação como:

$$\mathbf{w}^T \mathbf{u} + b \geq 0 \quad (130)$$

Figura 39 – Introdução de amostra cuja classe é desconhecida.



Fonte: Elaborado pelo autor.

Desta forma, também é possível estabelecer que:

$$\mathbf{x}_i^T \mathbf{w} + b \geq +1, \quad y_i = +1 \quad (131)$$

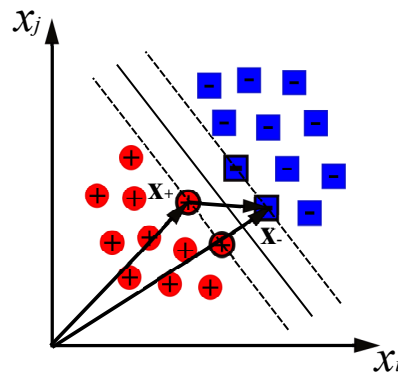
$$\mathbf{x}_i^T \mathbf{w} + b \leq -1, \quad y_i = -1 \quad (132)$$

onde y_i corresponde a cada classe. Ou, de modo a reduzir o número de equações,

$$y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1 \geq 0, \quad \forall i \quad (133)$$

A fim de se realizar a classificação, basta definir, portanto, os valores de \mathbf{w} e b . Para tal, é necessário mensurar o valor entre as margens, de modo que se possa maximizá-lo; esse processo é feito considerando-se a diferença entre os vetores suporte, como mostra a Figura 40.

Figura 40 – Definição da região entre margens para SVM.



Fonte: Elaborado pelo autor.

A projeção dessa diferença no vetor \mathbf{w} resulta na medida real entre as margens:

$$f(\mathbf{w}) = (\mathbf{x}_+ - \mathbf{x}_-)^T \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (134)$$

Isolando-se os termos \mathbf{x}_+ e \mathbf{x}_- na equação 133, tem-se:

$$\mathbf{x}_+ = \frac{1 - b}{\mathbf{w}} \quad (135)$$

$$\mathbf{x}_- = \frac{-1 - b}{\mathbf{w}} \quad (136)$$

Sendo equações que, combinadas a $f(\mathbf{w})$, tornam-se:

$$f(\mathbf{w}) = \frac{2}{\|\mathbf{w}\|} \quad (137)$$

Logo, para que se obtenha a maior distância entre as margens, é preciso que $f(\mathbf{w})$ seja o valor máximo. Em outras palavras, $\|\mathbf{w}\|$ deve ser o mínimo. Alternativamente, portanto, a distância pode ser escrita como uma função-objetivo de minimização:

$$f(\mathbf{w}) = \min \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\} \quad (138)$$

Nota-se que há uma adequação, considerada apenas para facilitar os cálculos seguintes. Assim como visto para a PCA, por exemplo, esse problema é abordado a partir dos multiplicadores de Lagrange, onde a função objetivo é $f(\mathbf{w})$, e a restrição, $g(\mathbf{w}) = y_i(\mathbf{x}_i \bullet \mathbf{w} + b) - 1$. Consequentemente, aplicando-se a Lagrangiana:

$$L(\mathbf{w}, \lambda_i, b) = f(\mathbf{w}) - \lambda_i g(\mathbf{w}) \quad (139)$$

onde busca-se por pontos em que ocorra:

$$\nabla L(\mathbf{w}, \lambda_i, b) = 0 \quad (140)$$

Assim, a Lagrangiana se torna:

$$L(\mathbf{w}, \lambda_i, b) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \lambda_i [y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1] \quad (141)$$

A partir do gradiente $\nabla L(w, \lambda_i)$, pode-se estabelecer novamente as condições KKT. No contexto de SVM, tem-se que:

$$\frac{\partial L(\mathbf{w}, \lambda_i, b)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i = 0 \quad (142)$$

$$\frac{\partial L(\mathbf{w}, \lambda_i, b)}{\partial \lambda_i} = y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1 = 0 \quad (143)$$

$$\frac{\partial L(\mathbf{w}, \lambda_i, b)}{\partial b} = \sum_{i=1}^N \lambda_i y_i = 0 \quad (144)$$

Esse conjunto de equações, quando resolvido simultaneamente, resulta nos valores para \mathbf{w} e b . Observa-se que $\mathbf{w} = \sum_{i=1}^N \lambda_i y_i \mathbf{x}_i$, ou seja, uma combinação linear das amostras. Assim, a Lagrangiana pode ser escrita como:

$$L(\mathbf{w}, \lambda_i, b) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^N \lambda_i \quad (145)$$

Finalmente, a regra de decisão se torna:

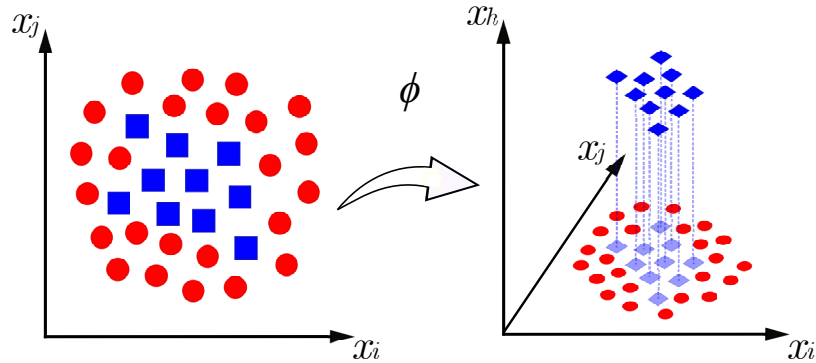
$$\sum_{i=1}^N \lambda_i y_i \mathbf{x}_i^T \mathbf{u} + b \geq 0, \quad \mathbf{u} \in i = + \quad (146)$$

$$\sum_{i=1}^N \lambda_i y_i \mathbf{x}_i^T \mathbf{u} + b < 0, \quad \mathbf{u} \in i = - \quad (147)$$

A fronteira de decisão, nesse caso, baseia-se no fato das amostras serem linearmente

separáveis. Caso isso não ocorra, pode-se aplicar funções *kernel* em SVM, de modo a realizar uma transformação para um novo espaço onde as amostras, por fim, são separáveis, como apresentado pela Figura 41.

Figura 41 – Transformação para novo espaço aplicando uma função *kernel* em SVM.



Fonte: Elaborado pelo autor.

Para o caso linear, a função *kernel* é definida como (BISHOP, 2006):

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (148)$$

Para os casos não lineares, funções *kernel* comumente usadas são a exponencial (também chamada de Gaussiana, ou *Radial Basis Function*), a polinomial, e a sigmoide, respectivamente (BISHOP, 2006; YEKKEHKHANY et al., 2014):

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\gamma}\right) \quad (149)$$

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + d)^n \quad (150)$$

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\mathbf{x}_i^T \mathbf{x}_j + d) \quad (151)$$

onde n , d e γ são constantes e $\exp()$ e $\tanh()$ são as funções exponencial e tangente hiperbólico, respectivamente. A variável d normalmente é escolhida para ser igual a 0 ou 1, n está ligado ao grau de controle e flexibilidade do classificador resultante considerando *kernel* polinomial, e γ desenvolve um papel similar, porém para o *kernel* Gaussiano, controlando seu comprimento (BEN-HUR et al., 2008). Em termos de implementação computacional, o SVM é originalmente desenvolvido para separar classes binárias (Cortes e Vapnik (1995), apud Ma e Guo (2014)), e deve-se adotar algumas estratégias para quando há mais de duas delas:

- abordagem um versus o restante (1VR) (TAX; DUIN, 1999): considera uma classe versus as demais; sua vantagem é reduzir o número de classes e simplificar a aplicação da técnica; a desvantagem, por sua vez, é que, embora torne o problema de otimização novamente binário, pode causar um problema de desbalanceamento do número de amostras, especialmente quando o número total de classes é grande, ignorando possíveis classificações corretas do grupo menor;
- abordagem um versus um (1V1) (Kreßel (1999), apud Ma e Guo (2014)): classifica as classes em pares; se existem, por exemplo, k classes, há $k(k-1)/2$ classificadores binários individuais. Essa abordagem considera como grupo correto aquele que obtém maior número de votos total;
- abordagem direta (Bredensteiner e Bennett (1999), Crammer e Singer (2002), Guermur (2002), Khan, Awad e Thuraisingham (2007), apud Ma e Guo (2014)): aplica um modelo multiclasse, tornando o problema de otimização apenas um; a vantagem é a sofisticação e generalização do método, porém a um custo de complexidade computacional relativamente alto.

Na seção de resultados, aplica-se a abordagem 1V1. O conceito de SVM conclui a etapa de apresentação de técnicas de classificação de dados. As seções seguintes, por fim, encerram o capítulo 3 com dois enfoques: i) função de densidade de probabilidade gaussiana, bem como uma análise para se mensurar, visual e matematicamente, o quão próxima dela uma função qualquer é, isto é, se pode-se ou não assumir como Gaussiana uma variável aleatória de interesse; esse estudo é necessário não só pela importância em si da Gaussiana, mas também porque algumas técnicas, como visto, consideram tal densidade de probabilidade (essa proposição é assumida para todos os sinais SSVEP deste trabalho); ii) validação de modelos de classificação estatística para se estimar erros de predição, etapa também considerada relevante no contexto de tratamento de dados.

4.3 CRITÉRIOS DE ADERÊNCIA À NORMALIDADE

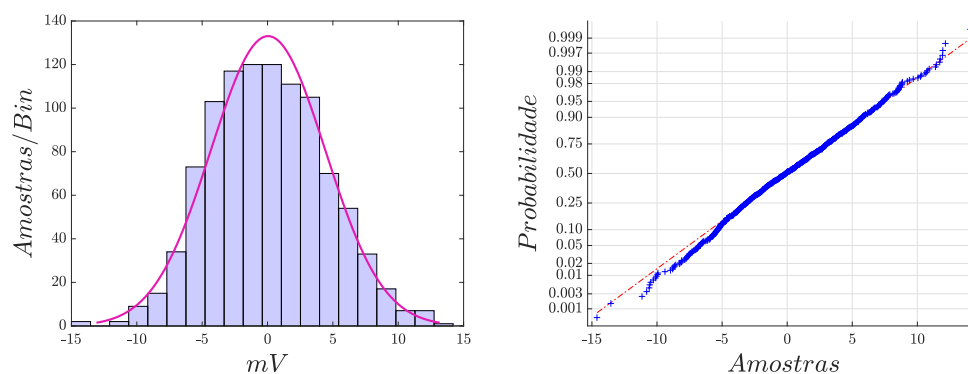
De acordo com Curran-Everett e Benos (2004), apud Ghasemi e Zahediasl (2012), os erros estatísticos são comuns na literatura científica, e cerca de 50% dos trabalhos publicados apresentam ao menos um erro. Muitos dos procedimentos adotados assumem que os dados analisados seguem uma distribuição normal, ou gaussiana, e nem sempre essa suposição pode ser levada em consideração. Baseado nesse contexto, existem meios para se mensurar o quão próxima da gaussiana uma distribuição de dados é, chamados de critérios de

aderência à normalidade. Dentre os existentes, pode-se citar: Anderson-Darling, Cramer-von Mises, Kolmogorov-Smirnov, Shapiro-Wilk, D'Agostino-Pearson, Jarque-Bera, Lilliefors, chi-quadrada, bem como recursos gráficos, como o histograma, box plot, e *normal plot*, por exemplo (YAP; SIM, 2011). O presente trabalho dá enfoque a duas abordagens: gráfica, a partir da análise de histogramas e curvas de distribuição, e formal, pela aplicação do teste de Shapiro-Wilk, considerado como a melhor escolha dentre os outros critérios formais (YAP; SIM, 2011; GHASEMI; ZAHEDIASL, 2012; RAZALI; WAH et al., 2011).

4.3.1 Critério Visual

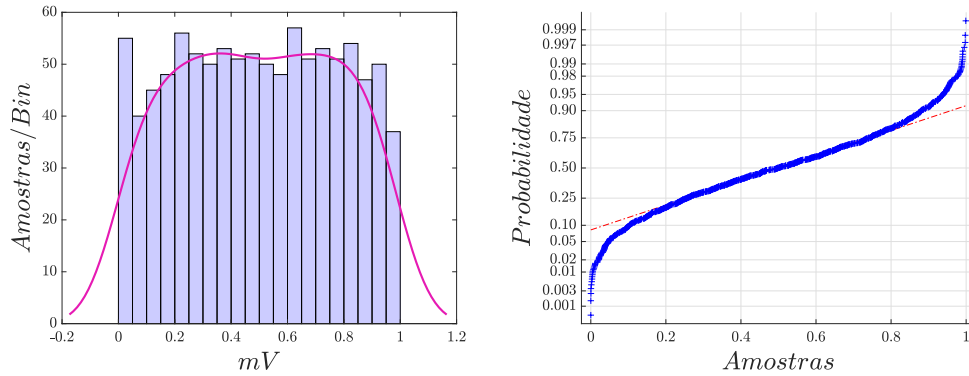
Para a primeira abordagem, tem-se que os histogramas estão relacionados à função de densidade de probabilidade, dividindo os dados em partições (*bins*) de acordo com o número de amostras contidas em cada faixa de valores Δx da amplitude do sinal $x(t)$; os *normal plots*, por sua vez, relacionam-se à função de distribuição de probabilidade, mapeando em valores de 0 a 1 as probabilidades de cada uma dessas partições. Considere duas séries de dados cuja distribuição é desconhecida a priori. A Figura 42 mostra os resultados dos histogramas obtidos para uma das séries, seja na forma de histograma, seja na forma de *normal plot*. Nota-se, para ambos os gráficos, uma aproximação expressiva em relação à gaussiana. O mesmo não ocorre para os gráficos da segunda série de dados (Figura 43), que divergem significativamente da distribuição gaussiana, tendendo, a princípio, a uma distribuição uniforme.

Figura 42 – Teste visual de aderência à normalidade para a primeira série de dados.



Fonte: Elaborado pelo autor.

Figura 43 – Teste visual de aderência à normalidade para a segunda série de dados.



Fonte: Elaborado pelo autor.

4.3.2 Critério Formal

O teste de Shapiro-Wilk, proposto por Shapiro e Wilk (1965), consistem em uma metodologia formal para testar a aderência à normalidade. Essencialmente, baseia-se no fato de que uma variável $Y \sim N(\mu, \sigma^2)$ (isto é, de média μ e variância σ qualquer) pode ser escrita na seguinte forma: $Y = \mu + \sigma X$, onde $X \sim N(0, 1)$ (YAP; SIM, 2011). Considerando $X_1 \leq X_2 \leq \dots \leq X_p$ uma amostra aleatória de tamanho p em ordem ascendente ($\mu = 0, \sigma = 1$), $\mathbf{m}' = (m_1, m_2, \dots, m_p)$ o vetor de valores esperados da estatística de ordem padrão normal, $\mathbf{V} = (v_{ij})$ a matriz de covariância $n \times n$ dessas estatísticas de ordem, e $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_n)$ o vetor ordenado das observações amostrais; se existe aderência à normalidade, pode-se escrever $Y_{(i)}$ como (YAP; SIM, 2011):

$$Y_{(i)} = \mu + \sigma X_{(i)}, \quad (i = 1, 2, \dots, n) \quad (152)$$

e o teste de Shapiro-Wilk (SW) pode ser definido como:

$$SW = \frac{[\sum_{i=1}^n a_i Y_{(i)}]^2}{\sum_{i=1}^n n(Y_{(i)} - \bar{Y})^2} = \frac{(\sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i))^2}{\sum_{i=1}^n n(Y_{(i)} - \bar{Y})^2} \quad (153)$$

onde $n = 2k$, se n for par, e $n = 2k + 1$, se n for ímpar, e os chamados carregamentos a_i são dados por:

$$\mathbf{a}' = \mathbf{m}'\mathbf{V}^{-1}(\mathbf{m}'\mathbf{V}^{-1}\mathbf{m})^{-1/2} \quad (154)$$

podendo ser obtidos a partir de tabelas em Shapiro e Wilk (1965). O valor de SW varia entre 0 e 1, sendo que valores próximos da unidade indicam aderência à normalidade, e zero, o contrário; o critério varia de acordo com a confiabilidade desejada. Baseado nesse

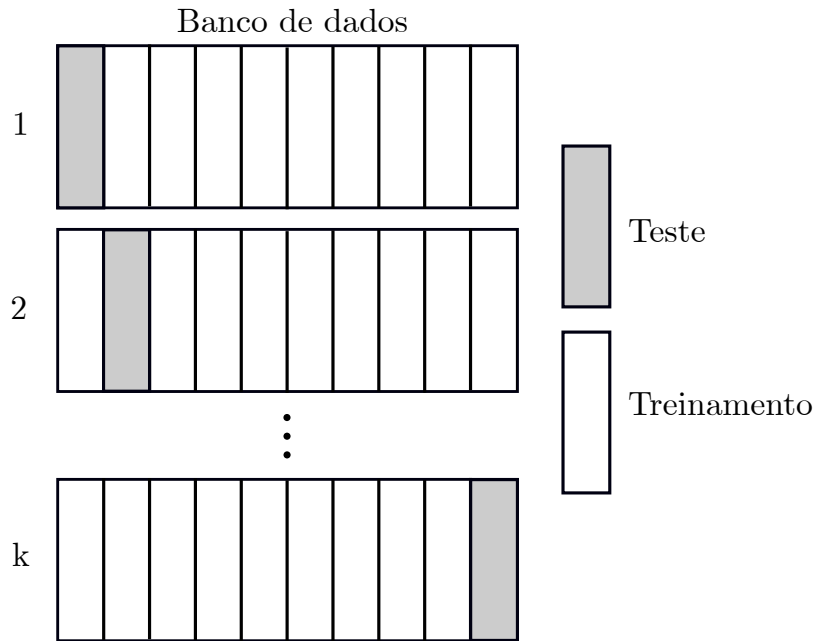
contexto, aplica-se o valor p , mínimo nível de significância a partir do qual pode-se rejeitar uma hipótese em questão baseado no resultado (BHATTACHARYA; HABTZGHI, 2002). A hipótese para o teste Shapiro-Wilk é sobre se considerar ou não uma distribuição como Gaussiana (hipótese H_0); assim, deve-se rejeitá-la caso $P_{H_0}(X \geq \alpha) = \alpha$; isto é, se x é o valor observado de uma variável aleatória X , considerando 95% de confiança, por exemplo, e a distribuição de X em H_0 é dada por P_0 , deve-se rejeitar a hipótese caso $p = 1 - P_0 < 0,05$ (BHATTACHARYA; HABTZGHI, 2002). Valores para p baseado nos valores de SW podem ser obtidos a partir de tabelas em Shapiro e Wilk (1965). Por fim, o teste de Shapiro-Wilk foi originalmente desenvolvido para $n \leq 50$; extensões desse teste para até $n \leq 5000$ podem ser vistas em Royston (1992) e Royston (1993). Os valores de SW e p para as distribuições das Figuras 42 e 43 são, respectivamente: $SW = 0,9972$ e $SW = 0,9487$; $p = 0,0794$ e $p = 0$. Como esperado, pode-se considerar a primeira distribuição como gaussiana, enquanto que essa hipótese deve ser rejeitada para a segunda.

4.4 VALIDAÇÃO CRUZADA

De acordo com Mahmood e Khan (2009), a validação cruzada (ou, do inglês, *cross-validation*), ferramenta padrão para se estimar erros de predição de modelos estatísticos, é uma técnica computacional de reamostragem intensiva utilizada com o intuito de avaliar modelos estatísticos, ou mesmo selecionar, entre aqueles que estão disponíveis, o modelo de regressão mais adequado. A validação cruzada pode ser considerada como um método eficaz de validação de modelo quando não é possível coletar novos dados sobre um fenômeno observado (SNEE, 1977). Dentre os tipos mais comuns, pode-se citar (KOHAVI et al., 1995; MAHMOOD; KHAN, 2009; FUSHIKI, 2011): *k-fold cross-validation* e *leave-one-out*. Neste trabalho, emprega-se a primeira dessas técnicas, pois esta pode ser preferível do ponto de vista computacional (FUSHIKI, 2011).

Na validação cruzada por *k-fold*, a partir do conjunto de dados originais D , separa-se k subconjuntos D_1, D_2, \dots, D_k mutuamente exclusivos de aproximadamente mesmo tamanho, chamados de *folds* (KOHAVI et al., 1995). Esses subconjuntos, em seguida, são categorizados em “treinamento” ou “teste”, sendo o menor subconjunto pertencente ao teste. Assim, são realizadas k iterações, cada qual retirando um subconjunto D_i qualquer, $i = 1, \dots, k$, e construindo o modelo de análise, separação ou classificação de dados com os $k - 1$ subconjuntos restantes. A Figura 44 apresenta uma ilustração do processo de validação cruzada.

Figura 44 – Ilustração do processo de validação cruzada por k -fold.



Fonte: Elaborado pelo autor.

A cada iteração, calcula-se um erro de predição ou porcentagem de acerto em classificação e, ao final do processo, faz-se uma média aritmética simples de todos os erros de predição/porcentagens de acerto encontrados. Desta forma, tem-se uma medida geral de desempenho do modelo, podendo-se aferir se este está bem construído. O erro total de predição é apresentado pela equação 155:

$$E_{total} = \frac{1}{k} \sum_{i=1}^k E_i \quad (155)$$

Outra ferramenta importante para melhor análise estatística chama-se curva ROC (do inglês, *Receiver Operating Characteristic*), que é abordada na seção seguinte.

4.5 CURVA ROC

Uma curva ROC (*Receiver Operating Characteristic*), cuja idealização deve-se à constatação de que uma simples porcentagem de acerto pode ser uma métrica insuficiente, consiste em uma técnica para se visualizar, organizar e selecionar classificadores baseado em seu desempenho (FAWCETT, 2006). Dadas duas classes i (positivos) e j (negativos) de dados, e um classificador estatístico capaz de alocar uma nova amostra a uma dessas classes (como no caso apresentado para a LDA/QDA ou SVM, por exemplo), há quatro possibilidades: a amostra positiva ser classificada, de fato, como positiva (o que é conhe-

cido como verdadeiro positivo, ou, do inglês, *true positive* – TP), a amostra positiva ser classificada como negativa (falso negativo, *false negative* – FN), a amostra negativa ser classificada como negativa (verdadeiro negativo, *true negative* – TN), e a amostra negativa ser classificada como positiva (falso positivo, *false positive* – FP) (FAWCETT, 2006). A partir dessas possibilidades, é possível construir uma tabela que ilustre o resultado da classificação, chamada da matriz de confusão, como mostra a Figura 45.

Figura 45 – Exemplo de uma matriz de confusão.

		Classes Verdadeiras	
		p	n
Classificação	p	Verdadeiro Positivo	Falso Negativo
	n	Falso Positivo	Verdadeiro Negativo

Fonte: Elaborado pelo autor.

A partir dessa matriz, estabelece-se relações de acurácia (acertos na classificação) e erros (“confusões” ocorridas na classificação); desta forma, pode-se obter mais informações sobre o desempenho do classificador em questão. Nesse contexto, estabelece-se relações conhecidas para a matriz de confusão: taxa de verdadeiros positivos (*true positive rate*, TPR) e taxa de falsos positivos (*false positive rate*, FPR), cujas expressões são apresentadas pelas equações 156 e 157, respectivamente (FAWCETT, 2006). Certamente, há outras relações possíveis, porém o presente trabalho dá enfoque para as citadas anteriormente, especialmente visando à construção de curvas ROC.

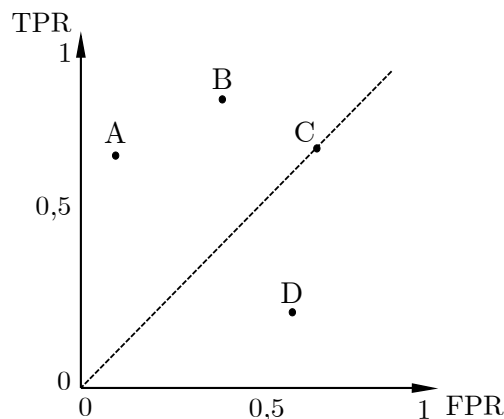
$$TPR = \frac{\text{Positivos corretamente classificados}}{\text{Total de positivos}} \quad (156)$$

$$FPR = \frac{\text{Negativos incorretamente classificados}}{\text{Total de negativos}} \quad (157)$$

Segundo Fawcett (2006), o plano conhecido como curva ROC, por sua vez, é obtido a partir das coordenadas FPR vs TPR , que mostra relações, portanto, de custo (falsos positivos) e benefícios (verdadeiros positivos) do classificador. Também, de acordo com o mesmo autor, um classificador discreto produz um ponto no plano FPR vs TPR , a partir

do qual se pode aferir se o classificador apresenta uma abordagem mais “conservadora” ou “liberal”. Por exemplo: se um ponto apresenta um valor alto em TPR e um baixo valor em FPR , diz-se que o classificador é mais conservador, pois realiza classificações positivas apenas se há uma forte evidência, de modo a reduzir erros de falsos positivos; se o ponto apresenta o mesmo valor de TPR , mas um valor mais alto de FPR , diz-se, por outro lado, que o classificador é mais “liberal”, pois realiza classificações positivas a partir de uma evidência mais fraca. Outra possibilidade é se o classificador apresentar $FPR = TPR$, conhecido como desempenho aleatório (*random performance*), o que implica que não se pode obter informações concretas sobre o classificador, pois ao mesmo tempo que apresenta uma alta taxa de acertos, o mesmo vale para a taxa de erros. Quando as coordenadas definem pontos no triângulo abaixo da classificação aleatória, diz-se que seu desempenho é pior que a própria reta definida por $FPR = TPR$. A Figura 46 apresenta pontos com as características mencionadas: A implica em um classificador mais conservador; B , outro mais liberal; C , classificação aleatória; D , classificação indesejável.

Figura 46 – Ilustração de uma curva ROC para valores discretos.

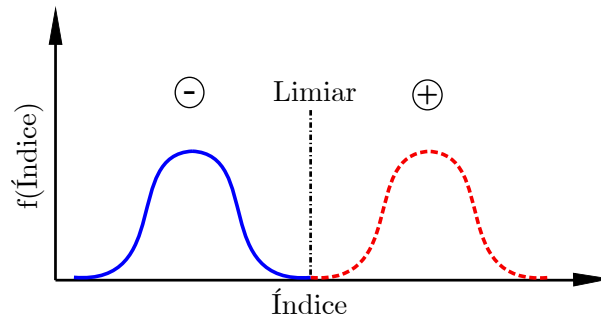


Fonte: Adaptado de Fawcett (2006).

Não obstante, há classificadores que não resultam em classificações binárias, mas sim em uma probabilidade, ou valor numérico associado, que representa as chances de uma amostra pertencer a uma determinada classe. A esses classificadores é associado um limite (*threshold*), a partir do qual se pode, novamente, produzir um classificador binário: se o resultado obtido é acima desse limite, classifica-se uma amostra à classe i ; caso contrário, à classe j (FAWCETT, 2006). Para clareza de entendimento, considere uma situação genérica onde se tenha duas distribuições de dados, uma referente à condição normal (positivo) de um fenômeno verificado, e outra à condição que apresente algum aspecto heteróclito (negativo), mapeadas numericamente por um índice qualquer, como apresentado na Figura 47. Nota-se que as distribuições são visivelmente separáveis, podendo-se,

portanto, classificar novas amostras facilmente.

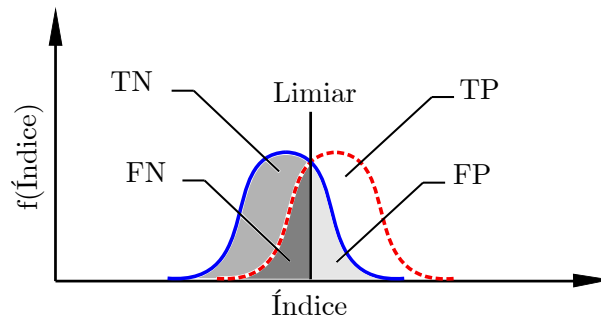
Figura 47 – Exemplo de distribuições de dados facilmente separáveis.



Fonte: Elaborado pelo autor.

Entretanto, problemas reais contam com distribuições menos visivelmente separáveis, como mostra a Figura 48, e com o propósito de se procurar a melhor configuração de separabilidade entre os dados, pode-se variar o limiar. Cada valor assumido por esse limiar implica em um ponto no plano ROC; assim, ao se estabelecer diversos valores, constrói-se, enfim, uma curva ROC.

Figura 48 – Exemplo de distribuições de dados mais próximas das reais.

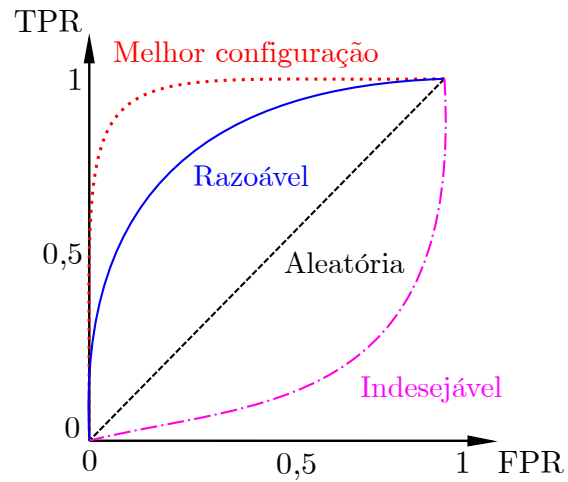


Fonte: Elaborado pelo autor.

Quanto mais próximo da configuração da Figura 47, melhor é a relação custo-benefício, o que implica, por conseguinte, em um melhor desempenho de classificação, e a área sob a curva (do inglês, *Area Under the Curve* – *AUC*) tende à unidade. Em outras palavras, se um modelo matemático-físico, ou de classificação, apresenta-se bem estruturado, bem como se é possível obter amostras cujas classes são distinguíveis com clareza, consegue-se obter classificações precisas. A partir das curvas ROC, portanto, tem-se informações mais refinadas sobre o fenômeno físico, a aplicação correta da técnica de classificação, e da distribuição de dados, o que fornece um critério mais elaborado em relação a simples porcentagens de acerto. Ilustrações de curvas ROC podem ser vistas na Figura 49; a curva pontilhada (em vermelho), representa uma curva mais próxima da ideal, equanto que as

outras apresentam curvas de classificação razoavelmente boa (linha contínua, em azul), aleatória ($FPR = TPR$, linha tracejada, em preto), e classificação considerada indesejável (ponto e traço, em magenta), respectivamente.

Figura 49 – Exemplo de distribuições de dados mais próximas das reais.



Fonte: Elaborado pelo autor.

5 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados obtidos para as técnicas de classificação. São incluídos resultados analíticos (como no caso da proposta de generalização da CCA) e baseados em sinais de EEG reais, que correspondem ao banco de dados disponíveis em Zhang (2018), referentes ao trabalho de Zhang et al. (2014), cuja licença é do tipo Creative Commons CC-BY. Essa licença permite que terceiros distribuam, remixem, adaptem e criem a partir de seu trabalho, mesmo para fins comerciais, desde que lhe atribuam o devido crédito pela criação original (COMMONS, 2018).

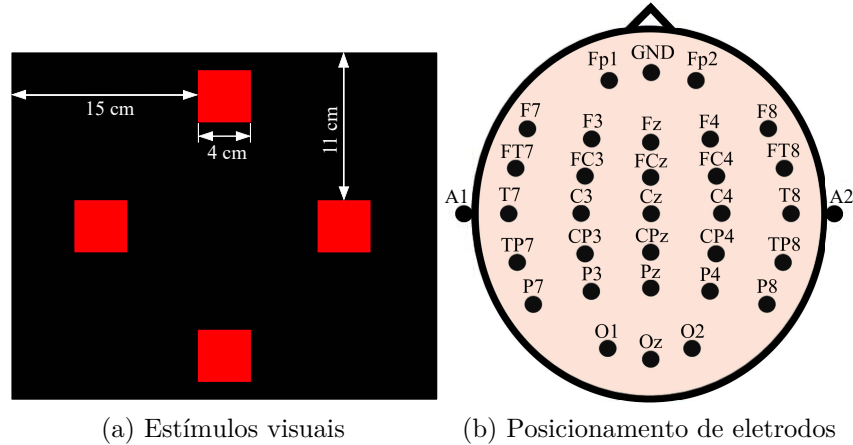
Os dados são obtidos para quatro frequências de estímulo: 9,75 Hz, 8,75 Hz, 7,75 Hz e 5,75 Hz. De acordo com os autores, os experimentos foram feitos por 10 indivíduos saudáveis, de 21 a 27 anos de idade, todos homens com visão normal ou corrigida para a normal. No experimento, os indivíduos se sentaram confortavelmente em uma cadeira a 60cm de distância da tela do computador, cuja taxa de atualização da imagem era de 85 Hz, em um quarto com luz reduzida. Quatro quadrados vermelhos, cada um de dimensões 4cm por 4cm, foram apresentados na tela como estímulo, como mostra a Figura 50a — nota-se que, neste experimento, os estímulos visuais foram aplicados simultaneamente, de modo que não se pode garantir que os voluntários, ao fixarem seu olhar sobre um deles, não sofreram qualquer influência de mais de um estímulo por rodada. No total, foram feitas 20 rodadas para cada frequência escolhida, com duração de 4 segundos por rodada precedida por um aviso de 2 segundos antes do estímulo ser efetuado.

Os sinais EEG foram gravados a partir de 30 eletrodos (dos quais apenas 8, representativos da região occipital do córtex cerebral, constam no banco; ou seja, eletrodos P7, P3, Pz, P4, P8, O1, Oz e O2) posicionados de acordo com o sistema 10-20, como mostra a Figura 50b, usando um amplificador Nuamps (NuAmp, Neuroscan, Inc.) a uma taxa de amostragem de 250 Hz, com filtros passa alta e passa baixa de 0.1 Hz e 70 Hz, respectivamente. Após a aquisição, aplicou-se um filtro Butterworth passa banda de 6ª ordem entre as frequências 4 e 45 Hz.

Devido à quantidade de dados, para a primeira etapa de resultados, que corresponde a colocar à prova a hipótese de aderência à normalidade, apenas um sinal de cada grupo é abordado para a verificação do critério proposto. Para clareza de entendimento, a Figura 51, por sua vez, apresenta uma ilustração geométrica do banco de dados utilizado. No

total, há, portanto, 8 eletrodos, 1000 pontos por eletrodo, 20 rodadas por frequência de estímulo, e 4 frequências de estímulo.

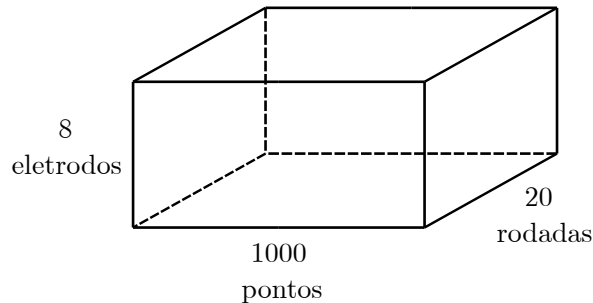
Figura 50 – Estímulos visuais e posicionamento utilizados no experimento.



Fonte: Zhang et al. (2014).

Figura 51 – Ilustração geométrica do banco de dados utilizado para os testes com EEG.

5,75Hz; 7,75Hz; 8,75Hz ou 9,75Hz



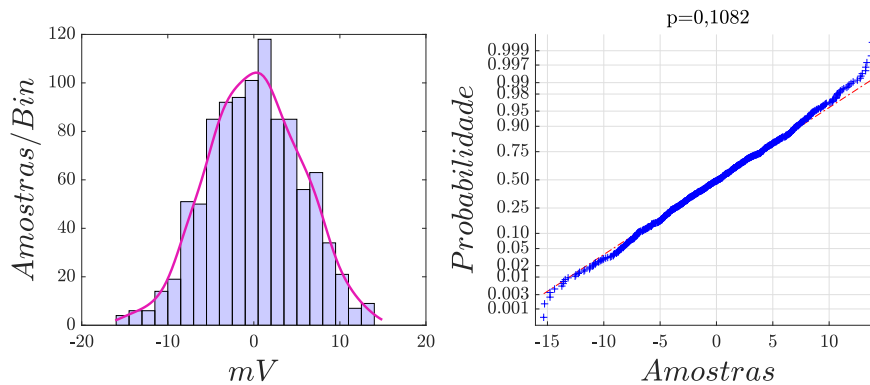
Fonte: Elaborado pelo autor.

5.1 TESTE DE ADERÊNCIA À NORMALIDADE

A seguir, apresenta-se os resultados obtidos para ambos os critérios de aderência à normalidade: visual e formal. As Figuras 52 a 55 mostram os histogramas, curvas de densidade de probabilidade aproximadas, e funções cumulativas de probabilidade para alguns testes selecionados de cada uma das quatro frequências de estímulo analisadas; acima de cada função cumulativa, há o valor obtido a partir do teste de Shapiro-Wilk adotando 95% de confiança. Como se observa, as distribuições se aproximam da distribuição gaussiana, bem como todos os valores de p são maiores que 0,05. Por fim, considerando todos

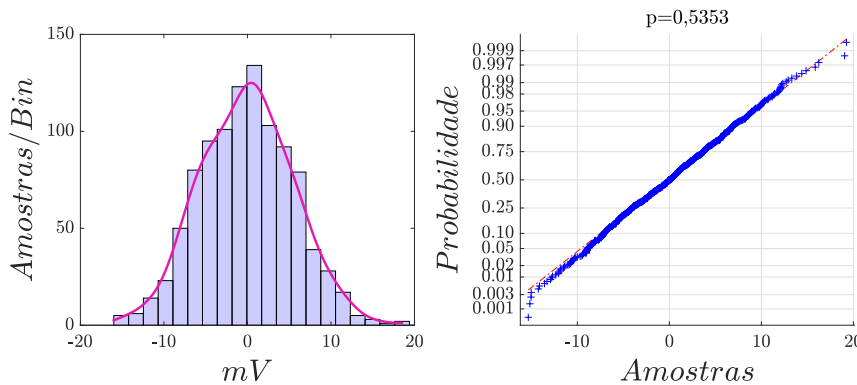
os dados da Figura 51, as porcentagens de testes de Shapiro-Wilk cujo valor é maior que 0,05 são, para as frequências de 9,75 Hz, 8,75 Hz, 7,75 Hz e 5,75 Hz, respectivamente: 35,63%, 36,88%, 30,63% e 29,38%. Embora essas porcentagens não sejam altas, salienta-se que elas se referem aos sinais originais, no domínio do tempo. Após a redução de dimensionalidade, forma na qual os dados são, de fato, utilizados para classificação, o teste é aplicado novamente, como visto nas próximas seções.

Figura 52 – Teste de aderência à normalidade para sinal de frequência de estímulo igual a 9,75 Hz.



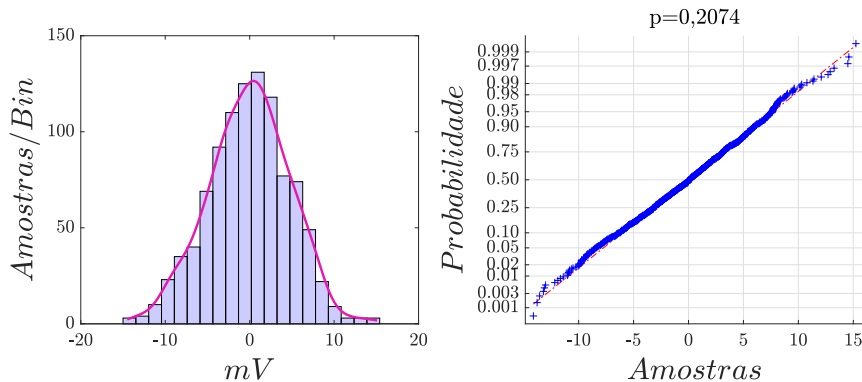
Fonte: Elaborado pelo autor.

Figura 53 – Teste de aderência à normalidade para sinal de frequência de estímulo igual a 8,75 Hz.



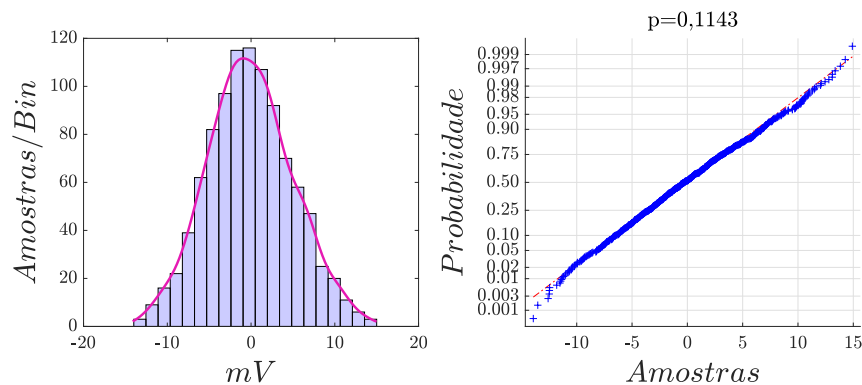
Fonte: Elaborado pelo autor.

Figura 54 – Teste de aderência à normalidade para sinal de frequência de estímulo igual a 7,75 Hz.



Fonte: Elaborado pelo autor.

Figura 55 – Teste de aderência à normalidade para sinal de frequência de estímulo igual a 5,75 Hz.



Fonte: Elaborado pelo autor.

Na seção seguinte, por sua vez, apresenta-se todos os resultados obtidos a partir da Análise de Correlação Canônica (CCA).

5.2 ANÁLISE DE CORRELAÇÃO CANÔNICA (CCA)

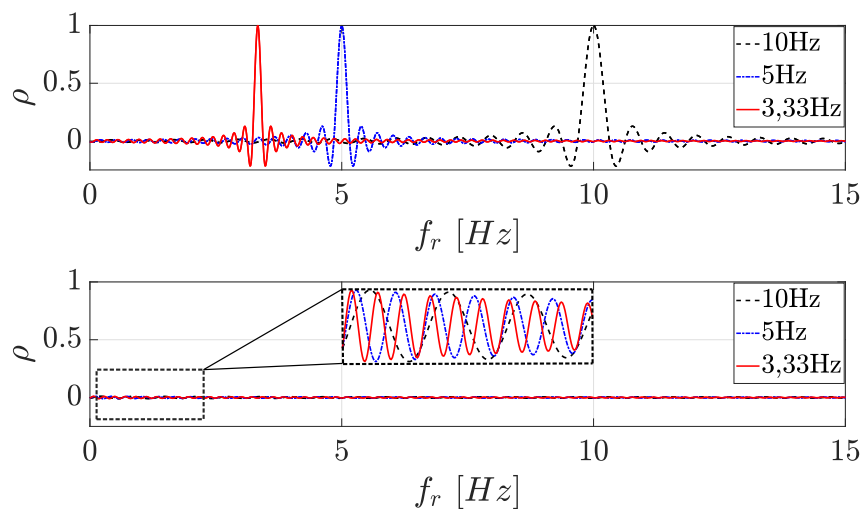
Para os resultados da Análise de Correlação Canônica, são realizadas simulações numéricas em três fases: a) soluções analíticas (baseadas na Equação 74, para senoides geradas), b) sinais reais (baseadas em um algoritmo CCA), obtidos do mesmo banco de sinais de domínio público conforme previamente mencionado neste texto, e c) análise de porcentagem de acerto do algoritmo para casos reais. A seguir, aborda-se a fase a), e todo parâmetro considerado importante é analisado separadamente: número de períodos (n_P), razão de frequências de amostragem (F), adição de harmônicos tanto no sinal de entrada

quanto no conjunto de sinais artificial (H_i e H_s , respectivamente), e adição de ruído ao sinal de entrada.

5.2.1 Soluções Analíticas

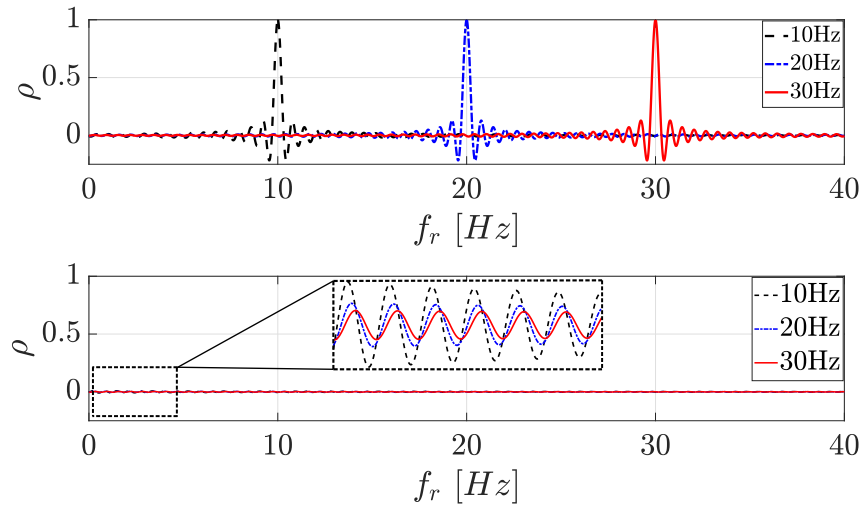
Inicialmente, apresenta-se nesta seção o comportamento geral da equação generalizada da CCA. A frequência de estímulo considerada é $f_m = 10$ Hz, as amplitudes das senoides são fixadas em 1 mV, e $F = 1$, o que implica em uma razão de frequências de amostragem constante. A Figura 56 mostra as contribuições de ambos os termos da Equação 74 em um gráfico f_r vs ρ , para $H_s = 3$, que representa a frequência reconhecida e o valor o coeficiente de correlação, respectivamente. Note a semelhança desse domínio com o tradicional da transformada de Fourier: o eixo das frequências se mantém, a amplitude se dá em termos de correlação. Pode-se observar que a função sinc causa a maior influência no resultado final, enquanto que o segundo termo adiciona uma contribuição pequena, conforme discutido no capítulo 4. No caso das sub-harmônicas, ρ é ainda menor, e as curvas tendem a tornar-se mais íngreme ao se aproximar da origem ($n\omega_x = m\omega_y = 0$), como comentado no capítulo 4. Um efeito similar ocorre quando harmônicas são adicionadas ao sinal de entrada ($H_i = 3$), como apresentado pela Figura 57: quando o valor da frequência reconhecida aumenta, a curva torna-se mais paralela ao eixo das abscissas. O efeito da adição de harmônicos é abordado em detalhes neste capítulo. Para as soluções seguintes, usa-se outro gráfico f_r vs ρ , que representa a frequência reconhecida pelo valor absoluto do coeficiente de correlação considerando ambos os termos da equação 74.

Figura 56 – Contribuição dos termos da equação generalizada CCA ($H_s = 3$).



Fonte: Elaborado pelo autor.

Figura 57 – Contribuição dos termos da equação generalizada CCA ($H_i = 3$).



Fonte: Elaborado pelo autor.

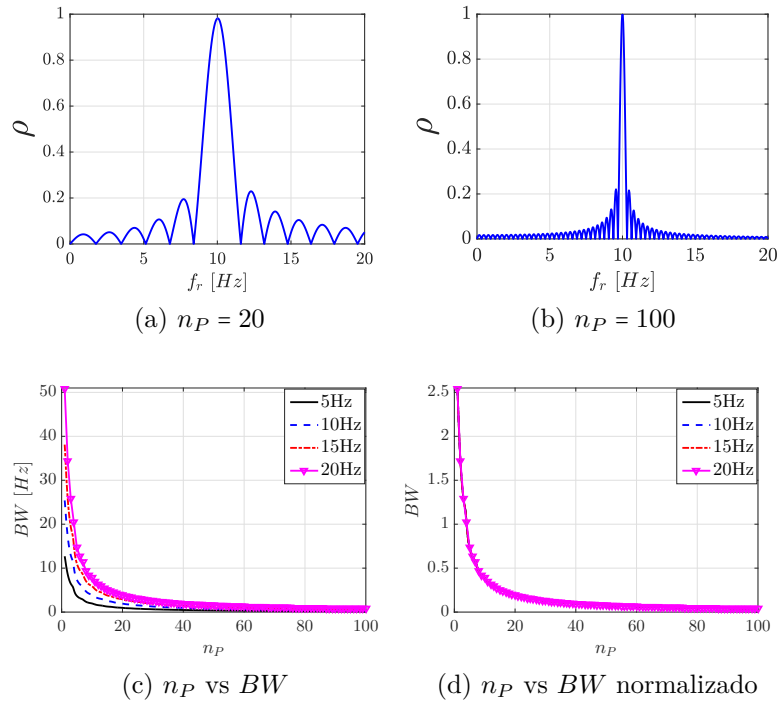
Número de Períodos

As Figuras 58a e b apresentam, respectivamente, as curvas para $n_P = 20$ e $n_P = 100$ com os parâmetros anteriores, exceto para n e m , ajustados para 1 (sem harmônicos), para auxiliar no entendimento. Observa-se que, conforme o número de períodos aumenta, a curva torna-se mais estreita em torno da frequência principal, o que acarreta uma melhor acurácia para o reconhecimento. A partir desse fato, novamente é possível estabelecer uma analogia com os efeitos observados na transformada de Fourier: o vazamento espectral (*spectral leakage*), causado pelo janelamento de um número não inteiro de períodos de um sinal harmônico, causa o espalhamento da energia da frequência principal para suas componentes vizinhas (WU; CHEN; CHEN, 2012), prejudicando a resolução em frequência; no caso da correlação, efeito similar ocorre por um número pequeno de períodos.

A Figura 58c, por sua vez, mostra como a largura de banda BW [Hz] em torno da frequência de estímulo varia com n_P , dadas diferentes frequências de estímulo; uma linha de referência foi estabelecida em $\rho = 0,5$ para calcular os limites dessa largura de banda. A Figura 58d, por fim, apresenta esse mesmo gráfico, porém normalizado pelas respectivas frequências de estímulo. A partir deste resultado, embora se note que BW tem pequena variação para $n_P > 30$, considera-se uma abordagem conservativa neste trabalho, com $n_P = 100$, especialmente visto o pequeno esforço computacional associado. Assim, usa-se uma janela temporal de 10 segundos para atingir a acurácia desejada. Essa análise esclarece que janelas temporais menores levam a maiores classificações incorretas em sistemas BCI,

como observado por Zhang et al. (2014).

Figura 58 – Efeito de aumento do número de períodos no valor de correlação.



Fonte: Elaborado pelo autor.

Uma discussão importante surge deste fato, pois deve-se sempre conciliar diversos aspectos práticos no protocolo SSVEP: janela temporal (quanto menor for, mais rápida é a resposta do sistema BCI), acurácia da técnica aplicada para tal janela temporal, graus de liberdade (isto é, intenções do usuário convertidas em ações mecânicas baseado no número de frequências de estímulo), e faixa de frequência disponível para a seleção de frequência de estímulo, entre outros. Uma vez que as principais restrições são fisiológicas, a fim de se obter sinais com amplitudes mais facilmente detectáveis, bem como evitar um maior risco de crises epiléticas, a faixa de frequência alpha, que compreende de 7,5 a 12,5 Hz, é a mais adequada. Baseado nisso, é possível definir um parâmetro que relaciona o tempo de resposta, número de períodos e frequência de estímulo, como dado pela equação 75 no capítulo 4: $t_{max} = n_P / f_m$.

Essencialmente, o parâmetro proposto é importante para escolher a aplicação do sistema BCI. Por exemplo, supondo que a seleção não envolva frequências duais, $f_m = 12$ Hz (limite da melhor frequência atingível, na transição entre as faixas alpha e beta) e o tempo de resposta deve ser igual a 4s $n_P = \frac{12}{s}(4s) = 48$. Se for requerida uma resposta mais rápida, então, tem-se $n_P = \frac{12}{s}(2s) = 24$. Nota-se que, para este primeiro caso, há uma melhor acurácia. Para o segundo, pode-se melhorar aumentando f_m ; entretanto, esse

valor está restrito pela faixa alpha, e a única possibilidade é utilizar outra frequência (cujo valor e harmônicas não sobreponham a primeira frequência e suas harmônicas); em outras palavras, por exemplo: $n_{P,1} = \frac{10}{s}(2s) = 20$ ou $n_{P,2} = \frac{7}{s}(2s) = 14$. Assim, para qualquer frequência, ou a resposta é mais rápida e, no entanto, menos exata, ou é mais lenta e, por conseguinte, mais precisa; do contrário, os parâmetros devem ser alterados para novas implicações. Se $n_{P,ot} = 100$ é fixado, desejando-se obter maior (neste caso), e as mesmas frequências são avaliadas: $t_{max,ot} = \frac{100}{12} = 8,33s$. De forma similar, $t_{max,1} = (100/10) = 10s$ e $t_{max,2} = (100/7) = 14,28s$. Baseado nessa análise, é factível definir-se o fator de acurácia FA , apresentado no capítulo 4 pela equação 76, para ambas situações, respectivamente, ou seja:

$$FA = \frac{n_P}{n_{P,otimo}} \Leftrightarrow FA = \frac{t_{max,otimo}}{t_{max}} = \frac{f_{m,otimo}}{f_m}$$

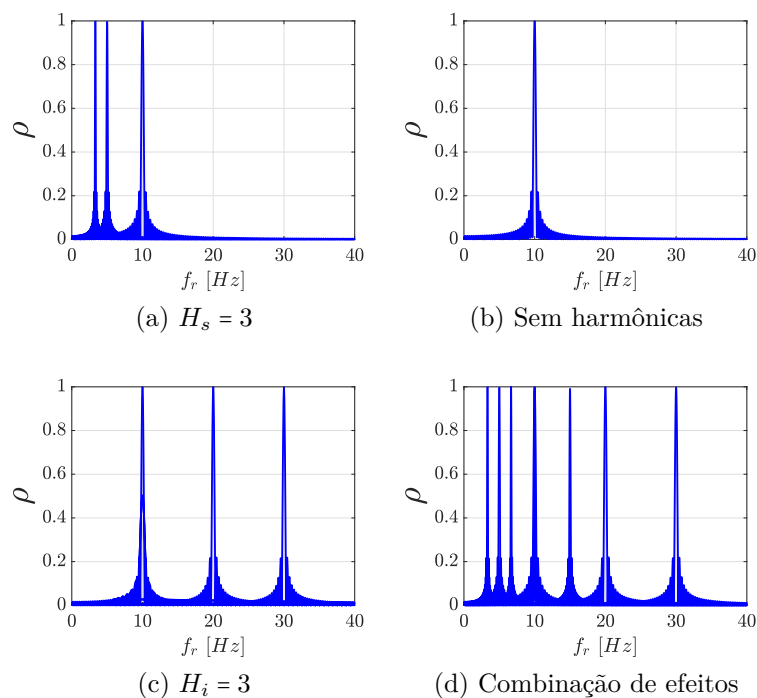
Nos exemplos apresentados, tem-se para o primeiro caso: $FA = (20/24) = 0,8333$ e $FA = (14/24) = 0,5833$ (83,33% e 58,33% do melhor resultado possível); para o segundo caso: $FA = (8,33/10) = (10/12) = 0,8330$ e $FA = (8,33/14,28) = (7/12) = 0,5833$. É importante enfatizar que esse tipo de protocolo é de difícil generalização devido às particularidades de cada sujeito e do eletroencefalograma obtido. Então, não se pode estabelecer relações definitivas no contexto de interface cérebro-máquina, pois o mais importante é verificar e garantir a maior robustez possível do protocolo, dadas as limitações. Portanto, essa equação é apenas um indicativo teórico de acurácia. Assim, com esses resultados conclui-se a etapa de análise do número de períodos. Na próxima seção apresenta-se a análise analítica da equação generalizada CCA, com uma discussão sobre a adição de harmônicos.

Adição de Harmônicos

A Figura 59 apresenta o efeito da adição de harmônicos tanto no sinal de entrada (H_i) e conjunto de sinais artificiais (H_s); cada um, a sua vez, fixado em 3 como um valor de referência (MÜLLER-PUTZ et al., 2005). Pode-se observar que, conforme harmônicas são adicionadas no conjunto artificial, sub-harmônicas surgem, de fato, no gráfico f_r vs ρ (5 Hz e 3,33 Hz), enquanto que super-harmônicas surgem quando o sinal de entrada tem $H_i = 3$ (20 Hz e 30 Hz); por fim, uma combinação desses dois efeitos resulta em mais componentes do que a soma de todas as harmônicas adicionadas; isso se deve, como esperado pela análise do capítulo 4, à superposição causada pelas sub-harmônicas de cada harmônica adicionada ao sinal de entrada, como visto na Figura 60. Deve-se salientar que, somando o número para cada caso, $n_{f_r} = nm = 9$ (há, de fato, 9 picos, previsto pela equação $n_{f_r} = nm$); contudo, após a superposição, apenas $n_{f_r} = 7$ podem ser identificados.

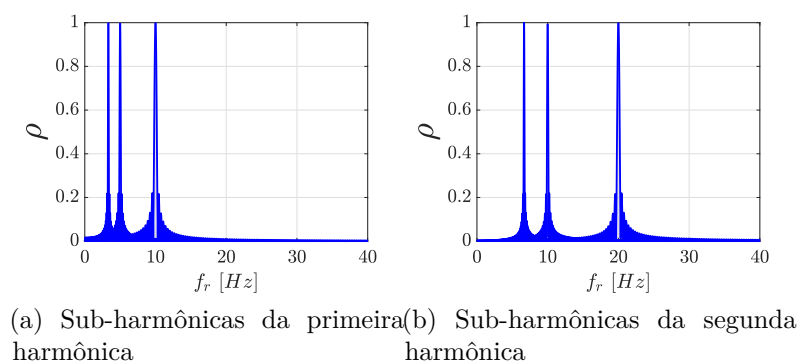
Não obstante, todas as classificações são realizadas corretamente.

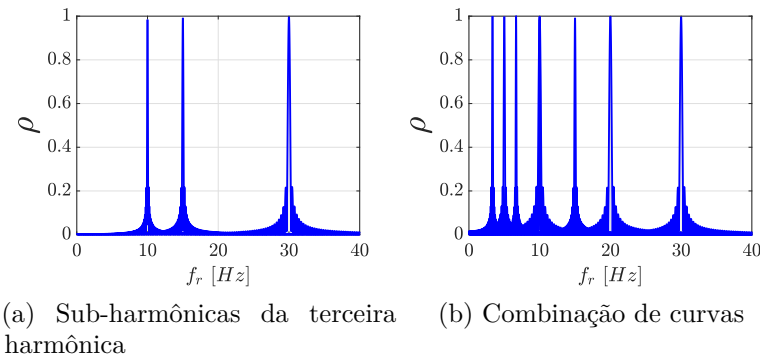
Figura 59 – Efeito de adição de harmônicas.



Fonte: Elaborado pelo autor.

Figura 60 – Sub-harmônicas de cada harmônica adicionada no sinal de entrada.





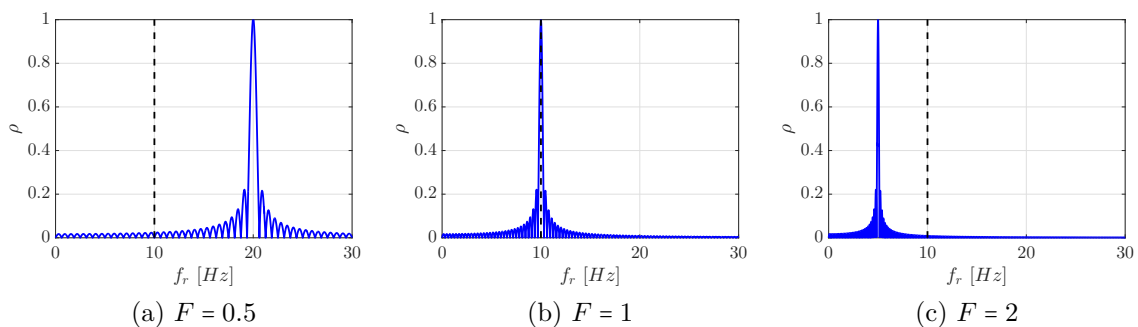
Fonte: Elaborado pelo autor.

Frequência de Amostragem

Em relação à razão de frequências de amostragem F , a Figura 61 mostra como alterações nesse parâmetro influenciam no resultado final da correlação: para qualquer valor real $F < 1$, a frequência reconhecida situa-se à direita da frequência de estímulo, e para $F > 1$, situa-se à esquerda. Essencialmente, isso se deve à sub-amostragem (*aliasing*); e para esses casos particulares (números inteiros), conforme F varia, torna-se uma sub-harmônica ou super-harmônica da frequência principal, como visto na equação 74. Deve-se salientar que, uma vez que essa variável está puramente ligada ao processamento do sinal, o comportamento verificado é válido para qualquer combinação na adição de harmônicas, o que faz de F um fator de correção, podendo ser expresso como visto pela equação 77, repetida a seguir para facilitar para o leitor:

$$\hat{f} = F[\operatorname{argmax}(\rho_m)], \quad m = 1, 2, \dots, M$$

Figura 61 – Efeito de alterações em F .

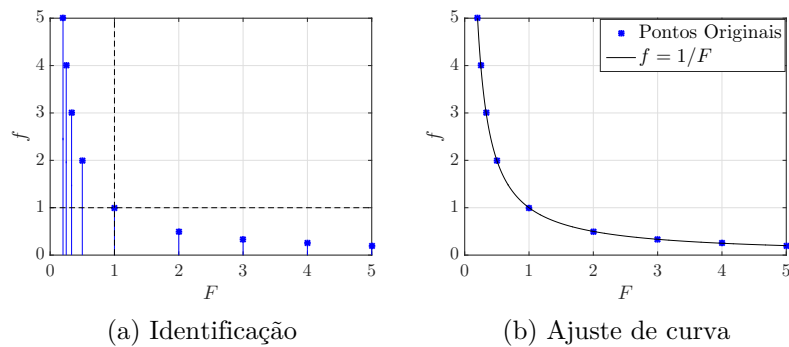


Fonte: Elaborado pelo autor.

Essa expressão torna possível a aquisição do sinal a qualquer taxa de amostragem que esteja de acordo com os requisitos mínimos do teorema de Nyquist-Shannon (SHANNON,

1949). Ela fornece não apenas maior flexibilidade ao se utilizar a CCA, mas também maior confiabilidade para o sistema. Baseado nisso, é possível estabelecer uma nova relação entre F e a frequência reconhecida normalizada $f = f_r/f_m$. A Figura 62 apresenta um gráfico de F vs f ; nota-se que, de fato, $F = 1$ é o único valor para o qual há identificação correta. Também, verifica-se que a curva é simétrica e seu comportamento é dado como visto pela equação 78, ou seja, $f = 1/F$.

Figura 62 – Relação entre F e f .



Fonte: Elaborado pelo autor.

o que é coerente: seja maior ou menor, a frequência reconhecida se distancia da frequência de estímulo se a razão de frequências de amostragem for diferente da unidade. Isso conclui a análise desse parâmetro, e a seção seguinte aborda o efeito da adição de ruído ao sinal de entrada.

Adição de Ruído

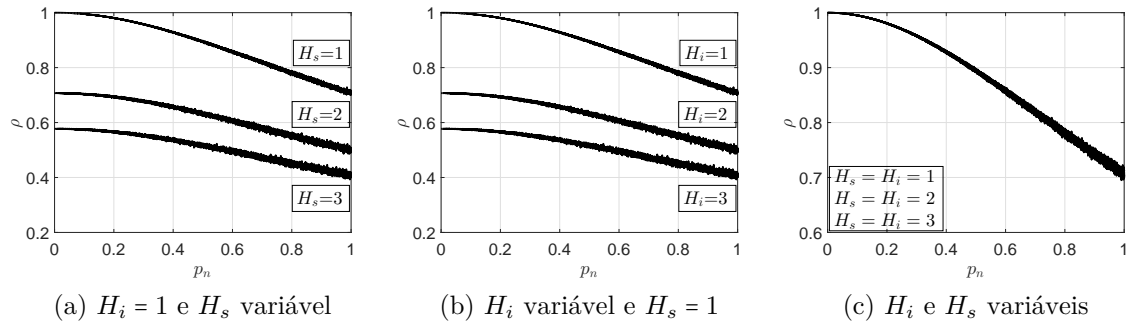
O efeito da adição de ruído ao sinal de entrada é mensurado aplicando uma senoide e ruído branco:

$$x_n(t) = x(t) + p_n r(t) \quad (158)$$

onde $x_n(t)$ é a curva do seno $x(t)$ após a adição de ruído, representado por $r(t)$, uma distribuição gaussiana normalizada multiplicada pelo valor RMS de $x(t)$, e um fator de porcentagem p_n . A Figura 63 apresenta valores de p_n em relação a ρ considerando: a) H_i fixo e H_s variável (maior quantidade de harmônicas no conjunto de sinais artificial); b) H_i variável e H_s fixo (maior quantidade de harmônicas no sinal de entrada); c) H_i e H_s variáveis simultaneamente. Observa-se que a adição de ruído é claramente um fator fundamental para se diminuir o valor total de correlação. Além disso, quando $H_i \neq H_s$, esse decréscimo é ainda mais severo, o que sugere que, para um melhor reconhecimento,

o número de harmônicas no sinal de entrada e no conjunto artificial de sinais deve ser o mesmo.

Figura 63 – Efeito da adição de ruído.



Fonte: Elaborado pelo autor.

Na seção seguinte, apresenta-se uma análise de sinais reais, que envolve a apresentação de gráficos semelhantes aos vistos nas seções anteriores, a fim de verificar o comportamento dos parâmetros propostos neste trabalho em aplicações de engenharia.

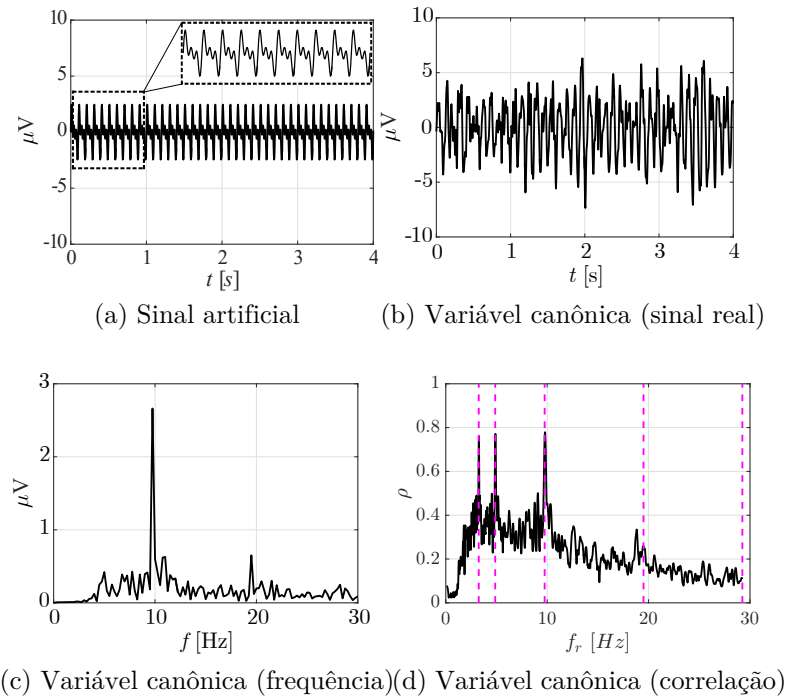
5.2.2 Sinais de EEG

Para os sinais de EEG, a princípio, apresenta-se seu comportamento em relação aos sinais considerados nas soluções analíticas. Para tal, como abordado no capítulo 4, as variáveis canônicas, escolhidas de uma das 20 rodadas existentes, são consideradas representativas da atividade geral cerebral sob estímulo de cada uma das frequências selecionadas por Zhang et al. (2014). Ou seja, um único sinal, após a devida aplicação dos pesos originados do maior autovetor de cada situação, é comparado com o sinal gerado artificialmente, constituído apenas por um somatório de senoides. O número de harmônicas no sinal de entrada artificial é igual a 3 (i.e., $H_i = 3$), o número de harmônicas no sinal de entrada real é desconhecido, e o número de harmônicas no conjunto de sinais artificial é, também, igual a 3 (i.e., $H_s = 3$).

As Figuras 64a e b ilustram a comparação desses sinais no domínio do tempo; a Figura 64c, por sua vez, apresenta o espectro em frequência da variável canônica. Observa-se duas faixas de frequência com componentes pronunciadas: a frequência de estímulo (9,75 Hz) e sua primeira super-harmônica (~ 20 Hz). A Figura 64d, por fim, apresenta a variável canônica no domínio f_r vs ρ , como nos casos vistos anteriormente, após a utilização de um algoritmo CCA, diferentemente dos sinais analíticos. Nota-se que, apesar do ruído (que, de fato, reduz o valor geral de correlação; a diferença entre o número de harmônicas

contidas nesses sinais é outra possível causa), as maiores amplitudes ocorrem em 9,75 Hz, duas de suas sub-harmônicas (4,875 Hz e 3,25 Hz), o que implica em $H_s = 3$ (estipulado inicialmente e, portanto, confirmado), e um pico começa a surgir em uma super-harmônica (~ 20 Hz), o que explicita o comportamento não linear do cérebro, e endossa a hipótese de superposição para o modelo: a frequência de estímulo é bem definida; suas sub-harmônicas surgem devido ao conjunto de sinais artificial (e, portanto, não estão diretamente ligadas à dinâmica neural), e são bem definidas também, com amplitude semelhante à principal; uma super-harmônica surge, porém sua amplitude é menor, e suas sub-harmônicas não são observáveis (possivelmente devido ao ruído e baixa amplitude); outras super-harmônicas não são visíveis nesse caso.

Figura 64 – Comparação entre variável canônica (sinal real) e resultado analítico para $f_m = 9,75$ Hz.

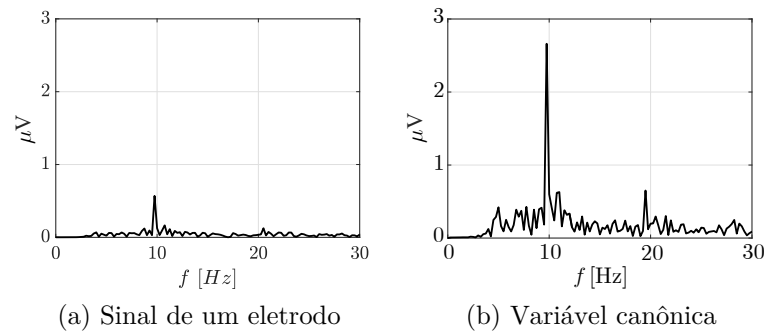


Fonte: Elaborado pelo autor.

Outro ponto importante é a amplitude em frequência vista nos sinais reais: note a diferença neste valor em comparação aos sinais das Figuras 15 a 18 (que provêm diretamente de eletrodos e não são, portanto, variáveis canônicas); por se tratarem de variáveis canônicas (e, conseqüentemente, pela melhora na relação sinal-ruído), evidencia-se melhor a componente em frequência. A Figura 65 mostra essa comparação para a frequência escolhida de 9,75 Hz (sinal da Figura 18 e variável canônica da Figura 64 na mesma escala). Esse fato é endossado quando se avalia os autovetores de projeção dessas variáveis canônicas, como mostrado na Figura 66. Nela, tem-se o valor absoluto dos pesos de cada

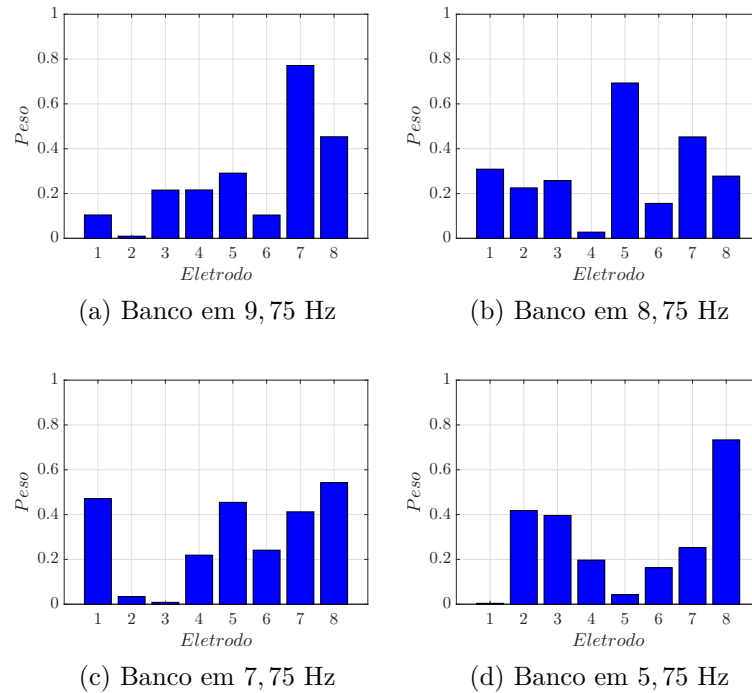
eletrodo utilizado; nota-se que, para a variável canônica cujo espectro de frequência é mostrado na Figura 65, o eletrodo de maior peso é o número 7, verificado, por inspeção (discutida em mais detalhes na seção sobre componentes principais), como mais próximo à região occipital (e, desta forma, mais significativo) em relação aos outros, o que não ocorre para outras variáveis canônicas de frequências distintas da frequência-alvo — isto é, quando banco de sinais artificial e sinal de entrada não têm a mesma frequência.

Figura 65 – Comparação de amplitudes em frequência entre sinal de um eletrodo e variável canônica para $f_m = 9,75$ Hz.



Fonte: Elaborado pelo autor.

Figura 66 – Pesos de cada eletrodo na classificação da frequência-alvo de 9,75 Hz.

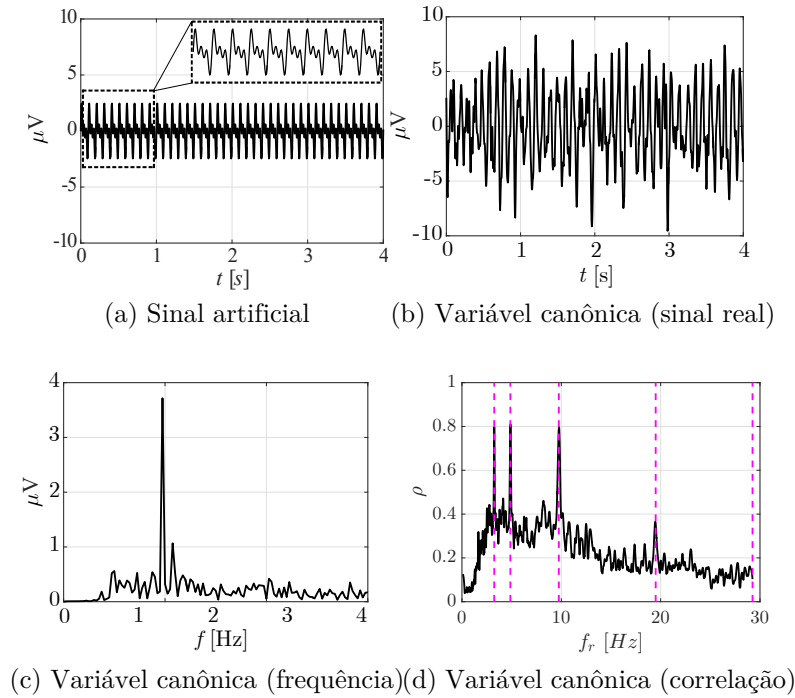


Fonte: Elaborado pelo autor.

É interessante analisar o comportamento dessas curvas nos três domínios apresentados, pois a informação contida em um deles complementa a dos outros. Por exemplo, a Figura

67 apresenta uma nova variável canônica referente a um outro teste realizado sob a mesma frequência de estímulo anterior. Nota-se que, neste caso, o domínio de frequência vs amplitude não explicita a componente super-harmônica que existe no sinal de entrada; entretanto, o domínio f_r vs ρ o faz, semelhante ao visto na Figura 64.

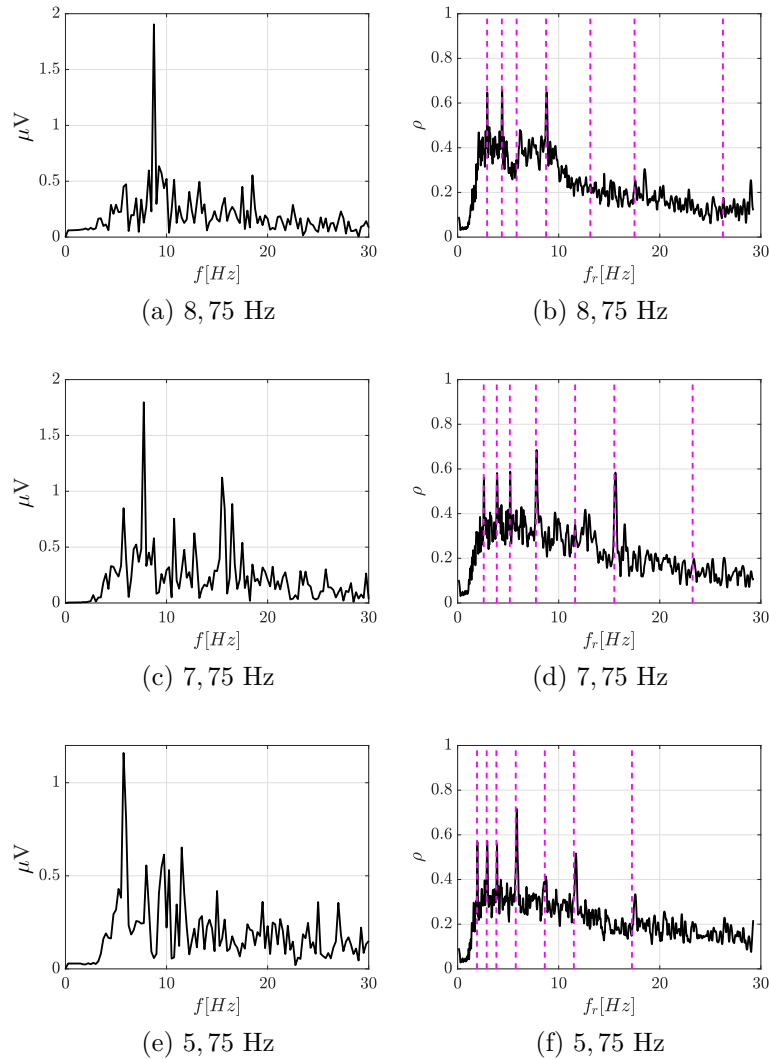
Figura 67 – Comparação entre variável canônica (sinal real) e resultado analítico para $f_m = 9,75$ Hz.



Fonte: Elaborado pelo autor.

Para fins de comparação, a Figura 68 apresenta os outros casos contidos no banco de sinais, para as frequências 8,75 Hz, 7,75 Hz e 5,75 Hz, respectivamente. Resultados similares aos das Figuras 64 e 67 são obtidos, com exceção do valor geral de correlação; os picos de $f_m = 9,75$ Hz são um pouco mais pronunciados que os outros casos, o que vem ao encontro do resultado obtido para análise de número de períodos: frequências mais próximas da faixa alpha, dado uma janela temporal fixa, acarretam em maior número de períodos, maior informação sobre o sinal e, portanto, maior coeficiente de correlação.

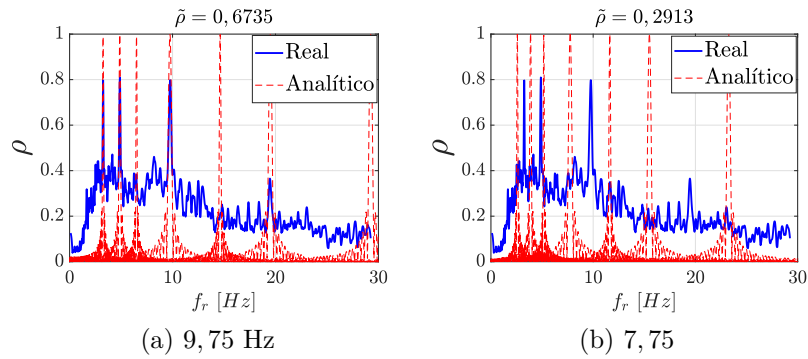
Figura 68 – Componentes em frequência e correlação para 8,75 Hz, 7,75 Hz e 5,75 Hz.



Fonte: Elaborado pelo autor.

É possível, finalmente, comparar-se diretamente os resultados de sinais reais com os analíticos visual e matematicamente. A Figura 69a apresenta os resultados para $f_m = 9,75$ Hz (Figura 67d). A comparação visual é indicada pela sobreposição de curvas, e a comparação matemática é dada pelo valor de $\tilde{\rho}$, correlação obtida após aplicar o algoritmo CCA entre ambas as curvas novamente, como apresentado pela equação 79 no capítulo 4. Observa-se que existe, de fato, uma boa concordância para ambos os critérios; os únicos picos “mal ajustados” são os referentes a sub-harmônicos das harmônicas, que não são significativos em termos de dinâmica, apenas caracterizam a técnica em si. Por fim, a mesma concordância não ocorre na Figura 69b, quando a frequência escolhida para a curva analítica é diferente da frequência de estímulo.

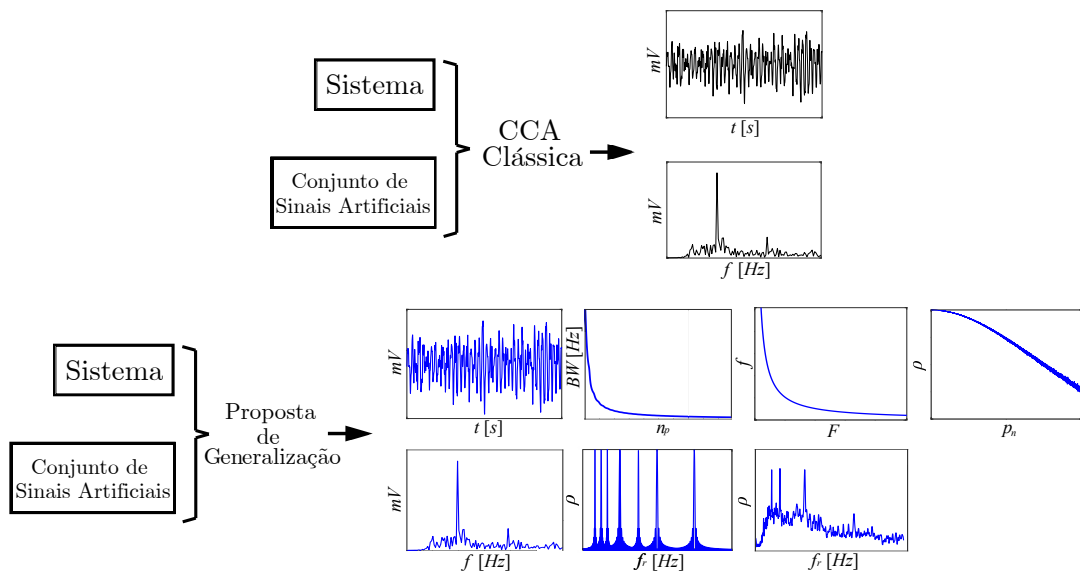
Figura 69 – Comparação entre curvas analíticas e de sinal real para 9,75 Hz.



Fonte: Elaborado pelo autor.

Como verificado, é possível se obter informações essenciais a partir da proposta de generalização da CCA. Para resumi-las e esclarecê-las, a Figura 70 apresenta um fluxograma comparando as abordagens clássica e a proposta deste trabalho. A partir da introdução de novos parâmetros, pode-se compreender melhor quais deles influenciam mais significativamente no reconhecimento correto e valor final de correlação, bem como realizar-se uma seleção de frequência de estímulo de forma mais cautelosa, incluindo fatores fisiológicos, técnicos e operacionais. Essencialmente, seja em termos de informação disponível do sistema, or introdução de parâmetros resultantes de uma compreensão mais ampla dessa técnica, A seguir, apresenta-se a análise de porcentagem de acerto da CCA, incluindo aplicações de validação cruzada e curvas ROC.

Figura 70 – Fluxograma de similaridades e diferenças entre a abordagem clássica da CCA e a proposta de generalização.



Fonte: Elaborado pelo autor.

5.2.3 Análise de Porcentagem de Acerto

A princípio, como a técnica CCA envolve comparações de sinais reais com sinais criados artificialmente, ela não requer treinamento, o que implica que um teste de validação cruzada, neste caso, não é aplicável. Esta seção aborda, portanto, apenas aspectos relacionados à porcentagem de acertos na classificação, bem como construção de curvas ROC; salienta-se, também, que o teste de validação cruzada é aplicado para todas as outras técnicas, porém a construção de matrizes de confusão e curvas ROC são empregados apenas para a CCA, visto que é a técnica de maior relevância no presente trabalho.

A Tabela 2 apresenta os resultados obtidos para a CCA, mostrando as frequências de estímulo, a porcentagem de acerto por frequência e geral, e valor de correlação médio. A porcentagem é calculada a partir da relação identificação correta/total de testes, onde o total de testes é igual ao número de rodadas realizadas no banco de sinais (20 por frequência de estímulo); isto é, cada sinal de dimensão 8×1000 retirado das matrizes cúbicas é comparado com 4 bancos artificiais de senos e cossenos, aquele que apresentar maior correlação indica a frequência de estímulo correta (Figura 31). O valor de correlação médio, por sua vez, é obtido a partir da média entre todos os valores individuais a cada rodada; isto é, para cada 1 dos 20 testes, atribui-se identificação correta, se este indicar, de fato, a frequência esperada, ou incorreta caso contrário; o somatório das identificações corretas dividido por 20 é a porcentagem final de acerto. A porcentagem geral de acerto, por sua vez, é a média aritmética das porcentagens de acerto por frequência. Percebe-se que, de forma geral, as porcentagens por frequência são altas, especialmente considerando o fato de que os sinais de comparação são artificiais. Também, nota-se que, conforme a frequência-alvo se distancia da faixa alpha, o valor de correlação médio torna-se mais baixo, fato que endossa, novamente, o resultado obtido na seção de análise de número de períodos. Por fim, o desempenho médio dessa técnica é de 83,75% de acertos, valor relativamente alto.

Tabela 2 – Porcentagem de acerto e correlação média da CCA clássica para cada frequência de estímulo.

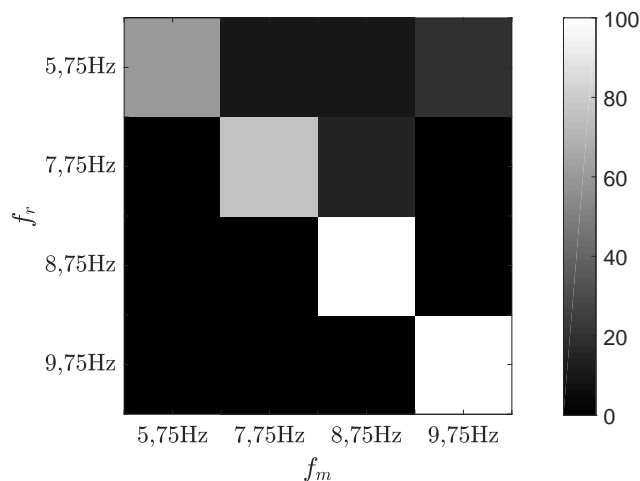
Frequência de estímulo (Hz)	Porcentagem de acerto (%)	Valor de correlação médio
9,75	100	0,7454
8,75	100	0,5696
7,75	75	0,5557
5,75	60	0,4567
Todas	83,75	0,5818

Fonte: Elaborado pelo autor.

A porcentagem geral de acerto, como abordado anteriormente, pode não ser uma informação suficiente para se avaliar o desempenho da técnica. Pelo fato da CCA ser a técnica de maior relevância neste trabalho, apresenta-se, a seguir, a matriz de confusão referente às frequências de estímulo, Figura 71. Essencialmente, essa matriz é um extensão da Tabela 2, considerando não só a porcentagem de acerto para a frequência esperada (alvo), mas também as porcentagens de classificações incorretas por frequência.

Nota-se que a matriz de confusão endossa o fato de que: quanto mais distante da faixa alpha, menos eficiente é a classificação. Também, os cenários menos favoráveis são, em ordem crescente, para 7,75 Hz e 5,75 Hz, respectivamente; para o primeiro, observa-se que há confusão com a frequência de 8,75 Hz, que é próxima à frequência de estímulo; para o segundo, há classificações incorretas com todas as outras frequências, mas especialmente para 9,75 Hz, frequência próxima à primeira harmônica de 5,75 Hz. Isso pode ajudar a explicar o porquê das menores porcentagens de acerto; não obstante, há de se lembrar que, como os estímulos foram apresentados simultaneamente no protocolo de execução (ZHANG et al., 2014), é possível que o indivíduo a partir do qual os sinais foram obtidos tenha sofrido uma influência indevida de um estímulo visual não relacionado ao alvo no qual fixou seu olhar. Para ilustrar essa hipótese, a Figura 72 apresenta sinais no domínio da frequência com duas identificações: frequência de estímulo e frequência reconhecida; nota-se que, de fato, essas frequências não coincidem para esses casos.

Figura 71 – Matriz de confusão utilizando CCA.

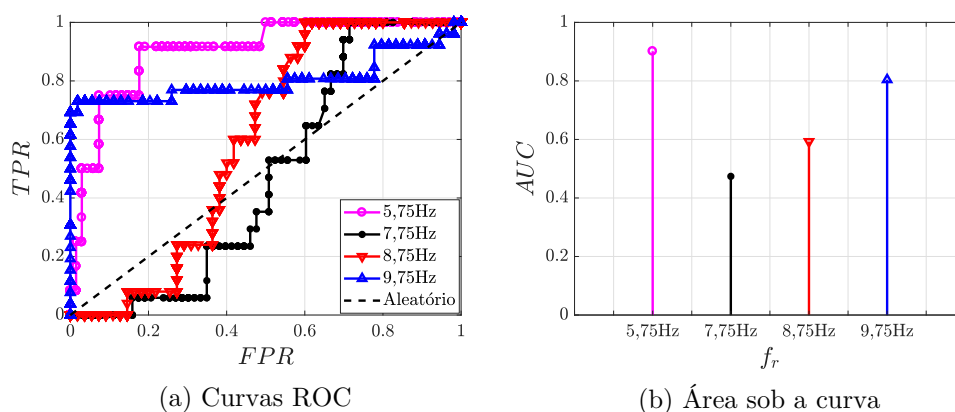


Fonte: Elaborado pelo autor.

dispersão dos dados), bem como para se avaliar a relação de custo-benefício de tal identificação, apresenta-se as curvas ROC para cada frequência na Figura 74a e as áreas sob as curvas na Figura 74b. Para sua construção, o limiar considerado é o próprio valor de ρ , que neste caso variou de 0, valor mínimo, até 1, valor máximo. Observa-se que as piores relações de FPR por TPR ocorrem para as frequências intermediárias, como esperado, devido à baixa separabilidade vista na Figura 73. Melhores separabilidades implicam em melhores relações custo-benefício, justificando as melhoras expressivas vistas para 5,75 Hz e 9,75 Hz. Desta forma, quanto maior o Δf existente entre uma frequência e outra, melhor a identificação. Neste caso, as frequências são avaliadas em conjunto, pois deseja-se maiores graus de liberdade; entretanto, também é necessário aumentar a diferença entre uma e outra.

Conclui-se que, assim como a questão de escolha de frequência de estímulo e tempo de resposta, a relação entre essa frequência e a identificação correta é assimétrica, implicando na priorização de um ganho em detrimento do outro. Por fim, os limiares que melhor separam os dados, isto é, os limiares a partir dos quais coordenadas no plano ROC são obtidas com a menor distância do ponto (0, 1) (melhor classificação possível), são: 0,5160 para 5,75 Hz, 0,6190 para 7,75 Hz, 0,5860 para 8,75 Hz e 0,6490 para 9,75 Hz. Esse valores podem auxiliar no processo de reconhecimento, visto que estabelecem faixas nas quais a relação entre falsos positivos e verdadeiros positivos é maximizada (para os dados considerados).

Figura 74 – Curvas ROC para as frequências consideradas.



Fonte: Elaborado pelo autor.

As análises feitas nessa seção evidenciam que: i) é indispensável, para a aplicação do protocolo SSVEP, conciliar uma seleção cautelosa de frequências de estímulo, bem como a quantidade dessas frequências apresentadas simultaneamente (e, portanto, de número

de graus de liberdade), com um reconhecimento preciso, de modo que a melhora em um desses fatores pode acarretar a piora do outro; ii) a dificuldade do reconhecimento correto de padrões (associada à necessidade de maiores graus de liberdade para perspectivas de aplicação real) é, desta forma, intrínseca ao protocolo SSVEP, e a técnica escolhida deve atender suficientemente essa demanda; iii) a CCA, a princípio, embora simples e precisa (em termos de reconhecimento individual de frequências), pode não conseguir contornar a questão da simultaneidade de estímulos visuais; iv) entretanto, sugere-se que o estudo aprofundado de seu comportamento e desempenho, como feito anteriormente, integrando aspectos de natureza fisiológica, técnica e operacional, pode colaborar para que o reconhecimento correto seja realizado apesar das limitações da técnica.

5.3 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

A Análise de Componentes Principais é feita em duas etapas: a) visualização das componentes principais considerando-se um número variável de eletrodos (neste caso, o número total de eletrodos do córtex visual, ou apenas aqueles localizados na parte mais posterior do cérebro), b) validação cruzada e verificação da porcentagem de acerto a partir das métricas de classificação de dados; a saber: distância Euclidiana (Equação 90), distância de Mahalanobis (Equação 91) e Naïve-Bayes (Equação 96).

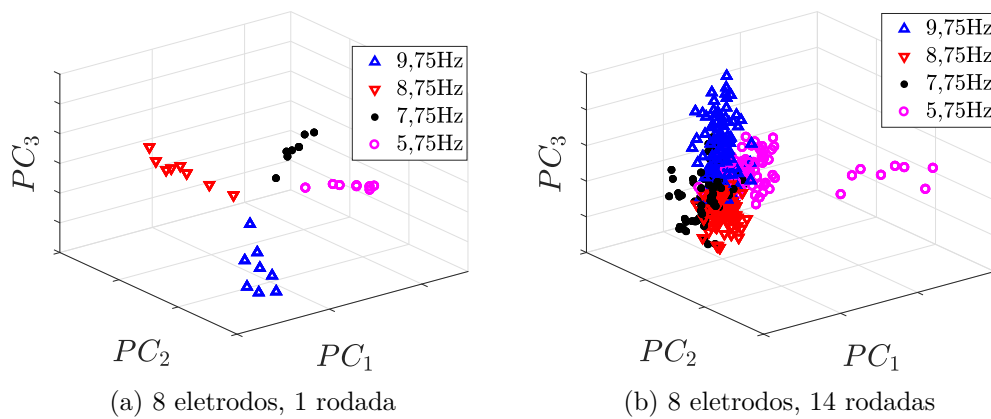
5.3.1 Visualização das Componentes Principais

Para a visualização das componentes principais, duas configurações gerais são consideradas, sendo subdivididas em sequência: número total de eletrodos do córtex visual (todos aqueles da região posterior do cérebro, como indica a Figura 11), ou apenas 3 dos 8 eletrodos totais (escolhidos por inspeção após sucessivas verificações no plano de componentes principais, de forma que esses eletrodos, segundo o Mapa de Brodmann, são aqueles mais próximos à região occipital, córtex visual). Para cada um deles, os dados de 1 das rodadas foi analisado primeiramente, a fim de se verificar a disposição dos eletrodos; em seguida, 70% das rodadas (14 das 20 existentes) foram separadas para construir uma nova disposição em componentes principais. Considera-se essa porcentagem para melhor se visualizar o efeito de separabilidade entre os grupos de dados, fato especialmente útil para a visualização de suas densidades de probabilidade, mostradas na seção seguinte.

A Figura 75a apresenta a primeira configuração, para umas das rodadas, e a Figura 75b, o mesmo caso, porém considerando 70% das rodadas. Nota-se que, quando há apenas

uma rodada, a separação visual é melhor definida; quando se adiciona mais dados, há perda de distinção entre os grupos. Possivelmente, isso se deve ao fato de que, quando se incluem mais dados, neste caso, apesar de haver mais informação sobre o sistema, os dados provêm de uma região mais distante daquela principal do córtex visual (isto é, eletrodos diferentes dos 3 mencionados inicialmente); isso é visível nas componentes principais, pois nota-se que, progressivamente, as amostras tendem a se tornar mais esparsas ao longo dos grupos, não se aglutinando, mas sim se sobrepondo.

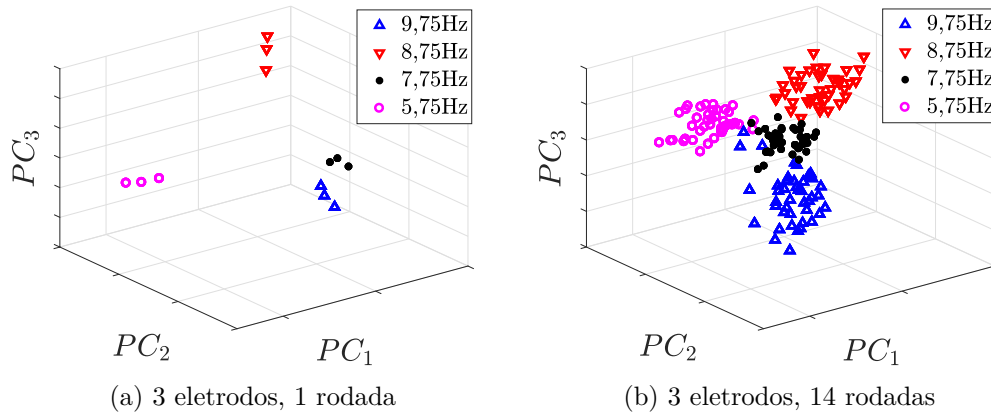
Figura 75 – Comparação entre curvas analíticas e de sinal real para 9,75 Hz.



Fonte: Elaborado pelo autor.

A Figura 76a, por sua vez, mostra a segunda configuração, para uma das rodadas, e a Figura 76b, a mesma situação, porém considerando 70% delas. Observa-se que os grupos de dados estão melhor separados, em ambas as situações, se comparados à configuração anterior. Essa melhora pode ser justificada pela inclusão de eletrodos mais específicos do córtex visual — os mais influenciados pelo estímulo visual e, portanto, com efeito mais representativo da sincronização de disparos dos neurônios.

Figura 76 – Comparação entre curvas analíticas e de sinal real para 9,75 Hz.



Fonte: Elaborado pelo autor.

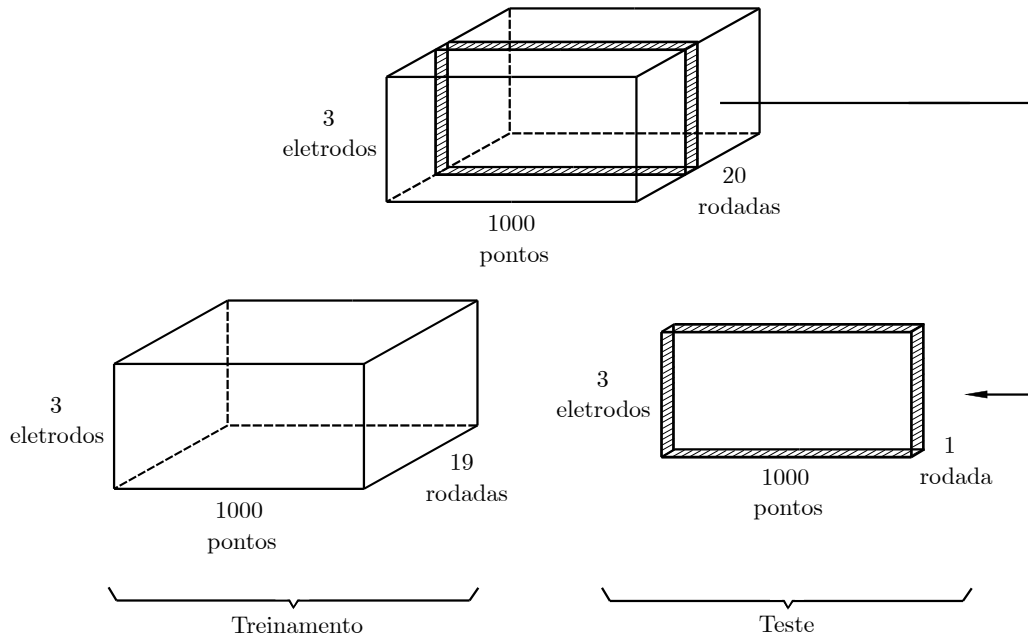
A partir desses resultados, as classificações são feitas baseadas na situação considerada como mais significativa, isto é, da Figura 76b, aplicando-se validação cruzada.

5.3.2 Validação Cruzada e Porcentagem de Acerto

Para se realizar a validação cruzada, utiliza-se o conceito apresentado pela Figura 44: o grupo original de dados é subdividido em k partições (neste caso, $k = 20$, devido à existência de 20 rodadas), e um processo iterativo é estabelecido: a cada iteração, retira-se 1 partição para ser utilizada como teste, e as restantes são utilizadas para a construção do modelo de componentes principais. Salienta-se, também, que apenas 3 dos 8 eletrodos são utilizados, considerados como os mais significativos, como visto na seção anterior. Uma ilustração dos dados utilizados para treinamento e teste é mostrada na Figura 77. Nela, representa-se a partição de um dos grupos de dados, para uma frequência. Cada grupo (Fig. 51), por sua vez, apresenta a mesma divisão, de modo que as quatro partições individuais são utilizadas como testes, enquanto que os quatro conjuntos de treinamento, juntos, fornecem o modelo de separabilidade de dados.

Após a realização das k iterações e aplicação da equação 155, calcula-se a porcentagem média de acerto, como mostrado na Tabela 3, utilizando as métricas: distâncias Euclidiana e Mahalanobis, e classificador Naïve-Bayes. A distância da amostra de teste é calculada entre as coordenadas da amostra e o centroide de cada grupo no subespaço de componentes principais. O classificador por Naïve-Bayes, por sua vez, considera que as distribuições desses dados no subespaço são do tipo Gaussiana. Observa-se que a distância Euclidiana tem um desempenho bom na classificação, isto é, a classificação aproxima-se de 100%,

Figura 77 – Dados utilizados de uma frequência de estímulo para validação cruzada em PCA.



Fonte: Elaborado pelo autor.

o que não ocorre para a distância de Mahalanobis. Essa diferença significativa pode ser devido à principal desvantagem da distância de Mahalanobis (YEAGER et al., 2018): uma vez que ela é caracterizada pela média das amostras e covariância, qualquer amostra não pertencente ao conjunto de referência de treinamento (*outlier*) apresenta uma influência indevida na estimativa da média e covariância; menores distâncias apresentadas pelas amostras de teste podem mascarar os *outliers* da mesma vizinhança estatística que o conjunto de treinamento, resultando em um maior número de falsos negativos. Em outras palavras, a distância de Mahalanobis é sensível à presença de *outliers*, e considerando que há uma quantidade pequena de amostras nas componentes principais, pode-se justificar o resultado obtido. O classificador Naïve-Bayes, por fim, obteve um resultado similar ao da distância Euclidiana. Pode-se justificar seu desempenho por levar em conta as probabilidades condicionais; ademais, a hipótese de independência condicional entre as componentes principais se demonstra adequada.

Tabela 3 – Porcentagem de acerto para PCA.

Número de Validações	E_c (%)	M_c (%)	NB (%)
20	97,92	49,58	97,92

Fonte: Elaborado pelo autor.

A próxima seção introduz os resultados da Análise de Discriminante Linear/Quadrático

(*Linear/Quadratic Discriminant Analysis*).

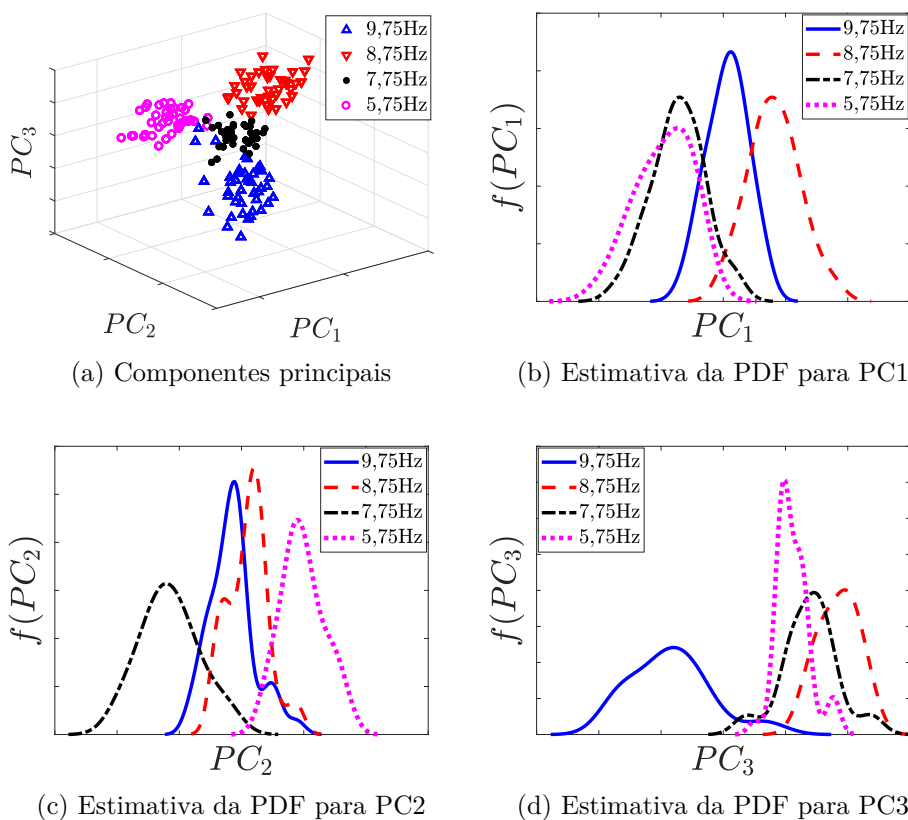
5.4 ANÁLISE DE DISCRIMINANTE LINEAR/QUADRÁTICO (LDA/QDA)

Para a análise desse classificador, adota-se três etapas: a) visualização espacial do efeito da LDA, b) aplicação do discriminante linear e quadrático para a classificação de fato, c) validação cruzada e porcentagem de acerto. Para sua utilização, é necessário que se tenha duas ou mais características distintas a serem comparadas; os dados não são, pois, utilizados em sua forma original, devido ao problema de dimensionalidade: para os dados originais, tem-se um vetor n -dimensional, o que implica em n características; após PCA, tem-se apenas 3 características, compostas pelas componentes principais, o que reduz a dimensionalidade e destaca as informações principais contidas nos sinais originais. Dessa forma, comumente aplica-se uma técnica de redução de dimensionalidade antes de se fazer uso da LDA/QDA, como a PCA, por exemplo. Uma comparação entre as componentes principais antes e depois da LDA é mostrada na próxima seção.

5.4.1 Visualização Espacial do Efeito da LDA

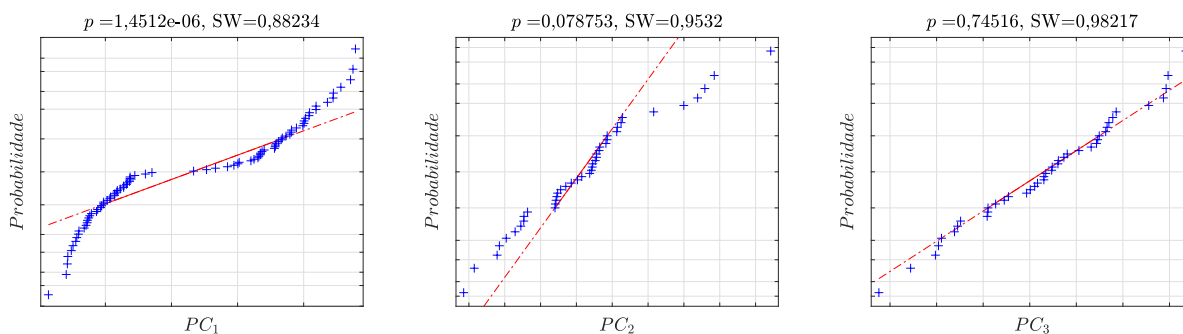
A configuração considerada neste ponto é a mesma da seção anterior: 3 eletrodos e 14 rodadas. A Figura 78 apresenta, além das componentes principais vistas na Figura 76b, as estimativas de função de densidade de probabilidade para cada uma dessas componentes. Antes da aplicação do algoritmo LDA (que requer, a princípio, que os dados sejam normalmente distribuídos, como visto neste trabalho), testa-se novamente a hipótese de aderência à normalidade para tais estimativas, que agora representam os dados em dimensão reduzida (a partir dos quais a classificação ocorre de fato). Embora não contando com as 20 rodadas totais, para fins demonstrativos, a configuração apresentada é considerada representativa. As Figuras 79 a 82 mostram ambos os critérios visual e formal para as 3 componentes principais de cada conjunto de frequências – isto é, considerando a distribuição de probabilidade tridimensional. Nota-se que, para 95% de confiança novamente, o critério é satisfeito para a maioria dos conjuntos, com exceção das componentes PC_1 para 9,75 Hz, 8,75 Hz e 5,75 Hz, o que é suficiente para os fins de teste e comparação deste trabalho.

Figura 78 – Estimativas de função densidade de probabilidade para as componentes principais consideradas.



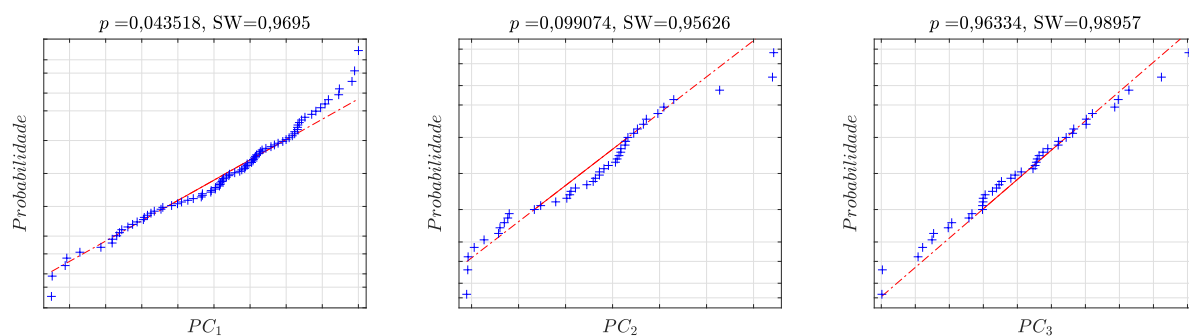
Fonte: Elaborado pelo autor.

Figura 79 – Teste visual e de Shapiro-Wilk para os dados de 9,75 Hz nas componentes principais.



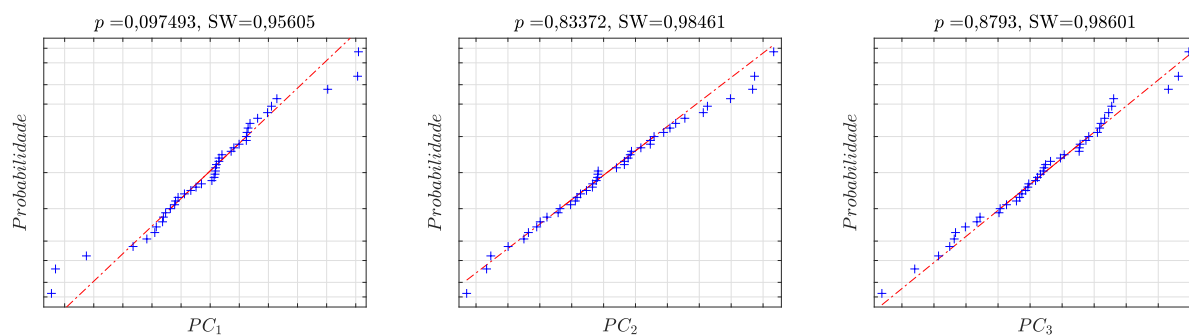
Fonte: Elaborado pelo autor.

Figura 80 – Teste visual e de Shapiro-Wilk para os dados de 8,75 Hz nas componentes principais.



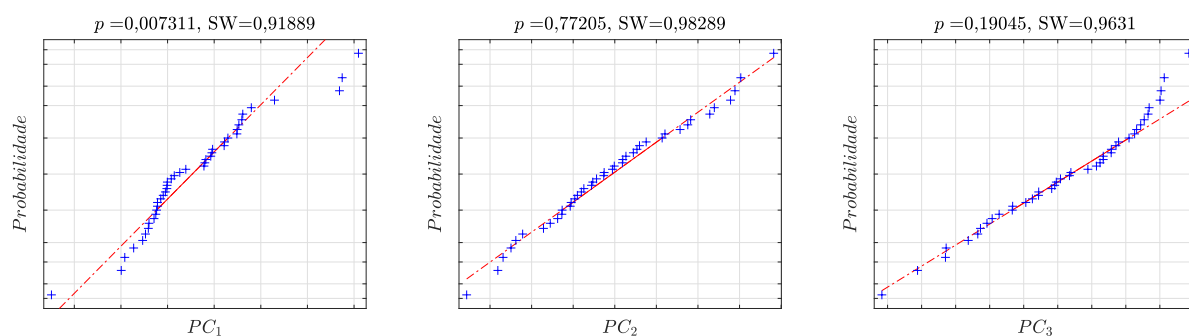
Fonte: Elaborado pelo autor.

Figura 81 – Teste visual e de Shapiro-Wilk para os dados de 7,75 Hz nas componentes principais.



Fonte: Elaborado pelo autor.

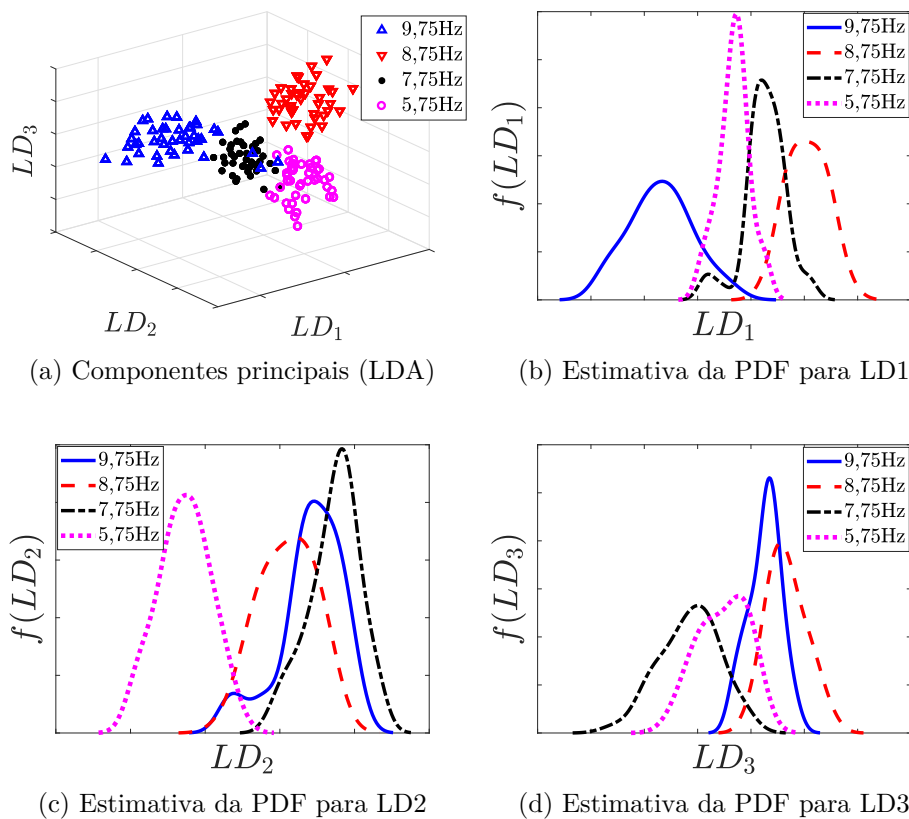
Figura 82 – Teste visual e de Shapiro-Wilk para os dados de 5,75 Hz nas componentes principais.



Fonte: Elaborado pelo autor.

A Figura 83, por sua vez, apresenta esses dados nos novos eixos após a transformação realizada pelo algoritmo LDA, juntamente com as novas estimativas. Vale lembrar que tais eixos são o resultado do processo de otimização que visa à projeção dos dados nos autovetores das componentes principais que melhor separam as distribuições entre os grupos de dados. Observa-se que há, não obstante um certo grau de superposição entre as curvas, uma melhor separação entre as estimativas das funções de densidade de probabilidade, especialmente para a componente LD_1 , representativa do maior autovetor LDA.

Figura 83 – Estimativas de função densidade de probabilidade para as componentes principais após aplicação da LDA.



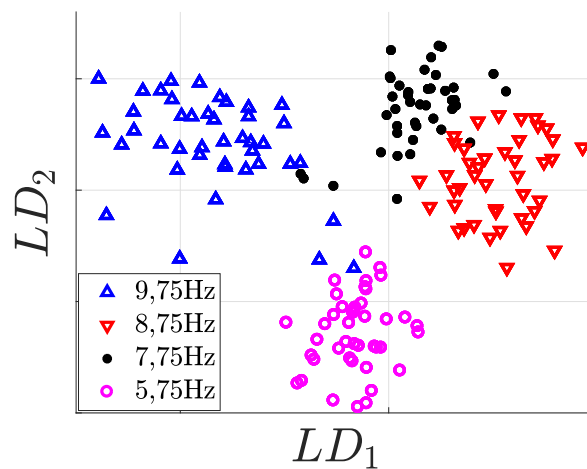
Fonte: Elaborado pelo autor.

A partir dessa mudança de eixos, é possível utilizar uma métrica, como a distância Euclidiana, de Mahalanobis, ou mesmo Naïve-Bayes. Em todo caso, optou-se por dar sequência à técnica como é originalmente definida (capítulo 4); os resultados são apresentados a seguir.

5.4.2 Discriminante Linear e Quadrático

Para os discriminantes linear e quadrático, as duas componentes principais após LDA foram escolhidas, como mostrado na Figura 84; assim, tem-se o equivalente a duas características apenas, e o processo de classificação fica mais acessível. Salienta-se que os casos apresentados abaixo representam apenas um exemplo de como a classificação é realizada. Para o cálculo da porcentagem geral de acerto, a validação cruzada com $k = 20$ é aplicada; desta forma, por conveniência, as figuras abaixo mostram apenas um dos casos ($k = 1$).

Figura 84 – Subespaço de duas dimensões com as componentes principais após LDA.



Fonte: Elaborado pelo autor.

As Figuras 85 a 90 apresentam as curvas obtidas para a classificação entre pares de grupos: à esquerda, LDA; à direita, QDA. As Tabelas 4 e 5, por fim, apresentam os coeficientes de suas respectivas equações gerais entre pares para uma iteração ($k = 1$), dadas a partir da definição da equação 123 vista no capítulo 4, e expressas pelas equações 159 e 160. O intuito é obter um sistema a partir do conjunto de equações lineares ou não lineares e, obedecendo-se às restrições, classificar uma nova amostra adicionada ao subespaço. Observa-se, de maneira geral, que ambas as técnicas demonstram ser eficazes em separar as amostras; não obstante, a QDA tem melhor desempenho, aparentemente, nos casos das Figuras 87 e 88, devido ao aumento no grau do polinômio.

$$\zeta_0 + \zeta_1 LD_1 + \zeta_2 LD_2 = 0 \quad (159)$$

$$\zeta_0 + \zeta_1 LD_1 + \zeta_2 LD_2 + \zeta_3 LD_1^2 + \zeta_4 LD_1 LD_2 + \zeta_5 LD_2^2 = 0 \quad (160)$$

Tabela 4 – Coeficientes para as equações obtidas a partir da LDA ($k = 1$).

Frequências (Hz)	ζ_0	ζ_1	ζ_2
9,75 e 8,75	3,3150	0,2267	0,0226
9,75 e 7,75	3,7953	0,1578	0,0640
9,75 e 5,75	-0,1842	0,0842	-0,2033
8,75 e 7,75	1,8554	-0,1588	0,1618
8,75 e 5,75	-1,5907	-0,2191	-0,1897
7,75 e 5,75	-5,0056	0,0364	-0,3674

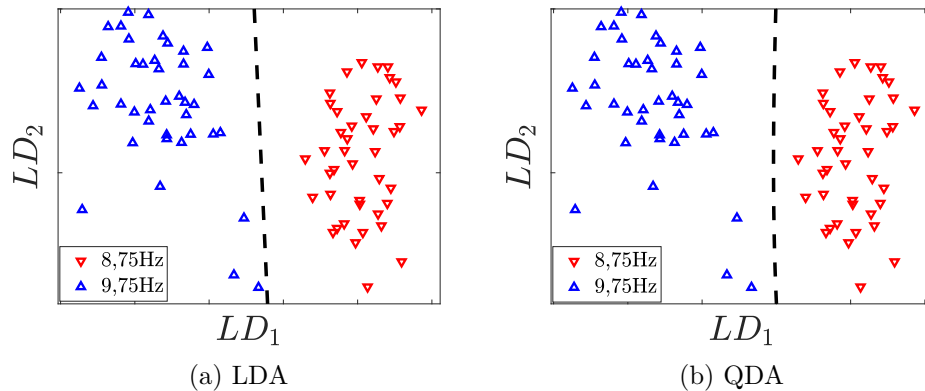
Fonte: Elaborado pelo autor.

Tabela 5 – Coeficientes para as equações obtidas a partir da QDA ($k = 1$).

Frequências (Hz)	ζ_0	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5
9,75 e 8,75	0,4874	0,2548	0,0037	-0,0008	0,0013	-0,0002
9,75 e 7,75	2,1972	0,0649	0,1637	-0,0011	0,0030	-0,0011
9,75 e 5,75	-2,2314	0,0688	-0,2035	-0,0019	0,0016	-0,0004
8,75 e 7,75	-2,2314	0,0688	-0,2035	-0,0019	0,0016	-0,0004
8,75 e 5,75	-2,7189	-0,1860	-0,2072	-0,0010	0,0003	-0,0002
7,75 e 5,75	-4,4286	0,0040	-0,3673	-0,0007	-0,0014	0,0007

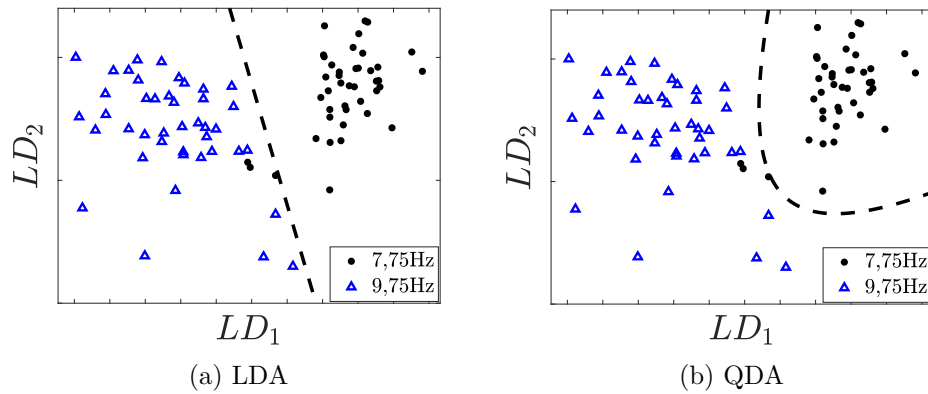
Fonte: Elaborado pelo autor.

Figura 85 – Aplicação de LDA e QDA para 9,75 Hz e 8,75 Hz.



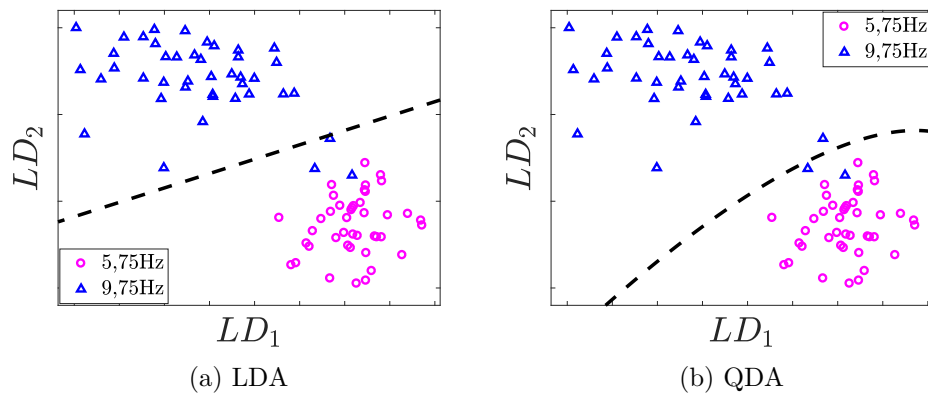
Fonte: Elaborado pelo autor.

Figura 86 – Aplicação de LDA e QDA para 9,75 Hz e 7,75 Hz.



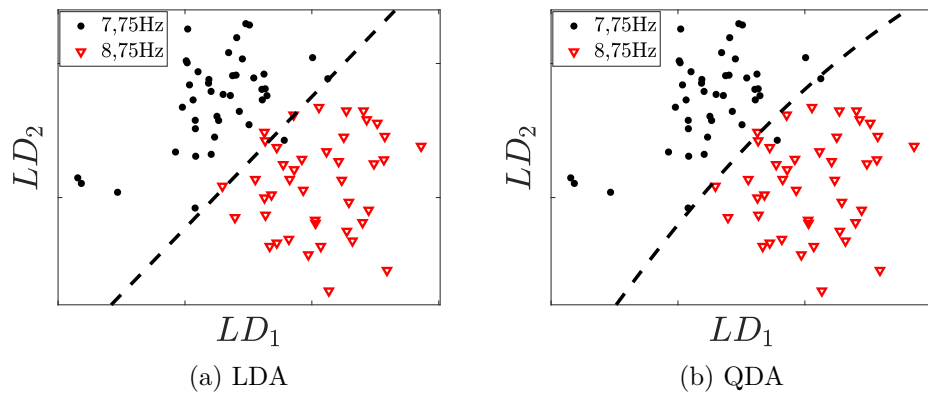
Fonte: Elaborado pelo autor.

Figura 87 – Aplicação de LDA e QDA para 9,75 Hz e 5,75 Hz.



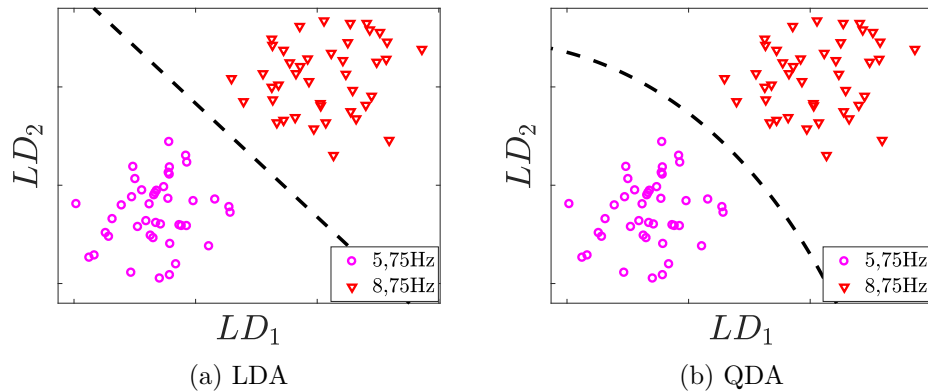
Fonte: Elaborado pelo autor.

Figura 88 – Aplicação de LDA e QDA para 8,75 Hz e 7,75 Hz.



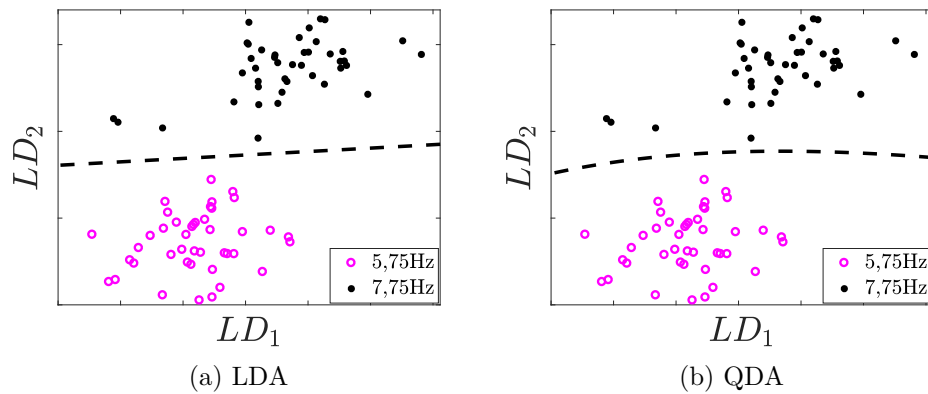
Fonte: Elaborado pelo autor.

Figura 89 – Aplicação de LDA e QDA para 8,75 Hz e 5,75 Hz.



Fonte: Elaborado pelo autor.

Figura 90 – Aplicação de LDA e QDA para 7,75 Hz e 5,75 Hz.



Fonte: Elaborado pelo autor.

A obtenção dos polinômios encerra essa etapa da análise. O próximo passo é realizar, de fato, a classificação das amostras do conjunto de testes, que é apresentada na próxima seção.

5.4.3 Validação Cruzada e Porcentagem de Acerto

Para a análise de porcentagem de acerto, utiliza-se o sistema obtido a partir das equações de fronteira da seção anterior, tanto para LDA quanto QDA. Seguindo as mesmas condições anteriores, aplica-se a validação cruzada utilizando $k = 20$. Para a classificação ser considerada correta, assume-se que todas as restrições devem ser obedecidas simultaneamente. Isto é, para cada frequência de estímulo, os 3 pares de conjuntos de dados obtidos devem acusar a mesma frequência reconhecida. A Tabela 6 apresenta as porcentagens de acerto obtidas pelas duas técnicas. Nota-se que, apesar da aparente superioridade

da QDA, ela não fornece melhoras significativas no resultado final nesta aplicação. Não obstante, as porcentagens são elevadas, ligeiramente superiores às obtidas por PCA.

Tabela 6 – Porcentagem de acerto para LDA e QDA.

Número de Validações	LDA (%)	QDA (%)
20	98,19	98,47

Fonte: Elaborado pelo autor.

5.5 MÁQUINA DE VETORES DE SUPORTE (SVM)

Para a análise desse classificador, apenas uma etapa é considerada: validação cruzada e análise de porcentagem de acerto. Como é recomendável se realizar redução de dimensionalidade dos dados antes de se aplicar SVM, especialmente quando há um banco de dados de tamanho considerável, essa etapa também é realizada, como visto nas seções anteriores; de modo que, para os testes a seguir, as amostras são extraídas da configuração apresentada pela Figura 76b, após aplicação de PCA. No presente trabalho, opta-se por empregar a abordagem 1V1, semelhante ao realizado na seção sobre LDA e QDA. A Tabela 7 apresenta as porcentagens de acerto obtidas para SVM, após validação cruzada para $k = 20$, considerando quatro tipos de funções *kernel*, como comentado no capítulo 4: linear, polinomial, sigmoide e gaussiano (ou exponencial, *Radial Basis Function*). Os parâmetros SVM são ajustados para $d = 0$, $\gamma = 1$, escolhas normalmente feitas, como visto, e $n = 3$ para a ordem do *kernel* polinomial, devido ao fato de que, para graus elevados do polinômio, ou mesmo para γ muito baixos, observa-se baixas porcentagens de acerto, podendo causar superajuste (*overfitting*) ou desempenho sub-ótimo (BEN-HUR et al., 2008). Observa-se que, de modo geral, as classificações são eficientes, e o *kernel* que fornece a melhor delas é o Gaussiano.

Tabela 7 – Porcentagem de acerto para SVM.

Número de Validações	Linear (%)	Sigmoide (%)	Polinomial (%)	Gaussiano (%)
20	93,75	87,50	93,75	97,92

Fonte: Elaborado pelo autor.

5.6 COMPARAÇÃO ENTRE CLASSIFICADORES

Esta seção compara o desempenho dos classificadores especialmente em relação ao principal classificador deste trabalho: CCA. Dessa forma, é possível fazer uma análise crítica de sua aplicação, visando a melhor utilizá-lo, e concluir, de fato, se as vantagens superam as desvantagens para a aplicação em estudo.

Primeiramente, de modo a resumir os resultados, apresenta-se a Tabela 8, onde constam as porcentagens gerais de acerto para todas as técnicas empregadas neste trabalho, ordenadas do melhor para o pior desempenho. Nota-se que a CCA clássica obteve a segunda menor porcentagem de acerto; todavia, salienta-se que, para os outros classificadores, apenas os eletrodos mais significativos foram escolhidos durante a etapa de redução de dimensionalidade (PCA) e aplicados às técnicas seguintes, enquanto que todos os eletrodos foram mantidos para a CCA. Também, considerando que o banco comparativo de sinais da CCA é totalmente artificial, considera-se que sua porcentagem é relativamente alta. Esse fato é discutido com mais detalhes a seguir.

Tabela 8 – Comparação geral de porcentagem de acerto entre classificadores.

Posição	Classificador	Acerto geral (%)
1º	QDA	98,47
2º	LDA	98,19
3º	PCA (Naïve-Bayes)	97,92
3º	PCA (Distância Euclidiana)	97,92
3º	SVM (<i>kernel</i> Gaussiano)	97,92
4º	SVM (<i>kernel</i> linear)	93,75
4º	SVM (<i>kernel</i> polinomial)	93,75
5º	SVM (<i>kernel</i> sigmoide)	87,50
6º	CCA	83,75
7º	PCA (Distância de Mahalanobis)	49,58

Fonte: Elaborado pelo autor.

Em seguida, para se comparar o resultado obtido a partir do banco de dados utilizado neste trabalho (ZHANG et al., 2014) com outros da literatura, apresenta-se a Tabela 9. Nela, configurações próximas (especialmente quanto à janela temporal) para a execução do protocolo foram utilizadas, a saber: 4 frequências de estímulo e 4 segundos de aquisição de sinal (ZHANG et al., 2013), 5 frequências de estímulo e 4 segundos de aquisição (WEI et al., 2013), 7 frequências de estímulo e 4 segundos de aquisição (HAKVOORT; REUDERINK; OBBINK, 2011), e 12 frequências de estímulo e 4 segundos de aquisição (NAKANISHI et al., 2015). Salienta-se que alguns resultados foram apresentados apenas

de forma gráfica no trabalho original, de forma que estes são denotados como aproximados na tabela. Nota-se que os dados utilizados no presente trabalho estão próximos dos encontrados na literatura, apresentando, em comparação, bom resultado.

Tabela 9 – Comparação de porcentagem de acerto entre CCAs de trabalhos distintos.

Posição	Trabalho	Acerto geral (%)
1 ^o	Nakanishi et al. (2015)	~ 90
2^o	Zhang et al. (2014)	83,75
3 ^o	Wei et al. (2013)	~ 80
4 ^o	Zhang et al. (2013)	76,6
5 ^o	Hakvoort, Reuderink e Obbink (2011)	76,3

Fonte: Elaborado pelo autor.

Por fim, uma análise do desempenho geral, vantagens e desvantagens dos classificadores utilizados neste trabalho são apresentados a seguir.

Análise de Componentes Principais (PCA)

A partir da PCA, é possível visualizar, de forma geral, a influência da posição dos eletrodos na disposição final do subespaço de componentes principais: quanto mais próximos estão da região de atividade neural senho empenhada (córtex visual, região occipital), maior é a separabilidade resultante entre os grupos originados por amostras de frequências distintas, principalmente ao se considerar mais rodadas. A distância Euclidiana e o classificador de Naïve-Bayes se mostram eficientes na classificação, enquanto que a distância de Mahalanobis tem um desempenho inferior.

Apesar dos resultados promissores, as desvantagens do uso da PCA são, essencialmente, três: i) não é um classificador por definição, e a escolha das componentes principais (o valor de m , neste caso) pode ser uma tarefa difícil, visto que as direções de maior variabilidade dos dados, se são muitas, nem sempre podem fornecer subsídios visuais para a interpretação destes; ii) a escolha da métrica de classificação pode ser um complicador, e deve ser feita com cautela, não sendo passível de generalização, como visto nos resultados; iii) por fim, o treinamento, isto é, a obtenção de um grupo de dados de treino, é uma etapa indispensável, o que implica na necessidade de validação de dados e modelo.

Análise de Discriminante Linear/Quadrático (LDA/QDA)

Com a aplicação da LDA/QDA, por sua vez, obtém-se neste um efeito interessante de separabilidade relacionada às distribuições, o que, a princípio, fornece maiores subsídios para uma classificação mais criteriosa. Por definição, LDA/QDA são, ambos, classificadores estatísticos, e forneceram bons resultados a partir da porcentagem de classificação.

Não obstante, especialmente quando há um número expressivo de dados, também é necessária uma etapa de redução de dimensionalidade antes de sua aplicação, caso contrário o tratamento das amostras, embora possível, torna-se impraticável; também, o treinamento do classificador é uma etapa indispensável. Ademais, ainda que mais criterioso a priori, LDA e QDA tiveram um desempenho apenas ligeiramente superior ao da PCA com uso de distância Euclidiana/Naïve-Bayes, o que evidencia que maior complexidade — mesmo com efeitos visual e qualitativamente ricos em informação — não necessariamente implica em melhores resultados.

Máquina de Vetores de Suporte (SVM)

O SVM é, da mesma forma, um classificador por definição. Sua concepção, equacionamento e aplicação são bem fundamentados e forneceu bons resultados na classificação dos dados, especialmente para o caso do *kernel* Gaussiano. Em todo caso, além de exigir, normalmente, redução de dimensionalidade, é de maior complexidade computacional e exige etapa de treinamento. Além disso, a escolha da função *kernel* em si é um complicador, gerando resultados bons ou ruins para os mesmos conjuntos de amostras. Ou seja, é uma técnica de alta eficiência devido à sua robustez de formulação, porém exige conhecimento e domínio maiores — principalmente para se aplicar a formulação multiclasse — que nem sempre é conveniente.

Análise de Correlação Canônica (CCA)

A CCA, por fim, é um classificador por definição. De forma geral, fornece boas porcentagens de acerto na identificação da frequência de estímulo para os casos abordados neste trabalho. Esse resultado merece especial destaque uma vez que a comparação entre os sinais provenientes do sistema BCI ocorre com um banco de sinais de senos e cossenos criados artificialmente. Além disso, não requer etapa de redução de dimensionalidade (os dados aplicados na técnica são utilizados em sua forma original, P -dimensional, onde P é o número de pontos por sinal), conhecimento prévio sobre a distribuição de probabilidade

dos dados (como se trata de uma correlação entre dois sinais, ou comparação direta, basta haver similaridade entre os comportamentos destes), ou mesmo um pré-processamento sofisticado. Essas vantagens, em conjunto, fornecem uma classificação simples, rápida e precisa. Com a generalização proposta, soma-se a essas vantagens a obtenção de informações quantitativa e qualitativamente importantes sobre o sistema BCI, bem como novos parâmetros que ajudam a melhor compreender quais fatores influenciam na degradação do coeficiente de correlação e são capazes de auxiliar na escolha mais criteriosa das frequências de estímulo, compreendendo fatores fisiológicos, técnicos e operacionais.

Por outro lado, como visto neste trabalho, há variações desta técnica que fornecem melhores porcentagens de acerto (NAKANISHI et al., 2015), especialmente aquelas que consideram o banco de comparação de sinais após calibração, isto é, após aquisição preliminar de sinais reais e construção do banco a partir destes (ZHANG et al., 2014). Não obstante, considera-se que a CCA clássica, juntamente com as informações agregadas a partir de sua generalização, é viável e vantajosa por dois motivos além dos já mencionados: i) as variações dessa técnica apresentam maior complexidade, seja pelos conceitos matemáticos aplicados, seja pela necessidade de incorporar uma etapa a mais no processo (no caso, construção do banco de comparação com sinais reais); ii) embora um banco composto por sinais reais (e, portanto, com características que representam mais fielmente a atividade neural real existente, levando em conta, inclusive, as peculiaridades do EEG de pessoa para pessoa) possa ser mais eficiente, a consideração de que a CCA clássica atua como um filtro, buscando correlações implícitas com componentes harmônicas contidas no EEG real e resultantes da estimulação visual, é aplicável a todos os EEG e todos os indivíduos.

Finalmente, no que tange à capacidade de classificação em si da CCA, embora ela possa ser significativamente deteriorada pela interferência de atividades cerebrais espontâneas contidas no EEG (NAKANISHI et al., 2015), ou simplesmente por ruídos de qualquer natureza, o presente trabalho mostra que um maior cuidado a todos os aspectos que envolvem o sistema BCI pode compensar a simplicidade da técnica, e que a classificação incorreta é expressivamente influenciada pelo modo como o protocolo em si é executado, não apenas por menor robustez de técnica. Por exemplo, a partir da generalização, pode-se estabelecer a melhor faixa de operação (considerando limitações de origem fisiológica) e aferir previamente a porcentagem de acurácia resultante caso uma frequência de estímulo seja escolhida, o que até então não é colocado em prática. Também, deve-se estabelecer um número de componentes harmônicas no banco artificial igual ao que se espera obter (dada a dificuldade de se saber quantas componentes harmônicas

existem em um sinal EEG, neste caso estabelece-se um número aproximadamente igual às harmônicas de maior intensidade, $n = 3$), com o objetivo de minimizar a degradação do coeficiente de correlação, já afetado pelo ruído presente.

A partir da matriz de confusão, endossa-se o fato de que frequências menores implicam em menor porcentagem de acerto na classificação; alternativamente, conciliando o fato de se evitar uma faixa onde a chance de uma crise epilética é maior com a necessidade de se ter frequências altas e de amplitudes altas o suficiente, deve-se se distanciar o mínimo possível da faixa alpha. A partir das curvas ROC, tem-se que, quanto maior a diferença entre uma frequência de estímulo escolhida e outra, maiores são as chances de classificação correta. Por fim, ainda há de se considerar que classificações incorretas foram feitas devido à influência indevida que um estímulo causou em outro enquanto ambos eram apresentados simultaneamente. Os estímulos simultâneos são necessários para se aumentar os graus de liberdade do sistema BCI, porém suas frequências moduladas devem ser analisadas cautelosamente antes da aplicação.

6 CONSIDERAÇÕES FINAIS

O presente trabalho tem como objetivo central o estudo detalhado da técnica de classificação estatística de dados Análise de Correlação Canônica, com o intuito de se obter uma generalização e maior flexibilidade ao se utilizá-la, bem como realizar uma análise e comparação críticas entre ela e outras técnicas comumente aplicadas, como PCA, LDA, QDA e SVM, para avaliar aplicações em sistemas BCI.

A abordagem analítica com inclusão de parâmetros de processamento de sinais permite uma análise individual de cada parâmetro considerado importante no sistema, fornecendo subsídios para o propósito apresentado e verificando a influência no resultado final de correlação. Quanto ao número de períodos, é possível estabelecer relações de tempo de resposta e precisão de reconhecimento da técnica para janelas de tempo variáveis; para a adição de componentes harmônicas, visualiza-se seu efeito no diagrama f_r vs ρ , confirmando o resultado esperado pela equação generalizada CCA, e posteriormente explicitando informações sobre o sistema BCI inicialmente não observadas no domínio da frequência, por exemplo, especialmente no caso de sinais reais; a razão de frequências de amostragem é incluída na formulação clássica da CCA como um fator de correção, o que permite flexibilização da técnica, tornando possível sua aplicação nos casos onde sinais de entrada e banco de sinais não têm a mesma taxa de amostragem; o efeito do ruído se mostra relevante na diminuição do valor de correlação, com maior severidade quando o número de harmônicas entre os grupos de sinais é distinto.

Para o caso de sinais reais, por fim, é possível confirmar a adequabilidade do modelo proposto de superposição de funções senoidais, o que é evidenciado pela boa concordância entre curvas analíticas e reais; além disso, obtém-se informações essenciais sobre o sistema BCI, como a característica não-linear do comportamento neural e a influência de ruído ou atividades cerebrais de fundo que não estão diretamente ligadas aos SSVEP, reduzindo, portanto, a amplitude dos picos de correlação, ou simplesmente adicionado aleatoriedade às curvas.

Finalmente, a partir dos resultados de classificação e porcentagem de acerto, verifica-se que, devido ao conjunto de fatores como necessidade de redução de dimensionalidade, treinamento a partir de banco de dados, e etapas de pré-processamento, a CCA se mostra a opção mais acessível, por conta de sua simplicidade e eficiência (mesmo em detrimento de

uma diminuição na porcentagem de acerto), para os casos estudados no presente trabalho. Ademais, demonstra-se extremamente flexível, seja para propósitos de reconhecimento ou mesmo para mensurar a concordância entre duas curvas.

A proposta de generalização apresentada neste trabalho, aliada às porcentagens de acerto e vantagens intrínsecas à CCA, colabora para avaliar estratégias de futuras aplicações em sistemas BCI reais *online* utilizando o protocolo SSVEP, tornando possível a integração desses sistemas a pacientes e, conseqüentemente, restabelecendo sua autonomia motora. Os resultados obtidos endossam a necessidade da indissociabilidade entre a engenharia e outras áreas do conhecimento para a melhora na qualidade de vida das pessoas.

6.1 CONTRIBUIÇÕES ALCANÇADAS

Com a conclusão deste trabalho, tem-se as seguintes contribuições:

- propôs-se uma abordagem alternativa para sistemas de interface cérebro-máquina usando potenciais visuais evocados, considerando uma generalização do coeficiente de correlação, conceito-base para a análise de correlação canônica;
- a partir dessa generalização, foi possível verificar quais fatores influenciam de forma mais expressiva no valor final de correlação e, conseqüentemente, no reconhecimento correto de padrões eletroencefalográficos;
- foram introduzidos novos parâmetros que colaboram para o melhor reconhecimento de padrões e melhores práticas de uso em sistemas ICM, incluindo fatores fisiológicos, técnicos e operacionais;
- foi possível realizar uma avaliação do desempenho da CCA perante outras técnicas de classificação no que diz respeito ao protocolo SSVEP;
- apresentou-se uma avaliação crítica, bem como subsídios, em termos de análise de sinais, para viabilizar a capacidade motora de pessoas com restrições a partir de protocolo SSVEP e generalização da técnica CCA.

6.2 LIMITAÇÕES DO ESTUDO

Complementar às contribuições alcançadas, tem-se que as limitações do estudo realizado são as seguintes:

- o protocolo SSVEP, apesar de apresentar facilidade de execução, sinais de amplitude expressiva e de relativa fácil obtenção, é complexo, não podendo ser generalizado a todos os indivíduos (cada sinal EEG apresenta particularidades dependendo do usuário do sistema ICM);
- desta forma, apesar das porcentagens de acerto relativamente altas verificadas, não se pode garantir que a aplicação dessas técnicas tenha desempenho alto para todos os sinais e conjuntos de dados, de modo que a calibração e treinamento (quando a técnica exigir, como no caso de PCA, por exemplo) são indispensáveis para todos os indivíduos;
- a intenção do usuário do sistema ICM está sujeita às limitações impostas pelo protocolo SSVEP; assim, apesar de ser útil e promover restabelecimento de autonomia a pacientes com restrições motoras severas, esse protocolo é restrito ao número de graus de liberdade estipulados pelos estímulos visuais (que são um complicador na escolha da seleção de frequências, como visto), e não contempla, portanto, a atividade imaginativa livre do indivíduo. Isso torna o protocolo mais apropriado a situações que envolvam dispositivos de escolha de ações simples;
- além da restrição dos graus de liberdade, salienta-se que, por se tratar de estímulos visuais, mesmo operando em uma faixa de frequências mais adequada e criteriosa, cujo risco de desencadeamento de crises epiléticas é menor (e melhor, em termos de boas amplitudes), pode haver desconforto e fadiga visual durante a execução do protocolo.

6.3 SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestões para trabalhos futuros, tem-se:

- aplicar a técnica de Análise de Correlação Canônica — considerando todo o conhecimento prévio adquirido e a proposta de generalização apresentada neste trabalho — a novos sinais obtidos a partir de experimentos, seja *offline*, seja *online*;
- explorar variações da Análise de Correlação Canônica buscando, também, melhor compreendê-las e aplicá-las a partir dos mesmos conceitos empregados neste trabalho;

- avaliar a integração de sujeito e sistema BCI para o caso de um atuador mecânico, como cadeira de rodas motorizada, por exemplo, utilizando o protocolo SSVEP para realização de tarefas distintas e, portanto, com diferentes graus de liberdade;
- realizar um estudo sobre o número máximo, ou o número mais viável, de graus de liberdade possíveis de serem obtidos a partir do sistema BCI com uso do protocolo SSVEP integrado a um atuador mecânico e aplicação da CCA, considerando as contribuições obtidas no presente trabalho.

REFERÊNCIAS

- AGGARWAL, C. C. Outlier analysis. In: SPRINGER. **Data mining**. [S.l.], 2015. p. 237--263.
- ALPAYDIN, E. **Introduction to machine learning**. [S.l.]: MIT press, 2009.
- ANDERSON, T. W.; ANDERSON, T. W.; ANDERSON, T. W.; ANDERSON, T. W.; MATHÉMATICIEN, E.-U. **An Introduction to Multivariate statistical analysis**. New York: Wiley, 1958.
- BAUER, G.; GERSTENBRAND, F.; RUMPL, E. Varieties of the locked-in syndrome. **Journal of Neurology**, Heidelberg, v. 221, n. 2, p. 77--91, Aug 1979. ISSN 1432-1459.
- BAZARAA, M. S.; JARVIS, J. J.; SHERALI, H. D. **Linear programming and network flows**. New Jersey: John Wiley & Sons, 2011.
- BEN-HUR, A.; ONG, C. S.; SONNENBURG, S.; SCHÖLKOPF, B.; RÄTSCH, G. Support vector machines and kernels for computational biology. **PLOS Computational Biology**, San Francisco, v. 4, n. 10, p. 1--10, 10 2008.
- BERTSEKAS, D. P. **Constrained optimization and Lagrange multiplier methods**. London: Academic press, 2014.
- BEVERINA, F.; PALMAS, G.; SILVONI, S.; PICCIONE, F.; GIOVE, S. et al. User adaptive BCIs: SSVEP and P300 based interfaces. **PsychNology Journal**, Citeseer, v. 1, n. 4, p. 331--354, 2003.
- BHATTACHARYA, B.; HABTZGHI, D. Median of the p value under the alternative hypothesis. **The American Statistician**, Alexandria, v. 56, n. 3, p. 202--206, 2002.
- BIN, G.; GAO, X.; YAN, Z.; HONG, B.; GAO, S. An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. **Journal of Neural Engineering**, Bristol, v. 6, n. 4, p. 046002, 2009.
- BISHOP, C. M. **Pattern recognition and machine learning**. London: Springer, 2006.
- BREDENSTEINER, E. J.; BENNETT, K. P. Multicategory classification by support vector machines. In: **Computational Optimization**. London: Springer, 1999. p. 53--79.
- BROGIN, J. A. F. **Análise de sinais neurais para interface cérebro-máquina**. [S.l.: s.n.], 2016.
- BRONZINO, J. D. **Biomedical engineering handbook**. Boca Raton: CRC press, 1999.

- CALHOUN, G. L.; MCMILLAN, G. R. EEG-based control for human-computer interaction. In: ANNUAL SYMPOSIUM ON HUMAN INTERACTION WITH COMPLEX SYSTEMS, 3., 1996, [S. l.]. **Proceedings [...]**. Piscataway: IEEE, 1996. p. 4--9.
- CHAUDHARY, U.; BIRBAUMER, N.; RAMOS-MURGUIALDAY, A. Brain-computer interfaces for communication and rehabilitation. **Nature Reviews Neurology**, London, v. 12, n. 9, p. 513, 2016.
- CHEN, J.; ZHANG, D.; ENGEL, A. K.; GONG, Q.; MAYE, A. Application of a single-flicker online SSVEP BCI for spatial navigation. **PloS One**, San Francisco, v. 12, n. 5, p. e0178385, 2017.
- CHEN, X.; WANG, Y.; NAKANISHI, M.; JUNG, T.; GAO, X. Hybrid frequency and phase coding for a high-speed ssvep-based bci speller. In: ANNUAL INTERNATIONAL CONFERENCE OF THE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY(EMBC), 36th., 2014, [S. l.]. **Proceedings [...]**. Piscataway: IEEE, 2014. p. 3993--3996. ISSN 1094-687X.
- CHOU, S. M.; NORRIS, F. H. Issues & opinions: Amyotrophic lateral sclerosis: Lower motor neuron disease spreading to upper motor neurons. **Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine**, Hoboken, v. 16, n. 8, p. 864--869, 1993.
- CHUMERIN, N.; MANYAKOV, N. V.; VLIET, M. V.; ROBBEN, A.; COMBAZ, A.; HULLE, M. V. Preprocessing and decoding steady-state visual evoked potentials for brain-computer interfaces. In: DURO, R.; PENA, F. L. **Digital image and signal processing for measurement systems**. [S.l.]: River, 2012. p. 1--33.
- COMMONS, C. **Attribution 4.0 International**. [S.l.], 2018. Disponível em: <http://creativecommons.org/licenses/by/4.0/>. Acesso em: 22 out. 2018.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, New York, v. 20, n. 3, p. 273--297, 1995.
- CRAMER, S. C.; BENSON, R. R.; KAPLAN, J. D.; PARKER, R. A.; KWONG, K. K.; KENNEDY, D. N.; FINKLESTEIN, S. P.; ROSEN, B. R. et al. A functional MRI study of subjects recovered from hemiparetic stroke. **Stroke**, Waltham, 1997.
- CRAMMER, K.; SINGER, Y. On the algorithmic implementation of multiclass kernel-based vector machines. **Journal of Machine Learning Research**, Brookline, v. 2, p. 265--292, mar. 2002. ISSN 1532-4435.
- CURRAN-EVERETT, D.; BENOS, D. J. **Guidelines for reporting statistics in journals published by the American Physiological Society**. [S.l.]: American Physiological Society, 2004.
- DOKKUM, L. V.; WARD, T.; LAFFONT, I. Brain-computer interfaces for neurorehabilitation--its current status as a rehabilitation strategy post-stroke. **Annals of Physical and Rehabilitation Medicine**, Issy les Moulineaux Cedex, v. 58, n. 1, p. 3--8, 2015.

- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. New York: John Wiley & Sons, 2012.
- EVA, O. D.; LAZAR, A. M. Comparison of classifiers and statistical analysis for EEG signals used in brain computer interface motor task paradigm. **International Journal of Advanced Research in Artificial Intelligence on IJARAI**, Citeseer, v. 4, n. 1, p. 8--12, 2015.
- FAUZAN, N.; AMRAN, N. H. Brain waves and connectivity of autism spectrum disorders. **Procedia - Social and Behavioral Sciences**, Amsterdam, v. 171, p. 882--890, 2015.
- FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, Amsterdam, v. 27, n. 8, p. 861--874, 2006.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, Chichester, v. 7, n. 2, p. 179--188, 1936.
- FISHER, R. S.; HARDING, G.; ERBA, G.; BARKLEY, G. L.; WILKINS, A. Photic-and pattern-induced seizures: a review for the Epilepsy Foundation of America Working Group. **Epilepsia**, Hoboken, v. 46, n. 9, p. 1426--1441, 2005.
- FRIEDMAN, J. H. Regularized discriminant analysis. **Journal of the American Statistical Association**, Alexandria, v. 84, n. 405, p. 165--175, 1989.
- FUSHIKI, T. Estimation of prediction error by using k-fold cross-validation. **Statistics and Computing**, New York, v. 21, n. 2, p. 137--146, 2011.
- GAREY, L. J. **Brodmann's Localisation in the Cerebral Cortex**. New Jersey: World Scientific, 2006.
- GHASEMI, A.; ZAHEDIASL, S. Normality tests for statistical analysis: a guide for non-statisticians. **International Journal of Endocrinology and Metabolism**, Iran, v. 10, n. 2, p. 486, 2012.
- GOLDBERGER, A. L.; AMARAL, L. A. N.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. **Circulation**, Waltham, v. 101, n. 23, p. e215--e220, 2000.
- GRAIMANN, B.; ALLISON, B.; PFURTSCHELLER, G. **Brain-Computer interfaces revolutionizing human-computer interaction (The Frontiers Collection)**. New York: Springer, 2010.
- GUERMEUR, Y. Combining discriminant models with new multi-class SVMs. **Pattern Analysis & Applications**, New York, v. 5, n. 2, p. 168--179, 2002.
- HAKVOORT, G.; REUDERINK, B.; OBBINK, M. Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system. **Centre for Telematics & Information Technology University of Twente**, Citeseer, 2011.

- HARMONY, T. The functional significance of delta oscillations in cognitive processing. **Frontiers in Integrative Neuroscience**, Frontiers, v. 7, p. 83, 2013.
- HATEM, S. M.; SAUSSEZ, G.; FAILLE, M. della; PRIST, V.; ZHANG, X.; DISPA, D.; BLEYENHEUFT, Y. Rehabilitation of motor function after stroke: a multiple systematic review focused on techniques to stimulate upper extremity recovery. **Frontiers in Human Neuroscience**, Frontiers, v. 10, p. 442, 2016.
- HERRMANN, C.; DEMIRALP, T. Human EEG gamma oscillations in neuropsychiatric disorders. **Clinical Neurophysiology**, Shannon, v. 116, n. 12, p. 2719--2733, 2005.
- HERRMANN, C. S. Human EEG responses to 1-100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. **Experimental Brain Research**, New York, v. 137, n. 3-4, p. 346--353, 2001.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, Washington, v. 24, n. 6, p. 417, 1933.
- _____. Relations between two sets of variates. **Biometrika**, Oxford, v. 28, n. 3/4, p. 321--377, 1936.
- HUANG, V. S.; KRAKAUER, J. W. Robotic neurorehabilitation: a computational motor learning perspective. **Journal of Neuroengineering and Rehabilitation**, London, v. 6, n. 1, p. 5, 2009.
- ISA, N. E. M.; AMIR, A.; ILYAS, M. Z.; RAZALLI, M. S. The performance analysis of K-nearest neighbors (K-NN) algorithm for motor imagery classification based on EEG signal. In: INTERNATIONAL CONFERENCE ON EMERGING ELECTRONIC SOLUTIONS FOR IOT (ICEESI), 2017, [S. l.]. **Proceedings [...]**. [S.I.]: MATEC Web Conf, 2017. v. 140, p. 1024.
- İŞCAN, Z.; NIKULIN, V. V. Steady state visual evoked potential (SSVEP) based brain-computer interface (BCI) performance under different perturbations. **PloS one**, San Francisco, v. 13, n. 1, p. e0191673, 2018.
- IZHIKEVICH, E. M. **Dynamical systems in neuroscience**. Cambridge: MIT press, 2007.
- JASPER, H. H. The ten-twenty electrode system of the international federation. **Electroencephalography and Clinical Neurophysiology**, Shannon, v. 10, p. 370--375, 1958.
- JÁUREGUI-LOBERA, I. Electroencephalography in eating disorders. **Neuropsychiatric Disease and Treatment**, Auckland, v. 8, p. 1, 2012.
- JOLLIFFE, I. Principal component analysis. In: **International encyclopedia of statistical science**. [S.l.]: Springer, 2002. p. 1094--1096.
- KALUNGA, E.; DJOUANI, K.; HAMAM, Y.; CHEVALLIER, S.; MONACELLI, E. SSVEP enhancement based on canonical correlation analysis to improve BCI performances. In: AFRICON, 2014, Pointe-Aux-Piments. **Proceedings [...]**. Piscataway: IEEE, 2013. p. 1--5. ISSN 2153-0033.

- KHAN, L.; AWAD, M.; THURASINGHAM, B. A new intrusion detection system using support vector machines and hierarchical clustering. **The VLDB journal**, New York, v. 16, n. 4, p. 507--521, 2007.
- KIM, T.-H.; WHITE, H. On more robust estimation of skewness and kurtosis. **Finance Research Letters**, Maryland Heights, v. 1, n. 1, p. 56--73, 2004.
- KLIMESCH, W. Alpha-band oscillations, attention, and controlled access to stored information. **Trends in Cognitive Sciences**, Oxford, v. 16, n. 12, p. 606--617, 2012.
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. **Ijcai**, Montreal, v. 14, n. 2, p. 1137--1145, 1995.
- KOTZ, S.; SEIER, E. Kurtosis of the top-levine distributions. **Interstat**, [S. I.], v. 1, p. 1--15, 2007.
- KRESSEL, U. **Pairwise classification and support vector machines**. Cambridge: MIT Press, 1999.
- KWAK, N.-S.; MÜLLER, K.-R.; LEE, S.-W. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. **PloS one**, San Francisco, v. 12, n. 2, p. e0172578, 2017.
- LAFLEUR, K.; CASSADY, K.; DOUD, A.; SHADES, K.; ROGIN, E.; HE, B. Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface. **Journal of Neural Engineering**, Bristol, v. 10, n. 046003, p. 15, 2013.
- LARRAIN-VALENZUELA, J.; ZAMORANO, F.; SOTO-ICAZA, P.; CARRASCO, X.; HERRERA, C.; DAIBER, F.; ABOITIZ, F.; BILLEKE, P. Theta and alpha oscillation impairments in autistic spectrum disorder reflect working memory deficit. **Scientific Reports**, London, v. 7, n. 1, p. 14328, 2017.
- LEBEDEV, M. A.; NICOLELIS, M. A. Brain-machine interfaces: past, present and future. **Trends in Neurosciences**, Elsevier, v. 29, n. 9, p. 536--546, 2006.
- LEVIN, M. F.; SVEISTRUP, H.; SUBRAMANIAN, S. K. Feedback and virtual environments for motor learning and rehabilitation. **Schedae**, Krakow, v. 1, p. 19--36, 2010.
- LI, Y.; BIN, G.; GAO, X.; HONG, B.; GAO, S. Analysis of phase coding SSVEP based on canonical correlation analysis (CCA). In: INTERNATIONAL IEEE/EMBS CONFERENCE ON NEURAL ENGINEERING, 5th., 2011, Cancun. **Proceedings [...]**. Piscataway: IEEE, 2011. p. 368--371. ISSN 1948-3554.
- LIN, Z.; ZHANG, C.; WU, W.; GAO, X. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. **IEEE Transactions on Biomedical Engineering**, New York, v. 53, n. 12, p. 2610--2614, 2006.
- LOPEZ, M. A.; PELAYO, F.; MADRID, E.; PRIETO, A. Statistical characterization of steady-state visual evoked potentials and their use in brain-computer interfaces. **Neural Processing Letters**, New York, v. 29, n. 3, p. 179--187, 2009.

- MA, Y.; GUO, G. **Support vector machines applications**. New York: Springer, 2014.
- MACHADO, A. **Neuroanatomia funcional**. New York: Aesculapius, 1993.
- MAHMOOD, Z.; KHAN, S. On the use of k-fold cross-validation to choose cutoff values and assess the performance of predictive models in stepwise regression. **The International Journal of Biostatistics**, Berlin, v. 5, n. 1, p. 1–21, 2009.
- MANOLAKIS, D. G.; INGLE, V. K.; KOGON, S. M. **Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing**. Boston: McGraw-Hill, 2000.
- MARPLE, S. L. **Digital spectral analysis: with applications**. Englewood Cliffs: Prentice-Hall, 1987.
- MARTIN, J. H. **Neuroanatomy: text and atlas**. Stamford: Appleton & Lange, 1998.
- MÜLLER-PUTZ, G. R.; SCHERER, R.; BRAUNEIS, C.; PFURTSCHELLER, G. Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. **Journal of Neural Engineering**, Bristol, v. 2, n. 4, p. 123, 2005.
- NAKANISHI, M.; WANG, Y.; WANG, Y.-T.; JUNG, T.-P. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. **PloS One**, San Francisco, v. 10, n. 10, p. e0140703, 2015.
- NEWLAND, D. E. **An introduction to random vibrations, spectral & wavelet analysis**. New York: John Wiley & Sons, 1993.
- NIEDERMEYER, E.; SILVA, F. L. da. **Electroencephalography: basic principles, clinical applications, and related fields**. Philadelphia: Lippincott Williams & Wilkins, 2005.
- OIKONOMOU, V. P.; LIAROS, G.; GEORGIADIS, K.; CHATZILARI, E.; ADAM, K.; NIKOLOPOULOS, S.; KOMPATSIARIS, I. Comparative evaluation of state-of-the-art algorithms for SSVEP-based BCIs. [**S.I.**]: **Technical Report**, 2016. 30 p.
- OOSTENVELD, R.; PRAAMSTRA, P. The five percent electrode system for high-resolution EEG and ERP measurements. **Clinical Neurophysiology**, Shannon, v. 112, n. 4, p. 713–719, 2001.
- PASTOR, M. A.; ARTIEDA, J.; ARBIZU, J.; VALENCIA, M.; MASDEU, J. C. Human cerebral activation during steady-state visual-evoked responses. **Journal of Neuroscience**, Washington, v. 23, n. 37, p. 11621–11627, 2003.
- RAZALI, N. M.; WAH, Y. B. et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. **Journal of Statistical Modeling and Analytics**, Maryland Heights, v. 2, n. 1, p. 21–33, 2011.
- RODGERS, J. L.; NICEWANDER, W. A. Thirteen ways to look at the correlation coefficient. **The American Statistician**, Alexandria, v. 42, n. 1, p. 59–66, 1988.

ROYSTON, P. Approximating the shapiro-wilk w-test for non-normality. **Statistics and Computing**, New York: Springer, v. 2, n. 3, p. 117--119, 1992.

_____. A toolkit for testing for non-normality in complete and censored samples. **Journal of the Royal Statistical Society. Series D (The Statistician)**, [S. I.]: JSTOR, p. 37--43, 1993.

RUGE, D.; LIOU, L.-M.; HOAD, D. Improving the potential of neuroplasticity. **Journal of Neuroscience**, Washington, v. 32, n. 17, p. 5705--5706, 2012.

SAGAHÓN-AZÚA, J.; TOVAR-CORONA, B.; ZÚNIGA-VALLADARES, J. A. Comparison between the canonical correlation analysis and the support vector machines as classification algorithms in an SSVEP-based brain-computer interface. In: INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING, COMPUTING SCIENCE AND AUTOMATIC CONTROL (CCE), 14th., 2017, Mexico City. **Proceedings [...]**. Piscataway: IEEE, 2017. p. 1--6.

SCHALK, G.; MCFARLAND, D. J.; HINTERBERGER, T.; BIRBAUMER, N.; WOLPAW, J. R. BCI2000: a general-purpose brain-computer interface (BCI) system. **IEEE Transactions on Biomedical Engineering**, Piscataway, v. 51, n. 6, p. 1034--1043, 2004.

SEPULVEDA, R.; MONTIEL, O.; DIAZ, G.; GUTIERREZ, D.; CASTILLO, O. Classification of encephalographic signals using artificial neural networks. **Computación y Sistemas**, [S. I.], v. 19, n. 1, p. 69--88, 2015.

SHANNON, C. E. Communication in the presence of noise. **Transactions of the IRE**, Piscataway, v. 37, n. 1, p. 10--21, 1949.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, Oxford, v. 52, n. 3/4, p. 591--611, 1965.

SHIAVI, R. **Introduction to applied statistical signal analysis: guide to biomedical and electrical engineering applications**. [S.l.]: Elsevier, 2006.

SINGI, G. **Fisiologia dinâmica**. [S.l.]: Atheneu, 1996.

SIULY, S. **Analysis and classification of EEG signals**. Tese (Doutorado) --- University of Southern Queensland, Australia, 2012.

SNEE, R. D. Validation of regression models: methods and examples. **Technometrics**, Philadelphia, v. 19, n. 4, p. 415--428, 1977.

STAWICKI, P.; GEMBLER, F.; VOLOSYAK, I. Evaluation of suitable frequency differences in SSVEP-based BCIs. In: INTERNATIONAL WORKSHOP ON SYMBIOTIC INTERACTION, 2015, [S. L.]. **Proceedings [...]**. [S. l.: s. n.], 2015. p. 159--165.

SUGIARTO, I.; ALLISON, B.; GRÄSER, A. Optimization strategy for SSVEP-based BCI in spelling program application. In: INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING AND TECHNOLOGY, 2009, Singapore. **Proceedings [...]**. Piscataway: IEEE, 2009. v. 1, p. 223--226.

- TAX, D. M.; DUIN, R. P. Data domain description using support vectors. In: EUROPEAN SYMPOSIUM ON ARTIFICIAL NEURAL NETWORKS, 1999, Bruges. **Proceeding** [...]. [S.l.]: D-Facto, 1999. v. 99, p. 251--256.
- UHLHAAS, P. J.; HAENSCHHEL, C.; NIKOLIĆ, D.; SINGER, W. The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. **Schizophrenia Bulletin**, Oxford, v. 34, n. 5, p. 927--943, 2008.
- VOLOSYAK, I.; CECOTTI, H.; GRÄSER, A. Optimal visual stimuli on lcd screens for ssvep based brain-computer interfaces. In: INTERNATIONAL IEEE/EMBS CONFERENCE ON NEURAL ENGINEERING, 4th., 2009, Antalya. **Proceedings** [...]. Piscataway: IEEE, 2009. p. 447--450.
- WANG, Y.; WANG, R.; GAO, X.; HONG, B.; GAO, S. A practical VEP-based brain-computer interface. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, Piscataway, v. 14, n. 2, p. 234--240, 2006.
- WEBSTER, J. G. **Encyclopedia of medical devices and instrumentation**. New York: John Wiley & Sons, 2006.
- WEI, C.; LIN, Y.; WANG, Y.; WANG, Y.; JUNG, T. Detection of steady-state visual-evoked potential using differential canonical correlation analysis. In: INTERNATIONAL IEEE/EMBS CONFERENCE ON NEURAL ENGINEERING (NER), 6th., 2013, San Diego. **Proceedings** [...]. Piscataway: IEEE, 2013. p. 57--60. ISSN 1948-3554.
- WOLPAW, J. R.; BIRBAUMER, N.; MCFARLAND, D. J.; PFURTSCHELLER, G.; VAUGHAN, T. M. Brain--computer interfaces for communication and control. **Clinical neurophysiology**, Shannon, v. 113, n. 6, p. 767--791, 2002.
- WORTZEL, H. S. The american psychiatric publishing textbook of neuropsychiatry and behavioral neurosciences. **American Journal of Psychiatry**, Washington, v. 165, n. 2, p. 267--268, 2008.
- WU, M.; CHEN, D.; CHEN, G. New spectral leakage-removing method for spectral testing of approximate sinusoidal signals. **IEEE Transactions on Instrumentation and Measurement**, Piscataway, v. 61, n. 5, p. 1296--1306, 2012.
- XIE, S.; LIU, C.; OBERMAYER, K.; ZHU, F.; WANG, L.; XIE, X.; WANG, W. Stimulator selection in SSVEP-based spatial selective attention study. **Computational Intelligence and Neuroscience**, Hindawi, v. 2016, p. 1--9, 2016.
- YAP, B. W.; SIM, C. H. Comparisons of various types of normality tests. **Journal of Statistical Computation and Simulation**, Abingdon, v. 81, n. 12, p. 2141--2155, 2011.
- YEAGER, M.; GREGORY, B.; KEY, C.; TODD, M. On using robust mahalanobis distance estimations for feature discrimination in a damage detection scenario. **Structural Health Monitoring**, London, v. 18, n. 1, p. 245--253, 2018.
- YEKKEHKHANY, B.; SAFARI, A.; HOMAYOUNI, S.; HASANLOU, M. A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry sar data. **The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, Goettingen, v. 40, n. 2, p. 281, 2014.

YU, H.; YANG, J. A direct LDA algorithm for high-dimensional data with application to face recognition. **Pattern Recognition**, Amsterdam, v. 34, n. 10, p. 2067--2070, 2001.

ZHANG, Y. **Yu Zhang Research**. 2018. Disponível em: <http://yuzhangresearch.weebly.com/research.html>. Acesso em: 22 out. 2018..

ZHANG, Y.; YIN, E.; LI, F.; ZHANG, Y.; TANAKA, T.; ZHAO, Q.; CUI, Y.; XU, P.; YAO, D.; GUO, D. Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 26, n. 7, p. 1314--1323, 2018.

ZHANG, Y.; ZHOU, G.; JIN, J.; WANG, M.; WANG, X.; CICHOCKI, A. L1-regularized multiway canonical correlation analysis for SSVEP-based BCI. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, Piscataway, v. 21, n. 6, p. 887--896, 2013.

ZHANG, Y.; ZHOU, G.; JIN, J.; WANG, X.; CICHOCKI, A. Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis. **International Journal of Neural Systems**, Singapore, v. 24, n. 04, p. 1450013, 2014.

ZHU, D.; BIEGER, J.; MOLINA, G. G.; AARTS, R. M. A survey of stimulation methods used in SSVEP-based BCIs. **Computational Intelligence and Neuroscience**, New York, v. 2010, p. 1--13, 2010.